ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 84, No. 4 (July, 2016), 1613–1633

GENERALIZED METHOD OF INTEGRATED MOMENTS FOR HIGH-FREQUENCY DATA

JIA LI
Duke University, Durham, NC 27708, U.S.A.

DACHENG XIU
University of Chicago Booth School of Business, Chicago, IL 60637, U.S.A.

Econometrica, Vol. 84, No. 4 (July, 2016), 1613–1633

NOTES AND COMMENTS

GENERALIZED METHOD OF INTEGRATED MOMENTS FOR HIGH-FREQUENCY DATA

BY JIA LI AND DACHENG XIU[1]

We propose a semiparametric two-step inference procedure for a finite-dimensional parameter based on moment conditions constructed from high-frequency data. The population moment conditions take the form of temporally integrated functionals of state-variable processes that include the latent stochastic volatility process of an asset. In the first step, we nonparametrically recover the volatility path from high-frequency asset returns. The nonparametric volatility estimator is then used to form sample moment functions in the second-step GMM estimation, which requires the correction of a high-order nonlinearity bias from the first step. We show that the proposed estimator is consistent and asymptotically mixed Gaussian and propose a consistent estimator for the conditional asymptotic variance. We also construct a Bierens-type consistent specification test. These infill asymptotic results are based on a novel empirical-process-type theory for general integrated functionals of noisy semimartingale processes.

KEYWORDS: High-frequency data, semimartingale, spot volatility, nonlinearity bias, GMM.

1. INTRODUCTION

IN THIS PAPER, we study a novel variant of the GMM (Hansen (1982)) for estimating moment equalities using high-frequency intraday data in certain types of derivative pricing and market microstructure models.[2] The moment conditions take the form of temporally integrated functionals of the sample paths of state variables, such as time, the asset price, and, importantly, the latent stochastic volatility. Volatility is the primary measure of risk in modern finance (Engle (2004)), and its unobservability poses a substantial challenge for inference.

[1]This paper supersedes our working paper previously circulated under the title "Spot Variance Regressions," containing substantially more general theoretical results. We are grateful to four anonymous referees and a co-editor for many comments and suggestions that have greatly improved the paper. We also thank Yacine Aït-Sahalia, Torben Andersen, Federico Bandi, Alan Bester, Tim Bollerslev, Federico Bugni, Marine Carrasco, Chris Hansen, Michael Jansson, Zhipeng Liao, Oliver Linton, Nour Meddahi, Ulrich Müller, Per Mykland, Andrew Patton, Eric Renault, Jeff Russell, George Tauchen, Viktor Todorov, Lan Zhang, as well as many seminar and conference participants at the University of Chicago, Brown University, Toulouse School of Economics, the 2012 Triangle Econometrics Conference, the 2013 Financial Econometrics Conference at Toulouse School of Economics, the 6th Annual SoFiE Conference, the 2013 workshop on "Measuring and Modeling Financial Risk with High Frequency Data" at EUI, and the 2014 conference on Inference in Nonstandard Problems for their helpful comments. Li's work was partially supported by NSF Grants SES-1227448 and SES-1326819. Xiu's work was supported in part by the FMC and IBM Corporation Faculty Scholar Funds at the University of Chicago Booth School of Business.

[2]Moment-based estimation can be dated back to Pearson (1894).

 DOI: 10.3982/ECTA12306

The common solution to the latent volatility problem in the classical time-series setting is to impose auxiliary parametric restrictions on volatility dynamics; see Bollerslev, Engle, and Nelson (1994) and Ghysels, Harvey, and Renault (1996) for reviews. However, Andersen, Bollerslev, and Lange (1999) found that standard parametric volatility models are unsatisfactory for modeling asset return and its volatility at high frequencies. It is therefore prudent to consider a nonparametric approach as a complement.[3] Indeed, the last decade has seen a large and burgeoning literature on nonparametric inference for volatility, which harnesses the rich information in high-frequency data.

We propose a semiparametric two-step estimation procedure based on integrated moment equalities. In the first step, we use a nonparametric high-frequency spot volatility estimator (Foster and Nelson (1996), Comte and Renault (1998)) to recover the volatility process. We then use it to form sample moment functions in the second step, where we estimate the finite-dimensional parameter of interest by minimizing a GMM criterion function. The inference procedure is justified in an infill (nonergodic) asymptotic setting for general semimartingales, which allows for an essentially unrestricted form of serial dependence and nonstationarity for state variables.

While our infill asymptotic setting is nonstandard, the proposed two-step estimation procedure formally resembles the semiparametric two-step GMM. As in conventional semiparametric settings, the resultant estimator for the finite-dimensional parameter of interest attains the $n^{1/2}$-rate of convergence (n is the sample size), although it is built on the nonparametric volatility estimates that converge at a slower rate.[4] That noted, there is an additional novel aspect in our analysis. On the one hand, it was common in prior work to consider sufficient conditions that ensure the nonparametric ingredient converges at a rate faster than $n^{1/4}$, so that the semiparametric estimator depends asymptotically linearly on the nonparametric ingredient, whereas the nonlinearity bias can be tuned to be asymptotically negligible.[5] On the other hand, it is well known in high-frequency literature that the optimal convergence rate for (pointwise) spot variance estimation is only $n^{1/4}$. As a result, we need to explicitly correct for the high-order nonlinearity bias term induced by the nonparametric volatility estimation. A bias-corrected sample moment function is hence used in our second-step GMM estimation.

[3]Although it is subject to the risk of misspecification, a tight parametric specification may have several advantages over a nonparametric approach, such as better statistical efficiency, better finite and out-of-sample performance, simplicity of interpretation and real-time control, etc.

[4]See, for example, Newey (1994), Section 8 in Newey and McFadden (1994), and Section 4 in Chen (2007), as well as many references therein, for results on semiparametric two-step estimation.

[5]Correcting the nonlinearity bias has been emphasized by Cattaneo, Crump, and Jansson (2013) in the study of estimators for weighted average derivatives with independent and identically distributed data.

We show that the proposed estimator is consistent and has a mixed Gaussian asymptotic distribution. Overidentification tests and Anderson–Rubin type confidence sets (Stock and Wright (2000), Andrews and Soares (2010)) are also discussed as by-products. We also construct a Bierens (1982)-type consistent specification test based on testing a continuum of integrated moment equalities, for which we develop a novel empirical-process-type asymptotic theory.

This paper extends the empirical scope of high-frequency econometrics on volatility estimation in an important direction. Although prior work has focused on the inference for the volatility process per se, the current paper goes one step further towards economic applications by proposing a general econometric framework for studying the economic relationship between volatility and other economic variables such as derivative prices and volume, all under the guidance of economic theory. We model these dependent variables as semimartingales contaminated (possibly nonadditively) by noise, where the noise terms are allowed to be conditionally weakly dependent. Technically speaking, these complications set our empirical-process-type asymptotic theory for general integrated functionals apart from other recent work such as that by Jacod and Rosenbaum (2013). Our analysis for the functional heteroscedasticity and autocorrelation consistent (HAC) estimation of the asymptotic covariance function associated with the dependent noise terms is also new because of the nonstandard infill asymptotic setting.

This paper is organized as follows. Section 2 presents the setting and Section 3 presents the main theory. The Supplemental Material to this paper (Li and Xiu (2016)) contains all proofs.

2. THE SETTING

Section 2.1 formalizes the probabilistic setting underlying our analysis. Section 2.2 introduces the econometric model of interest. The following notations are used in the sequel. The transpose of a matrix A is denoted by $A^\intercal$. The (i, j) elements of A, A_t, A_n are denoted by A_{ij}, $A_{ij,t}$, and $A_{ij,n}$, respectively. Let $\lambda_{\min}(\cdot)$ denote the smallest eigenvalue. All vectors are column vectors. We write (a, b) in place of $(a^\intercal, b^\intercal)^\intercal$ for simplicity. We denote the d-dimensional identity matrix and the d-vector of 1s by $\mathbf{I}_d$ and $\mathbf{J}_d$, respectively. The Euclidean norm is denoted by $\|\cdot\|$. We use $\lfloor\cdot\rfloor$ to denote the largest smaller integer function. A function $(x, y) \mapsto f(x, y)$ is said to be in $\mathcal{C}^{j,k}$ if it is j (resp. k) times continuously differentiable in x (resp. y). The symbol $\otimes$ denotes the Kronecker product and the product of σ-fields. The Hadamard product is denoted by $\odot$. All limits are for $n \to \infty$. We write $a_n \asymp b_n$ if for some $c \geq 1$, $b_n/c \leq a_n \leq cb_n$ for all n. We use $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{\mathcal{L}\text{-}s}$ to denote convergence in probability and stable convergence in law, respectively. We use $\mathcal{MN}$ to denote the mixed normal distribution.

2.1. *The Underlying Processes*

We observe a data sequence $(X_{i\Delta_n}, Z_{i\Delta_n}, Y_{i\Delta_n})$ at discrete times $i\Delta_n$, $0 \le i \le n \equiv \lfloor T/\Delta_n \rfloor$, within a *fixed* time span $[0, T]$, with the sampling interval $\Delta_n \to 0$ asymptotically. In applications, X typically denotes the (logarithmic) asset price, Z denotes observable state variables, and Y denotes dependent variables such as prices of derivative contracts, trading volumes, etc. In the analysis below, we assume that Z_t includes time (i.e., t) as its first component without loss of generality.

Formally, we consider a filtered probability space $(\Omega^{(0)}, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P}^{(0)})$ and, without loss, set $\mathcal{F} = \mathcal{F}_T$. We endow this probability space with càdlàg (i.e., right continuous with left limit) adapted processes X_t, Z_t, and β_t which, respectively, take values in open sets $\mathcal{X}$, $\mathcal{Z}$, and $\mathcal{B}$. The process β_t is not observable; instead, we observe its noisy transform $Y_{i\Delta_n}$ in discrete time.

In order to introduce the noise terms, we consider another probability space $(\Omega^{(1)}, \mathcal{G}, \mathbb{P}^{(1)})$ that is endowed with a stationary ergodic sequence $(\chi_i)_{i\in\mathbb{Z}}$, where $\mathbb{Z}$ denotes the set of integers and χ_i takes value in a Polish space with its marginal law denoted by $\mathbb{P}_\chi$. We stress from the outset that we do not assume the sequence $(\chi_i)_{i\ge 0}$ to be serially independent. Let $\Omega = \Omega^{(0)} \times \Omega^{(1)}$ and $\mathbb{P} = \mathbb{P}^{(0)} \otimes \mathbb{P}^{(1)}$. Processes defined in each space, $\Omega^{(0)}$ or $\Omega^{(1)}$, are extended in the usual way to the product space $(\Omega, \mathcal{F} \otimes \mathcal{G}, \mathbb{P})$, which serves as the probability space underlying our analysis. For the sake of notational simplicity, we identify the σ-field $\mathcal{F}_t$ with its trivial extension $\mathcal{F}_t \otimes \{\emptyset, \Omega^{(1)}\}$ in the product space. By construction, the sequence $(\chi_i)_{i\in\mathbb{Z}}$ is independent of $\mathcal{F}$.

We model $Y_{i\Delta_n}$ as a noisy transform of $\beta_{i\Delta n}$ given by

$$Y_{i\Delta_n} = \mathscr{Y}(\beta_{i\Delta_n}, \chi_i), \quad i = 0, \dots, n, \tag{2.1}$$

where $\mathscr{Y}(\cdot)$ is a deterministic transform taking values in a finite-dimensional real space $\mathcal{Y}$. For example, if $Y_{i\Delta_n}$ is the observed price of a derivative contract, (2.1) often has a location-scale form $Y_{i\Delta_n} = \beta_{1,i\Delta_n} + \beta_{2,i\Delta_n}\chi_i$, where $\beta_{1,t}$ represents the efficient price and $\beta_{2,t}$ captures the heteroscedasticity of the pricing error component $\beta_{2,i\Delta_n}\chi_i$ in the observed price.

The basic regularity condition for the underlying processes is the following.

ASSUMPTION 1: *For some constant $r \in [0, 1)$ and a sequence $(T_m)_{m\ge 1}$ of stopping times increasing to ∞, we have the following*:

(i) *The process X_t is a one-dimensional Itô semimartingale on $(\Omega^{(0)}, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P}^{(0)})$ with the form*

$$X_t = X_0 + \int_0^t b_s\, ds + \int_0^t \sqrt{V_s}\, dW_s + \int_0^t \int_{\mathbb{R}} \delta(s, u)\mu(ds, du),$$

where the process b_t is locally bounded and adapted; *the spot variance process V_t takes values in $\mathcal{V} \equiv (0, \infty)$*; *$W_t$ is a standard Brownian motion*; $\delta : \Omega \times \mathbb{R}_+ \times \mathbb{R} \mapsto$

$\mathbb{R}$ is a predictable function and μ is a Poisson random measure with compensator ν of the form $\nu(dt, du) = dt \otimes \lambda(du)$ for some σ-finite measure λ on $\mathbb{R}$. Moreover, for a sequence $(J_m)_{m\geq 1}$ of λ-integrable deterministic functions, we have $|\delta(\omega^{(0)}, t, u)|^r \wedge 1 \leq J_m(u)$ for all $\omega^{(0)} \in \Omega^{(0)}$, $t \leq T_m$, and $u \in \mathbb{R}$.

(ii) *For a sequence $(\mathcal{K}_m)_{m\geq 1}$ of convex compact subsets of $\mathcal{V}$, $V_t \in \mathcal{K}_m$ for all $t \leq T_m$. For a sequence $(\mathcal{K}'_m)_{m\geq 1}$ of compact subsets of $\mathcal{B} \times \mathcal{Z}$, $(\beta_t, Z_t) \in \mathcal{K}'_m$ for all $t \leq T_m$.*

(iii) *The process $\tilde{Z}_t \equiv (\beta_t, Z_t, V_t)$ is also an Itô semimartingale on $(\Omega^{(0)}, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}^{(0)})$ with the form*

$$\tilde{Z}_t = \tilde{Z}_0 + \int_0^t \tilde{b}_s\, ds + \int_0^t \tilde{\sigma}_s\, d\tilde{W}_s + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s, u) 1_{\{\|\tilde{\delta}(s,u)\|\leq 1\}}(\mu - \nu)(ds, du) + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s, u) 1_{\{\|\tilde{\delta}(s,u)\|> 1\}}\mu(ds, du),$$

where $\tilde{b}_t$ and $\tilde{\sigma}_t$ are locally bounded adapted processes, $\tilde{W}_t$ is a (multivariate) Brownian motion, and $\tilde{\delta}$ is a predictable function such that for some deterministic λ-integrable function $\tilde{J}_m : \mathbb{R} \mapsto \mathbb{R}$, $\|\tilde{\delta}(\omega^{(0)}, t, u)\|^2 \wedge 1 \leq \tilde{J}_m(u)$ for all $\omega^{(0)} \in \Omega^{(0)}$, $t \leq T_m$, and $u \in \mathbb{R}$.

Assumption 1 accommodates many models in finance and is commonly used for deriving infill asymptotic results for high-frequency data; see, for example, Jacod and Protter (2012) and many references therein. This assumption allows for price and volatility jumps and imposes no restriction on the dependence among various components of studied processes. In particular, the Brownian shocks dW_t and $d\tilde{W}_t$ can be correlated, which accommodates the leverage effect (Black (1976)). The constant r serves as an upper bound for the generalized Blumenthal–Getoor index (i.e., the activity) of price jumps. Condition (iii) says that the spot variance process V_t is an Itô semimartingale with general forms of volatility-of-volatility and volatility jumps.[6] Although this condition admits many volatility models in finance, it does exclude an important class of long-memory volatility models that are driven by fractional Brownian motion; see Comte and Renault (1996, 1998). The generalization in this direction seems to deserve its own research. It is also important in future work to allow X to be contaminated with microstructure noise, which is particularly relevant for studying illiquid stocks; see, for example, Zhang, Mykland, and Aït-Sahalia (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008),

[6]Stochastic volatility models with multiple factors are also allowed, provided that each factor is an Itô semimartingale.

Jacod, Li, Mykland, Podolskij, and Vetter (2009) and Xiu (2010) for ways to handle such complications in the data in the context of volatility estimation.

2.2. *Integrated Moment Equalities and Examples*

The primary interest of this paper is the asymptotic inference for a finite-dimensional parameter θ^* that satisfies the following conditional moment equality:

$$\mathbb{E}\big[\psi\big(Y_{i\Delta_n}, Z_{i\Delta_n}, V_{i\Delta_n}; \theta^*\big)|\mathcal{F}\big] = 0 \quad \text{almost surely (a.s.)}, \tag{2.2}$$

where $\psi : \mathcal{Y} \times \mathcal{Z} \times \mathcal{V} \mapsto \mathbb{R}^{q_1}$, $q_1 \geq 1$, is a measurable function with its functional form known up to the unknown parameter θ^*, and the conditional expectation integrates out the error term χ_i in $Y_{i\Delta_n}$ (recall (2.1)). We suppose that the true parameter θ^* is deterministic and takes value in a compact parameter space $\Theta \subset \mathbb{R}^{\dim(\theta)}$. In the sequel, we use θ to denote a generic element in Θ.

We remark two differences between the conditional moment restriction (2.2) and that in the classical GMM setting in time series.[7] On the one hand, the conditioning information set in (2.2) is the entire σ-field $\mathcal{F} = \mathcal{F}_T$, instead of the smaller information set $\mathcal{F}_{i\Delta_n}$. Whereas conditional moment restrictions using the latter often arise from Euler equations in structural asset pricing models, (2.2) imposes a stronger exogeneity requirement on the state variables over the time interval $[0, T]$. It should be noted, however, that because the fixed sample span of our high-frequency analysis is much shorter than that in the classical long-span setting, the exogeneity requirement in the former is much weaker than it would be in the latter. We further note that the exogeneity requirement is quite standard in the literature on noisy high-frequency data.

On the other hand, in the high-frequency setting, we should and do allow the state variable processes X_t, β_t, Z_t, and V_t to exhibit essentially unrestricted serial dependence, which is more general than the classical GMM setting for weakly dependent data. Consequently, we do not use limit theorems for weakly dependent data as in the classical setting.[8] In our setting, the limiting distributions of sample moments and estimators are mixed Gaussian, which means that

[7]See Hansen and Singleton (1982); also see the recent extension in the asset pricing setting by Gagliardini, Gouriéroux, and Renault (2011) and references therein.

[8]In this aspect, our setting can be related to the common-shock regression of Andrews (2005). Andrews (2005) considered a setting where data are unconditionally strongly dependent, but weakly dependent (indeed independent and identically distributed by Assumption 1 there) conditional on a "common shock" σ-field, the role of which is played by $\mathcal{F}$ here. Andrews (2005) also illustrated the necessity of exogeneity conditions, which are analogous to (2.2), for the asymptotic validity of least-square estimators; see Assumptions CU and CMZ, as well as Corollary 1 in that paper.

their sampling variability depends on the realization of state variables, even in the asymptotic limit.

For concreteness, we consider two empirical examples for model (2.2).

EXAMPLE 1—Derivative Pricing: Let X_t be the price process of an underlying asset and Y_t be the price of a derivative contract written on it. We set $Z_t = (t, X_t, r_t, d_t)$, where r_t is the short interest rate and d_t is the dividend yield. If, under the risk-neutral measure, the process (Z_t, V_t) is Markovian,[9] then the theoretical price of the derivative can be written as a real-valued function $f(Z_t, V_t; \theta^*)$, where θ^* arises from the risk-neutral model for the dynamics of the state variables. Empirically, it is common to model the observed derivative price Y_t as the theoretical price plus a pricing error, that is,

$$Y_{i\Delta_n} = f\big(Z_{i\Delta_n}, V_{i\Delta_n}; \theta^*\big) + a_{i\Delta_n}\chi_i, \quad \mathbb{E}[\chi_i|\mathcal{F}] = 0, \quad \mathbb{E}\big[\chi_i^2|\mathcal{F}\big] = 1, \tag{2.3}$$

where a_t is the stochastic volatility of the pricing errors and the condition $\mathbb{E}[\chi_i^2|\mathcal{F}] = 1$ is a scale normalization. Note that (2.3) can be written in the form of (2.1) with $\beta_t \equiv (f(Z_t, V_t; \theta^*), a_t)$. Setting $\psi(Y_t, Z_t, V_t; \theta) = Y_t - f(Z_t, V_t; \theta)$, we can verify (2.2).

EXAMPLE 2—Volume-Volatility Relationship: Andersen (1996) proposed a Poisson model for the volume-volatility relationship for daily data, in which the conditional distribution of daily volume, given the return variance, is a scaled Poisson distribution. Here, we consider a version of his model for intraday data. Let $Y_{i\Delta_n}$ denote the trading volume of an asset within the interval $[i\Delta_n, (i+1)\Delta_n)$. Suppose that $Y_{i\Delta_n}|V_{i\Delta_n} \sim \theta_1^* \cdot \text{Poisson}(\theta_2^* + \theta_3^* V_{i\Delta_n})$. To cast this model in the form (2.1), we represent the Poisson distribution with a time-varying mean in terms of a time-changed Poisson process: let $\chi_i = (\chi_i(\beta))_{\beta\geq 0}$ be a standard Poisson process indexed by β and then set $\beta_t \equiv \theta_2^* + \theta_3^* V_t$ and $Y_{i\Delta_n} = \theta_1^* \chi_i(\beta_{i\Delta_n})$. This model can be estimated by using the first two conditional moments of Y_t. This amounts to setting

$$\psi(Y_t, V_t; \theta) = \begin{pmatrix} Y_t - \theta_1(\theta_2 + \theta_3 V_t) \\ Y_t^2 - \theta_1^2(\theta_2 + \theta_3 V_t)^2 - \theta_1^2(\theta_2 + \theta_3 V_t) \end{pmatrix}, \tag{2.4}$$

which readily verifies (2.2).

[9]Assuming that V_t is the only unobservable Markov state variable, this excludes derivative pricing models with multiple volatility factors under the risk-neutral measure. That said, this assumption does not imply that (Z_t, V_t) is Markov under the physical measure (i.e., $\mathbb{P}$), because the equivalence between measures imposes little restriction on drift and jump components of (Z_t, V_t). Hence, it is useful to consider the general Itô semimartingale setting (Assumption 1) under the physical measure even if one imposes additional restrictions under the risk-neutral measure. See Garcia, Ghysels, and Renault (2010) for a review of empirical option pricing.

As seen from these examples, empirical applications using high-frequency data naturally involve the latent spot variance process V_t, but are agnostic regarding the precise form of its dynamics (under the probability measure $\mathbb{P}$). This reaffirms the relevance of including V_t in (2.2) and treating it nonparametrically in our econometric theory. We also note that it is desirable to allow the studied processes to be nonstationary in these empirical settings. For example, option pricing usually includes time and the underlying asset price as observed state variables, both of which render the process Z_t nonstationary. Moreover, although it may be reasonable to assume that the stochastic volatility process is stationary in the classical long-span setting for daily or weekly data, the stationarity assumption is more restrictive for high-frequency data because of intradaily seasonalities.

We now describe the integrated moment conditions. To simplify notations, we define

$$\bar{\psi}(\beta, z, v; \theta) \equiv \int \psi\big(\mathscr{Y}(\beta, \chi), z, v; \theta\big)\mathbb{P}_\chi(d\chi), \tag{2.5}$$

so that we can rewrite (2.2) as $\bar{\psi}(\beta_{i\Delta_n}, Z_{i\Delta_n}, V_{i\Delta_n}; \theta^*) = 0$. Under the mild maintained assumption that the process $(\bar{\psi}(\beta_t, Z_t, V_t; \theta^*))_{t\geq 0}$ is càdlàg, the fact that (2.2) holds for all n is equivalent to the (seemingly stronger) condition

$$\bar{\psi}\big(\beta_t, Z_t, V_t; \theta^*\big) = 0, \quad t \in [0, T]. \tag{2.6}$$

In order to form integrated moment conditions, we consider a weight function $\varphi : \mathcal{Z} \times \mathcal{V} \times \Theta \mapsto \mathbb{R}^{q_2}$ for some $q_2 \geq 1$. Let $q = q_1 q_2$. We then set the q-dimensional integrated moment function to be

$$G(\theta) \equiv \int_0^T \bar{\psi}(\beta_s, Z_s, V_s; \theta) \otimes \varphi(Z_s, V_s; \theta)\, ds. \tag{2.7}$$

Clearly, (2.6) *implies* the integrated moment condition:

$$G\big(\theta^*\big) = 0. \tag{2.8}$$

Our two-step estimation procedure is to solve a sample version of (2.8). In the first step, we nonparametrically estimate the latent spot variance process V_t. In the second step, we use this nonparametric estimator to form a sample moment function $G_n(\cdot)$ for $G(\cdot)$. As mentioned in the Introduction, the nonparametric spot variance estimator converges at a rate no faster than $n^{1/4}$ and, hence, leads to a nonlinearity bias that needs to be corrected for obtaining asymptotic mixed normality.[10] The construction and analysis of this

[10]The aggregated estimation error in spot variance affects the asymptotic variance (at the $n^{1/2}$ convergence rate) in the second-step estimation through the first-order expansion term of the

bias-corrected sample moment function is the key to our asymptotic theory; see Section 3.1. With $G_n(\cdot)$ in hand, we then choose a sequence of positive semidefinite weighting matrices Ξ_n and estimate θ^* by minimizing the GMM objective function

$$(2.9) \qquad \hat{\theta}_n \equiv \operatorname*{argmin}_{\theta \in \Theta} Q_n(\theta), \quad \text{where} \quad Q_n(\theta) \equiv G_n(\theta)^\intercal \Xi_n G_n(\theta).$$

Section 3.2 presents the asymptotic property of $\hat{\theta}_n$, along with feasible inference procedures.

A further question, which we address in Section 3.3, is to conduct a specification test for (2.6). Following the insight of Bierens (1982), we carry out this test by examining whether a continuum of integrated moment conditions holds. Indeed, by properly choosing a continuum of weight functions, we can rewrite (2.6) *equivalently* as a continuum of integrated moment conditions. Our test is based on their joint asymptotic distribution, for which we need an empirical-process-type convergence result for the continuum of moment conditions.

We close this section with a few remarks on possible extensions in future research. First, it is possible to improve the estimation efficiency by using a continuum of weight functions in the spirit of Carrasco and Florens (2000). A substantive complication in doing so is that it involves an ill-posed problem and needs regularization (Carrasco, Florens, and Renault (2007)). Ill-posed problems remain a very open question in the infill high-frequency setting. In addition, since our setting also involves a nonparametric first-step estimation with a convergence rate no faster than $n^{1/4}$, the extension via Carrasco and Florens (2000) appears rather nontrivial. More generally, we note that the semiparametric efficient estimation in our setting is also an interesting but very challenging problem. Indeed, semiparametric efficient estimation in the high-frequency setting is well known to be nonstandard (cf. Bickel, Klaassen, Ritov, and Wellner (1998)) because of the lack of locally asymptotically normal likelihood ratios.[11] The efficient estimation of general integrated volatility functionals has been recently studied by Clément, Delattre, and Gloter (2013) and Renault, Sarisoy, and Werker (2014) under certain models, but these results

estimation equation. The nonlinearity bias arises from the second-order expansion term with respect to the spot variance estimate, where the latter needs "undersmoothing" so as to ensure that local diffusive and jump moves in volatility have asymptotically negligible effects; see Jacod and Rosenbaum (2013) for further discussion.

[11]That noted, the insight of some recent work on GMM may shed light on future research. In a setting with IID data, the semiparametric efficiency of the semiparametric two-step GMM is established by a recent paper of Ackerberg, Chen, Hahn, and Liao (2014). Their efficiency result is conditional on a given set of moments, which does not concern the optimal choice of instrument function; see footnote 13 of that paper. For dependent data, Carrasco and Florens (2014) address the semiparametric efficient GMM estimation in a Markovian setting, but they do not consider the nonparametric first-step estimation.

do not have direct implications in our setting because, here, the volatility process plays the role of a nuisance component for estimating θ^*, rather than the estimand of interest.

3. MAIN RESULTS

3.1. *The Bias-Corrected Sample Moment Function and Its Asymptotic Properties*

We now construct the sample moment function $G_n(\theta)$ associated with a generic function $g(y, z, v; \theta)$; for example, in the setting of (2.9), $g(\cdot)$ has the form $g(y, z, v; \theta) = \psi(y, z, v; \theta) \otimes \varphi(z, v; \theta)$. We derive an empirical-process-type stable convergence for $G_n(\theta)$. Because the results of this subsection are also useful in other applications, we present them in a general form. In particular, we interpret θ as a generic index of a random function, which is not necessarily the parameter that will be estimated in later applications (see Section 3.3).

We first nonparametrically recover the spot variance $V_{i\Delta_n}$ by using a spot truncated realized variation estimator (Jacod and Protter (2012)). To this end, we consider an integer sequence k_n of block sizes and a real sequence u_n of truncation threshold for eliminating jumps (Mancini (2001)). The spot variance estimator is then given as follows:[12] for each $0 \le i \le N_n \equiv \lfloor T/\Delta_n \rfloor - k_n$,

$$\widehat{V}_{i\Delta_n} \equiv \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \left(\Delta_{i+j}^n X\right)^2 1_{\{|\Delta_{i+j}^n X| \le u_n\}}, \quad \text{where} \tag{3.1}$$

$$\Delta_{i+j}^n X \equiv X_{(i+j)\Delta_n} - X_{(i+j-1)\Delta_n}.$$

We further set

$$\begin{aligned} \hat{g}_{n,i}(\theta) &\equiv g(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta), \\ \hat{g}_{n,i}''(\theta) &\equiv \partial_v^2 g(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta), \quad i \ge 0. \end{aligned} \tag{3.2}$$

We then define the bias-corrected sample moment function as

$$G_n(\theta) \equiv \Delta_n \sum_{i=0}^{N_n} \hat{g}_{n,i}(\theta) - \frac{\Delta_n}{k_n} \sum_{i=0}^{N_n} \hat{g}_{n,i}''(\theta) \widehat{V}_{i\Delta_n}^2. \tag{3.3}$$

As shown below, the limiting counterpart of $G_n(\theta)$ is

$$G(\theta) \equiv \int_0^T \bar{g}(\beta_s, Z_s, V_s; \theta)\, ds, \tag{3.4}$$

[12] The estimation of spot variance can be dated at least back to Foster and Nelson (1996) and Comte and Renault (1998), in a setting without jumps; also see Kristensen (2010).

where we denote

$$\bar{g}(\beta, z, v; \theta) \equiv \int g\big(\mathscr{Y}(\beta, \chi), z, v; \theta\big)\mathbb{P}_\chi(d\chi). \tag{3.5}$$

Comparing (3.3) with (3.4), we note that the first term on the right-hand side of (3.3) is a natural sample-analogue estimator of $G(\theta)$. However, this "raw" estimator does not admit a central limit theorem due to a high-order nonlinearity bias, which arises from the nonlinear dependency of $g(y, z, v; \theta)$ on v, combined with the fact that the spot variance can be estimated at a rate no faster than $n^{1/4}$. The second term in (3.3) corrects this nonlinearity bias in its closed form.

We now collect some regularity conditions for studying the asymptotic behavior of $G_n(\cdot)$. To this end, it is convenient to introduce a conditional norm and a conditional semimetric as follows: for $p \geq 1$, we set

$$\begin{cases} \bar{g}_p(\beta, z, v; \theta) \equiv \left(\displaystyle\sum_{j=0}^{2} \int \big\|\partial_v^j g\big(\mathscr{Y}(\beta, \chi), z, v; \theta\big)\big\|^p \mathbb{P}_\chi(d\chi)\right)^{1/p}, \\ \rho_p\big((\beta, z, v), (\beta', z', v'); \theta\big) \\ \quad \equiv \left(\displaystyle\int \big\|g\big(\mathscr{Y}(\beta, \chi), z, v; \theta\big) - g\big(\mathscr{Y}(\beta', \chi), z', v'; \theta\big)\big\|^p \mathbb{P}_\chi(d\chi)\right)^{1/p}. \end{cases} \tag{3.6}$$

In particular, $\rho_p(\cdot)$ is useful for quantifying the smoothness of $\mathcal{F}$-conditional moments (such as the autocovariance) of the sequence $(g(\mathscr{Y}(\beta_{i\Delta_n}, \chi_i), Z_{i\Delta_n}, V_{i\Delta_n}; \theta))_{i\geq 0}$ as functions of the state variables.

ASSUMPTION 2: *The following conditions hold for some constants $k > 2$ and $\kappa \in (0, 1]$*: (i) *$g(y, z, v; \theta)$ is twice continuously differentiable in v*; (ii) *the function $\bar{g}(\cdot) : \mathcal{B} \times \mathcal{Z} \times \mathcal{V} \times \Theta \mapsto \mathbb{R}^q$ is in $\mathcal{C}^{2,2,3,2}$*; (iii) *for each (β, z, v, θ), $\partial_\theta \bar{g}(\beta, z, v; \theta) = \int \partial_\theta g(\mathscr{Y}(\beta, \chi), z, v; \theta)\mathbb{P}_\chi(d\chi)$ and $\partial_v^j \bar{g}(\beta, z, v; \theta) = \int \partial_v^j g(\mathscr{Y}(\beta, \chi), z, v; \theta)\mathbb{P}_\chi(d\chi)$ for $j \in \{0, 1, 2\}$*; (iv) *$\bar{g}_{2k}(\cdot; \cdot)$ is bounded on bounded sets*; (v) *for each bounded set $\mathcal{K} \subseteq \mathcal{B} \times \mathcal{Z} \times \mathcal{V}$, there exists some finite $K > 0$ such that $\rho_k(\tilde{z}, \tilde{z}'; \theta) \leq K\|\tilde{z} - \tilde{z}'\|^\kappa$ for all $\theta \in \Theta$ and $\tilde{z}, \tilde{z}' \in \mathcal{K}$ with $\|\tilde{z} - \tilde{z}'\| \leq 1$, where $\tilde{z} \equiv (\beta, z, v)$ and $\tilde{z}' \equiv (\beta', z', v')$*; (vi) *$\sum_{j=0}^{2} \|\partial_v^j g(y, z, v; \theta) - \partial_v^j g(y, z, v; \theta')\| \leq B(y, z, v)\|\theta - \theta'\|$ for all θ, θ' and $(y, z, v) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{V}$ and the function $(\beta, z, v) \mapsto \bar{B}_k(\beta, z, v) \equiv (\int B(\mathscr{Y}(\beta, \chi), z, v)^k \mathbb{P}_\chi(d\chi))^{1/k}$ is bounded on bounded sets*.

ASSUMPTION 3: *$k_n \asymp \Delta_n^{-\varsigma}$ and $u_n \asymp \Delta_n^{\varpi}$ for some $\varsigma \in (\frac{r}{2} \vee \frac{1}{3}, \frac{1}{2})$ and $\varpi \in [\frac{1-\varsigma}{2-r}, \frac{1}{2})$.*

ASSUMPTION 4: *The sequence $(\chi_i)_{i\in\mathbb{Z}}$ is stationary and α-mixing with mixing coefficient $\alpha_{\mathrm{mix}}(\cdot)$ such that $\sum_{j\geq 1} j\alpha_{\mathrm{mix}}(j)^{(k-2)/k} < \infty$ for some $k > 2$.*

ASSUMPTION 5: *We have* (i) *for some* $k' \in [2, k)$ *with* $\dim(\theta) < k'$, $\sum_{j\geq 1} \alpha_{\text{mix}}(j)^{1/k'-1/k} < \infty$; (ii) *either the process* V_t *is continuous or* $\dim(\theta) < 2(1-\varsigma)/\varsigma$.

Assumption 2 mainly concerns the smoothness of the function $g(\cdot)$, and it is easy to verify in applications. Note that condition (iii) of Assumption 2 concerns the interchangeability of integration and differentiation, for which sufficient conditions are well known. Unlike in Jacod and Rosenbaum (2013), the function $g(\cdot)$ is not restricted to having polynomial growth in the spot variance. This generality is important in applied work. We achieve this by using a proof technique which is notably different from that used by Jacod and Rosenbaum (2013).

Assumption 5 is used only for proving the functional central limit theorem (FCLT) (see Theorem 1(c)). Condition (i) is inspired by Hansen (1996) and is used for establishing stochastic equicontinuity for α-mixing sequences. Condition (ii) is sufficient for the stochastic equicontinuity of aggregated error terms that arise from spot variance estimation. While Assumption 5 restricts the dimension of the index θ, we actually only apply the FCLT in this paper with a one-dimensional index (i.e., τ in Section 3.3), for which Assumption 5 holds trivially (because $\varsigma < 1/2$ and $k' \geq 2$). Nevertheless, we prove the FCLT under this general setting, which may be useful in future work.

Theorem 1, below, shows that $\Delta_n^{-1/2}(G_n(\cdot) - G(\cdot))$ converges stably in law to an $\mathcal{F}$-conditionally centered Gaussian process.[13] We now describe the asymptotic conditional covariance function of the limiting process. We set, for $\theta, \theta' \in \Theta$,

$$\bar{S}_g(\theta, \theta') \equiv 2\int_0^T \partial_v \bar{g}(\beta_s, Z_s, V_s; \theta)\partial_v \bar{g}(\beta_s, Z_s, V_s; \theta')^\intercal V_s^2\, ds, \tag{3.7}$$

and, for $(\beta, z, v) \in \mathcal{B} \times \mathcal{Z} \times \mathcal{V}$,

$$\begin{aligned}\gamma_{g,l}(\beta, z, v; \theta, \theta') \\ \equiv \operatorname{Cov}\big(g(\mathscr{Y}(\beta, \chi_i), z, v; \theta), g(\mathscr{Y}(\beta, \chi_{i-l}), z, v; \theta')\big), \quad l \geq 0,\end{aligned} \tag{3.8}$$

$$\begin{aligned}\bar{\gamma}_g(\beta, z, v; \theta, \theta') \\ \equiv \gamma_{g,0}(\beta, z, v; \theta, \theta') + \sum_{l=1}^{\infty}\big(\gamma_{g,l}(\beta, z, v; \theta, \theta') + \gamma_{g,l}(\beta, z, v; \theta', \theta)^\intercal\big).\end{aligned} \tag{3.9}$$

[13] Stable convergence in law is stronger than the usual notion of weak convergence. It requires that the convergence holds jointly with any bounded $\mathcal{F}$-measurable random variable defined on the original probability space. Its importance for our problem stems from the fact that the limiting variable of our estimator is an $\mathcal{F}$-conditionally Gaussian process, and stable convergence allows for feasible inference using a consistent estimator for its $\mathcal{F}$-conditional variance–covariance function. See Jacod and Shiryaev (2003) for further details on stable convergence.

We then set

$$\bar{\Gamma}_g(\theta,\theta') \equiv \int_0^T \bar{\gamma}_g(\beta_s, Z_s, V_s; \theta, \theta')\,ds. \tag{3.10}$$

We note that $\bar{\gamma}_g(\beta, z, v; \theta, \theta')$ is the "long-run" covariance between the two sequences $g(\mathscr{Y}(\beta, \chi_i), z, v; \theta)$ and $g(\mathscr{Y}(\beta, \chi_i), z, v; \theta')$.[14]

With these notations, the $\mathcal{F}$-conditional asymptotic covariance function of $\Delta_n^{-1/2}(G_n(\cdot) - G(\cdot))$ can be written as

$$\Sigma_g(\theta,\theta') \equiv \bar{S}_g(\theta,\theta') + \bar{\Gamma}_g(\theta,\theta'), \quad \text{for} \quad \theta, \theta' \in \Theta. \tag{3.11}$$

We note that the term $\bar{S}_g(\cdot,\cdot)$ captures the sampling variability due to the estimator $\widehat{V}_{i\Delta_n}$, and $\bar{\Gamma}_g(\cdot,\cdot)$ captures the sampling variability due to the serially dependent random errors χ_i.

THEOREM 1: *Under Assumptions 1–4, the following statements hold*:

(a) $G_n(\theta) \xrightarrow{\mathbb{P}} G(\theta)$ *uniformly in* $\theta \in \Theta$;

(b) *for each* $\theta \in \Theta$, $\Delta_n^{-1/2}(G_n(\theta) - G(\theta)) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, \Sigma_g(\theta,\theta))$;

(c) *if Assumption 5 holds in addition, then the sequence* $\Delta_n^{-1/2}(G_n(\cdot) - G(\cdot))$ *converges stably in law under the uniform metric to a process which, conditional on* $\mathcal{F}$, *is centered Gaussian with covariance function* $\Sigma_g(\cdot,\cdot)$ *given by* (3.11).

3.2. *Asymptotic Inference for* θ^*

In this subsection, we study the estimator $\hat{\theta}_n$ given by (2.9), where the sample moment function $G_n(\cdot)$ is defined by (3.3) with

$$g(y, z, v; \theta) = \psi(y, z, v; \theta) \otimes \varphi(z, v; \theta).$$

Hence, $\bar{g}(\beta, z, v; \theta) = \bar{\psi}(\beta, z, v; \theta) \otimes \varphi(z, v; \theta)$ and $G(\theta)$ in (3.4) coincides with that in (2.7). Below, we maintain Assumption 6, where we denote $H(\theta) \equiv \int_0^T \partial_\theta \bar{g}(\beta_s, Z_s, V_s; \theta)\,ds$ and $H \equiv H(\theta^*)$.

ASSUMPTION 6: (i) $\Xi G(\theta) = 0$ *a.s. only if* $\theta = \theta^*$; (ii) $\Xi_n \xrightarrow{\mathbb{P}} \Xi$, *where* Ξ *is an* $\mathcal{F}$-*measurable* (*random*) *matrix that is positive semidefinite a.s.*; (iii) θ^* *is in the interior of the compact set* Θ; (iv) *the random matrix* $H^\intercal \Xi H$ *is nonsingular a.s.*

[14]The process $\bar{\gamma}_g(\beta_t, Z_t, V_t)$ may be more properly referred to as the *local* long-run covariance matrix, in that it is evaluated locally at time t. It arises from a large number of adjacent observations that are serially dependent (through χ_i), but all these observations are sampled from an asymptotically shrinking time window. In other words, $\bar{\gamma}_g(\beta_t, Z_t, V_t)$ is long run in tick time, but local in calendar time.

Assumption 6 is essentially standard for GMM estimation, though it is slightly modified so as to accommodate our pathwise inference setting. We note that condition (i) ensures identification, but that it takes a somewhat nonstandard form because the population moment function $G(\cdot)$ is itself a random function. It is instructive to illustrate the nature of this condition in a linear regression setting. Consider Example 1 with $f(z, v; \theta) = \theta_1 + \theta_2 v$ and $\varphi(v) = (1, v)$. In this case,

$$G(\theta) = \begin{pmatrix} T & \int_0^T V_s\,ds \\ \int_0^T V_s\,ds & \int_0^T V_s^2\,ds \end{pmatrix} \begin{pmatrix} \theta_1^* - \theta_1 \\ \theta_2^* - \theta_2 \end{pmatrix}. \tag{3.12}$$

It is easy to see that the identification of θ^* is achieved as soon as the process V_t does not remain constant on the interval $[0, T]$.

The asymptotic behavior of the estimator $\hat{\theta}_n$ is summarized by Theorem 2 below. With Theorem 1 in hand, we can derive the asymptotics of $\hat{\theta}_n$ by using a linearization argument from the classical GMM literature. Indeed, it can be shown that $\hat{\theta}_n$ has an asymptotically linear representation:

$$\Delta_n^{-1/2}\left(\hat{\theta}_n - \theta^*\right) = -\left(H^\intercal \Xi H\right)^{-1} H^\intercal \Xi G_n\left(\theta^*\right) + o_p(1). \tag{3.13}$$

From Theorem 1(b), it follows that $\Delta_n^{-1/2}(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, \Sigma^*)$, where

$$\Sigma^* \equiv \left(H^\intercal \Xi H\right)^{-1} H^\intercal \Xi \Sigma_g\left(\theta^*, \theta^*\right) \Xi H \left(H^\intercal \Xi H\right)^{-1}. \tag{3.14}$$

We now describe a consistent estimator of the $\mathcal{F}$-conditional asymptotic covariance matrix Σ^*. We set, for $i \geq 0$,

$$\hat{m}_{n,i}(g, \theta) \equiv \frac{1}{k_n} \sum_{j=0}^{k_n - 1} g(Y_{(i+j)\Delta_n}, Z_{(i+j)\Delta_n}, \widehat{V}_{i\Delta_n}; \theta),$$

$$\hat{\delta}_{n,i}(g, \theta) \equiv \hat{g}_{n,i}(\theta) - \hat{m}_{n,i}(g, \theta),$$

$$\hat{m}'_{n,i}(g, \theta) \equiv \frac{1}{k_n} \sum_{j=0}^{k_n - 1} \partial_v g(Y_{(i+j)\Delta_n}, Z_{(i+j)\Delta_n}, \widehat{V}_{i\Delta_n}; \theta).$$

The local estimates $\hat{m}_{n,i}(g, \theta)$ and $\hat{m}'_{n,i}(g, \theta)$ are used to approximate the processes $\bar{g}(\beta_t, Z_t, V_t; \theta)$ and $\partial_v \bar{g}(\beta_t, Z_t, V_t; \theta)$, respectively. The variable $\hat{\delta}_{n,i}(g, \theta)$ is the locally centered counterpart of $\hat{g}_{n,i}(\theta)$; removing the local mean is useful for analyzing the asymptotic behavior of the HAC estimator under misspecification, which is needed in Section 3.3. Under correct specification, we can use $\hat{g}_{n,i}(\theta)$ in place of $\hat{\delta}_{n,i}(g, \theta)$ without affecting our results.

We also consider a sequence B_n of integers and a kernel function $w(l, B_n)$ that satisfy the following.

ASSUMPTION 7: (i) *The kernel function* $w(\cdot,\cdot)$ *is uniformly bounded and, for each* $l \geq 1$, $\lim_{B_n\to\infty} w(l, B_n) = 1$; (ii) $B_n \to \infty$ *and* $B_n k_n^{-\kappa/2} \to 0$.

We estimate $\Sigma_g(\theta^*, \theta^*)$ by using $\widehat{\Sigma}_{g,n}(\hat{\theta}_n)$ defined as

$$\text{(3.15)}\quad \begin{cases} \widehat{\Sigma}_{g,n}(\hat{\theta}_n) \equiv \widehat{S}_{g,n}(\hat{\theta}_n) + \widehat{\Gamma}_{g,n}(\hat{\theta}_n), \quad \text{where} \\ \widehat{S}_{g,n}(\hat{\theta}_n) \equiv 2\Delta_n \sum_{i=0}^{N_n} \hat{m}'_{n,i}(g,\theta)\hat{m}'_{n,i}(g,\theta)^\intercal \widehat{V}^2_{i\Delta_n}, \\ \widehat{\Gamma}_{g,n}(\hat{\theta}_n) \equiv \widehat{\Gamma}_{g,0,n}(\hat{\theta}_n) + \sum_{l=1}^{B_n} w(l, m_n)\big(\widehat{\Gamma}_{g,l,n}(\hat{\theta}_n) + \widehat{\Gamma}_{g,l,n}(\hat{\theta}_n)^\intercal\big), \\ \widehat{\Gamma}_{g,l,n}(\hat{\theta}_n) \equiv \Delta_n \sum_{i=l}^{N_n} \hat{\delta}_{n,i}(g,\hat{\theta}_n)\hat{\delta}_{n,i-l}(g,\hat{\theta}_n)^\intercal, \quad l \geq 0. \end{cases}$$

The estimator for Σ^* is given by

$$\text{(3.16)}\quad \widehat{\Sigma}^*_n \equiv \big(H_n^\intercal \Xi_n H_n\big)^{-1} H_n^\intercal \Xi_n \widehat{\Sigma}_{g,n}(\hat{\theta}_n) \Xi_n H_n \big(H_n^\intercal \Xi_n H_n\big)^{-1}, \quad \text{where}$$
$$H_n \equiv \partial_\theta G_n(\hat{\theta}_n).$$

THEOREM 2: *Under Assumptions* 1–4, 6, *and* 7, *the following statements hold*:

(a) $\Delta_n^{-1/2}(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, \Sigma^*)$;

(b) $\widehat{\Sigma}^*_n \xrightarrow{\mathbb{P}} \Sigma^*$;

(c) *if* $\Sigma_g(\theta^*, \theta^*)$ *is nonsingular a.s. and* $\Xi = \Sigma_g(\theta^*, \theta^*)^{-1}$, *then* $\Delta_n^{-1} Q_n(\hat{\theta}_n) \xrightarrow{\mathcal{L}\text{-}s} \chi^2_{q\text{-}\dim(\theta)}$.

COMMENTS: (i) Part (a) shows the asymptotic mixed normality of the estimator $\hat{\theta}_n$ at the $n^{1/2}$-rate. Part (b) shows that the asymptotic $\mathcal{F}$-conditional covariance matrix Σ^* can be consistently estimated by $\widehat{\Sigma}^*_n$. An intermediate step of the proof is to show $\widehat{\Sigma}_{g,n}(\tilde{\theta}_n) \xrightarrow{\mathbb{P}} \Sigma_g(\theta^*, \theta^*)$ for any estimator $\tilde{\theta}_n$ of θ^* such that $\tilde{\theta}_n = \theta^* + o_p(B_n^{-1})$. Since $B_n = o(\Delta_n^{-\varsigma\kappa/2}) = o(\Delta_n^{-1/4})$ under maintained assumptions, any $n^{1/4}$-consistent estimator of θ^* is a valid choice for $\tilde{\theta}_n$. Part (c) shows that Hansen's (1982) J-statistic for the overidentification test has the familiar chi-squared asymptotic distribution in the current setting as well. This test can be implemented by setting $\Xi_n \equiv \widehat{\Sigma}_{g,n}(\tilde{\theta}_n)^{-1}$.

(ii) The HAC estimator $\widehat{\Gamma}_{g,n}(\hat{\theta}_n)$ is valid for weakly dependent error terms $(\chi_i)_{i\geq 0}$. If it is known a priori that $(\chi_i)_{i\geq 0}$ are mutually independent, we can instead define $\widehat{\Gamma}_{g,n}(\hat{\theta}_n)$ simply as $\widehat{\Gamma}_{g,0,n}(\hat{\theta}_n)$.

Theorem 2, parts (a) and (b), justify constructing confidence sets for θ^* in the usual way. We further note that Anderson–Rubin type confidence sets for θ^* can also be constructed by inverting tests. This type of confidence sets is known to be robust to weak identification issues (Stock and Wright (2000), Andrews and Soares (2010)).[15] For concreteness, we sketch the procedure using the test statistic $Q_n^{\mathrm{CU}}(\theta) \equiv \Delta_n^{-1} G_n(\theta)^\intercal \widehat{\Sigma}_{g,n}^{-1}(\theta) G_n(\theta)$, where CU stands for continuous-updating (Stock and Wright (2000)). By Theorem 1(b) and $\widehat{\Sigma}_{g,n}(\theta^*) \xrightarrow{\mathbb{P}} \Sigma_g(\theta^*, \theta^*)$, $Q_n^{\mathrm{CU}}(\theta^*)$ converges stably in law to a chi-squared distribution with degree of freedom q. For $\alpha \in (0, 1)$, let $cv_{1-\alpha}$ denote the $(1-\alpha)$-quantile of the limiting chi-squared distribution. It follows that the sequence of confidence sets $\mathrm{CS}_n \equiv \{\theta \in \Theta : Q_n^{\mathrm{CU}}(\theta) \leq cv_{1-\alpha}\}$ has asymptotic level $1-\alpha$, that is, $\mathbb{P}(\theta^* \in \mathrm{CS}_n) \to 1-\alpha$. Other test statistics can also be used (see Andrews and Soares (2010) for many examples), but they generally have nonpivotal null asymptotic distributions. In such cases, the critical values can be obtained via simulation.

3.3. *A Consistent Specification Test*

In this subsection, we provide a consistent specification test for the restriction (2.2). Since correct specification is no longer imposed, we now introduce the pseudo-true parameter $\theta^\dagger$ through Assumption 8 below, which generalizes Assumption 6.

ASSUMPTION 8: (i) *For some Θ-valued $\mathcal{F}$-measurable random variable $\theta^\dagger$, the function $Q(\theta) \equiv G(\theta)^\intercal \Xi G(\theta)$ is uniquely minimized at $\theta^\dagger$ a.s.*; (ii) $\Xi_n \xrightarrow{\mathbb{P}} \Xi$, *where Ξ is an $\mathcal{F}$-measurable* (*random*) *matrix that is positive semidefinite a.s.*; (iii) *under correct specification, $\theta^\dagger = \theta^*$ is in the interior of the compact set Θ*; (iv) *the random matrix $H(\theta^\dagger)^\intercal \Xi H(\theta^\dagger)$ is nonsingular a.s.*

Clearly, $\theta^\dagger$ coincides with the true parameter θ^* under correct specification. In the case of misspecification, $\theta^\dagger$ in general depends on the realization of the paths of studied processes and the choice of the weight function $\varphi(\cdot)$. It can be shown that $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta^\dagger$ by using Theorem 1(a).

Formally, the specification testing problem is to decide in which of the following two sets the observed sample path falls:[16]

$$\Omega_{H_0} \equiv \left\{\bar{\psi}\left(\beta_t, Z_t, V_t; \theta^\dagger\right) = 0 \text{ for all } t \in [0, T]\right\}, \quad \Omega_{H_a} \equiv \Omega \setminus \Omega_{H_0}. \tag{3.17}$$

[15] Andrews and Soares (2010) showed that this type of confidence sets has a valid size uniformly over a large class of data generating processes under high-level conditions in the standard long-span setting. The formal study of uniformity demands local asymptotics in the current nonstandard setting, which is beyond the scope of this paper.

[16] It is now standard in the high-frequency setting to form hypotheses in terms of collections of sample paths; see Aït-Sahalia and Jacod (2014) for many examples.

We construct a Bierens-type (Bierens (1982)) consistent specification test by testing whether a continuum of integrated moment conditions holds.[17] To this end, we consider a family of $\mathbb{R}$-valued weight functions on $\mathbb{R}$ of the form $t \mapsto \phi(\tau t)$, which is indexed by $\tau \in \mathcal{T}$ and the function $\phi(\cdot)$ satisfies the following assumption.

ASSUMPTION 9: *$\phi : \mathbb{R} \mapsto \mathbb{R}$ is a power series such that the set $\{k \in \mathbb{N} : (d/du)^k \phi(u)|_{u=0} = 0\}$ is finite and $\mathcal{T}$ is a compact subset of $\mathbb{R}$ with nonzero Lebesgue measure.*

We set

$$M(\theta, \tau) = \int_0^T \bar{\psi}(\beta_s, Z_s, V_s; \theta)\phi(\tau s)\,ds, \quad \tau \in \mathcal{T}.$$

Under Assumption 9, the set Ω_{H_0} can be equivalently written as $\Omega_{H_0} = \{M(\theta^\dagger, \tau) = 0 \text{ for all } \tau \in \mathcal{T}\}$. Below, we carry out the specification test by testing whether the process $M(\theta^\dagger, \tau)$ is identically zero over $\tau \in \mathcal{T}$.

We consider a functional estimator for $(M(\theta^\dagger, \tau))_{\tau\in\mathcal{T}}$ given by

$$M_n(\hat{\theta}_n, \tau) \equiv \Delta_n \sum_{i=0}^{N_n} \biggl(\psi(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \hat{\theta}_n, \tau) - \frac{1}{k_n} \partial_v^2 \psi(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \hat{\theta}_n, \tau)\widehat{V}_{i\Delta_n}^2 \biggr) \phi(i\Delta_n \tau).$$

We note that $M_n(\hat{\theta}_n, \tau)$ is a bias-corrected sample moment function like (3.3), except that it also depends on the estimator $\hat{\theta}_n$.

The asymptotic property of the process $M_n(\hat{\theta}_n, \tau)$ can be derived by using Theorem 1, after accounting for the sampling error in $\hat{\theta}_n$. In the proof of Theorem 3 below, we show that $M_n(\hat{\theta}_n, \tau) \xrightarrow{\mathbb{P}} M(\theta^\dagger, \tau)$ uniformly in τ, so that $M_n(\hat{\theta}_n, \tau)$ can be used to detect violations of the null hypothesis. Furthermore, in restriction to Ω_{H_0} (i.e., under the null hypothesis of correct specification), the sequence $\Delta_n^{-1/2} M_n(\hat{\theta}_n, \tau)$ of τ-indexed processes converges stably in law to a centered mixed Gaussian process $\tilde{\zeta}(\cdot)$ with the $\mathcal{F}$-conditional covariance function

$$C(\tau, \tau') \equiv \bigl[\mathbf{I}_{q_1} \vdots -D(\tau)\bigr] \Sigma_{\tilde{g}}(\tau, \tau') \bigl[\mathbf{I}_{q_1} \vdots -D(\tau')\bigr]^\top,$$

[17]In the standard statistical setting, an important alternative approach to consistent specification testing was studied by Härdle and Mammen (1993). The extension in this direction involves technical tools that are very distinct from ours and, hence, is left for future study.

for which we set

$$D(\tau) \equiv \left(\int_0^T \partial_\theta \bar{\psi}(\beta_s, Z_s, V_s; \theta^\dagger) \phi(\tau s) \, ds \right) \times \left(H(\theta^\dagger)^\intercal \Xi H(\theta^\dagger) \right)^{-1} H(\theta^\dagger)^\intercal \Xi$$

and define $\Sigma_{\tilde{g}}(\tau, \tau')$ as $\Sigma_g(\theta, \theta')$ in (3.11), but with $g(y, z, v; \theta)$ in the latter replaced by $\tilde{g}(y, z, v; \tau) \equiv h(y, z, v; \theta^\dagger) \odot \Phi(t\tau)$, where

$$h(y, z, v; \theta) \equiv \left(\psi(y, z, v; \theta), g(y, z, v; \theta) \right), \quad \Phi(t\tau) \equiv \left(\mathbf{J}_{q_1} \phi(t\tau), \mathbf{J}_q \right). \tag{3.18}$$

We estimate the $\mathcal{F}$-conditional covariance function $C(\tau, \tau')$ using

$$\widehat{C}_n(\tau, \tau') \equiv \left[\mathbf{I}_{q_1} \vdots - \widehat{D}_n(\tau) \right] \widehat{\Sigma}_{\tilde{g},n}(\hat{\theta}_n, \tau, \tau') \left[\mathbf{I}_{q_1} \vdots - \widehat{D}_n(\tau') \right]^\intercal,$$

where we set

$$\widehat{D}_n(\tau) \equiv \partial_\theta M_n(\hat{\theta}_n, \tau) \left(H_n^\intercal \Xi_n H_n \right)^{-1} H_n^\intercal \Xi_n$$

and, recalling (3.18), we set

$$\begin{cases} \widehat{\Sigma}_{\tilde{g},n}(\hat{\theta}_n, \tau, \tau') \equiv \widehat{S}_{\tilde{g},n}(\hat{\theta}_n, \tau, \tau') + \widehat{\Gamma}_{\tilde{g},n}(\hat{\theta}_n, \tau, \tau'), \\ \widehat{S}_{\tilde{g},n}(\hat{\theta}_n, \tau, \tau') \equiv 2\Delta_n \sum_{i=0}^{N_n} \left(\hat{m}'_{n,i}(h, \hat{\theta}_n) \odot \Phi(i\Delta_n \tau) \right) \times \left(\hat{m}'_{n,i}(h, \hat{\theta}_n) \odot \Phi(i\Delta_n \tau') \right)^\intercal \widehat{V}_{i\Delta_n}^2, \\ \widehat{\Gamma}_{\tilde{g},n}(\hat{\theta}_n, \tau, \tau') \equiv \widehat{\Gamma}_{\tilde{g},0,n}(\hat{\theta}_n, \tau, \tau') + \sum_{l=1}^{B_n} w(l, B_n) \left(\widehat{\Gamma}_{\tilde{g},l,n}(\hat{\theta}_n, \tau, \tau') + \widehat{\Gamma}_{\tilde{g},l,n}(\hat{\theta}_n, \tau', \tau)^\intercal \right), \\ \widehat{\Gamma}_{\tilde{g},l,n}(\hat{\theta}_n, \tau, \tau') \equiv \Delta_n \sum_{i=l}^{N_n} \left(\hat{\delta}_{n,i}(h, \hat{\theta}_n) \odot \Phi(i\Delta_n \tau) \right) \times \left(\hat{\delta}_{n,i-l}(h, \hat{\theta}_n) \odot \Phi((i-l)\Delta_n \tau') \right)^\intercal. \end{cases} \tag{3.19}$$

Turning to the test, we consider a Kolmogorov-type test statistic of the form

$$\widehat{K}_n \equiv \sup_{\tau \in \mathcal{T}} \max_{1 \le j \le q_1} \frac{\Delta_n^{-1/2} \left| M_{j,n}(\hat{\theta}_n, \tau) \right|}{\sqrt{\widehat{C}_{jj,n}(\tau, \tau)}}.$$

At significance level $\alpha \in (0, 1)$, we reject the null hypothesis of correct specification when $\widehat{K}_n$ is greater than a critical value cv_n^α that consistently estimates the $\mathcal{F}$-conditional $(1 - \alpha)$-quantile of the asymptotic null distribution

1 Simulate a centered Gaussian process $(\tilde{\zeta}_n^{\text{Sim}}(\tau))_{\tau\in\mathcal{T}}$ with covariance function $\widehat{C}_n(\cdot,\cdot)$.
2 Compute

$$\widehat{K}_n^{\text{Sim}} = \sup_{\tau\in\mathcal{T}} \max_{1\le j\le q_1} \frac{\left|\tilde{\zeta}_{j,n}^{\text{Sim}}(\tau)\right|}{\sqrt{\widehat{C}_{jj,n}(\tau,\tau)}}.$$

3 Repeat step 1 and step 2 to generate a large Monte Carlo sample and set cv_n^α to be the $(1-\alpha)$-quantile of $\widehat{K}_n^{\text{Sim}}$ in this simulated sample.

Algorithm 1: Critical value of the consistent specification test.

of $\widehat{K}_n$. Since the asymptotic null distribution is nonstandard, cv_n^α does not have a closed-form expression, but it can be constructed via simulation as detailed in Algorithm 1. Theorem 3, below, summarizes the asymptotic properties of the test.

THEOREM 3: *Let* $\alpha\in(0,1/2)$ *be a constant. Suppose* (i) *Assumptions* 1–4 *and* 7–9; (ii) *the function* ψ *satisfies Assumption* 2; (iii) $\inf_{\tau\in\mathcal{T}}\lambda_{\min}(C(\tau,\tau))>0$ *a.s.*; (iv) $\hat{\theta}_n-\theta^\dagger=o_p(B_n^{-1})$. *Then the test associated with the critical region* $\{\widehat{K}_n>cv_n^\alpha\}$ *has asymptotic size* α *under the null hypothesis and is consistent under the alternative hypothesis, that is,*

$$\mathbb{P}\big(\widehat{K}_n>cv_n^\alpha|\Omega_{H_0}\big)\to\alpha,\quad \mathbb{P}\big(\widehat{K}_n>cv_n^\alpha|\Omega_{H_a}\big)\to 1.$$

COMMENT: Given the results developed in Sections 3.1 and 3.2, the key additional technical component underlying Theorem 3 is the analysis of the asymptotic behavior of the test, including that of the estimator $\widehat{C}_n(\cdot,\cdot)$, for possibly misspecified models. This analysis is done under the (mild) convergence-rate condition that $\hat{\theta}_n-\theta^\dagger=o_p(B_n^{-1})$, for which $n^{1/4}$-consistency of $\hat{\theta}_n$ toward $\theta^\dagger$ suffices.

REFERENCES

ACKERBERG, D., X. CHEN, J. HAHN, AND Z. LIAO (2014): "Asymptotic Efficiency of Semiparametric Two-Step GMM," *Review of Economic Studies*, 81, 919–943. [1621]

AÏT-SAHALIA, Y., AND J. JACOD (2014): *High-Frequency Financial Econometrics*. Princeton: Princeton University Press. [1628]

ANDERSEN, T. (1996): "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance*, 51, 169–204. [1619]

ANDERSEN, T. G., T. BOLLERSLEV, AND S. LANGE (1999): "Forecasting Financial Market Volatility: Sample Frequency vis-á-vis Forecast Horizon," *Journal of Empirical Finance*, 6 (5), 457–477. [1614]

ANDREWS, D. W. K. (2005): "Cross-Section Regression With Common Shocks," *Econometrica*, 73 (5), 1551–1585. [1618]

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78 (1), 119–157. [1615, 1628]

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 1481–1536. [1617]

BICKEL, P. J., C. A. J. KLAASSEN, Y. RITOV, AND J. A. WELLNER (1998): *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer-Verlag. [1621]

BIERENS, H. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134. [1615,1621,1629]

BLACK, F. (1976): "Studies of Stock Price Volatility Changes," in *Proceedings of the 1976 Meetings of the American Statistical Association*. Washington, DC: American Statistical Association, 171–181. [1617]

BOLLERSLEV, T., R. ENGLE, AND D. NELSON (1994): "ARCH Models," in *Handbook of Econometrics*, Vol. 4. Amsterdam: Elsevier. [1614]

CARRASCO, M., AND J.-P. FLORENS (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16 (6), 797–834. [1621]

——— (2014): "On the Asymptotic Efficiency of GMM," *Econometric Theory*, 30, 372–406. [1621]

CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, Vol. 6B. Amsterdam: Elsevier. [1621]

CATTANEO, M. D., R. K. CRUMP, AND M. JANSSON (2013): "Generalized Jackknife Estimators of Weighted Average Derivatives," *Journal of the American Statistical Association*, 108, 1243–1256. [1614]

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, Vol. 6. Amsterdam: Elsevier. [1614]

CLÉMENT, E., S. DELATTRE, AND A. GLOTER (2013): "An Infinite Dimensional Convolution Theorem With Applications to the Efficient Estimation of the Integrated Volatility," *Stochastic Processes and Their Applications*, 123, 2500–2521. [1621]

COMTE, F., AND E. RENAULT (1996): "Long Memory Continuous Time Models," *Journal of Econometrics*, 73, 101–149. [1617]

——— (1998): "Long Memory in Continuous-Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291–323. [1614,1617,1622]

ENGLE, R. F. (2004): "Risk and Volatility: Econometric Models and Financial Practice," *American Economic Review*, 94 (3), 405–420. [1613]

FOSTER, D., AND D. B. NELSON (1996): "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica*, 64, 139–174. [1614,1622]

GAGLIARDINI, P., C. GOURIÉROUX, AND E. RENAULT (2011): "Efficient Derivative Pricing by the Extended Method of Moments," *Econometrica*, 79 (4), 1181–1232. [1618]

GARCIA, R., E. GHYSELS, AND E. RENAULT (2010): "The Econometrics of Option Pricing," in *Handbook of Financial Econometrics*, ed. by Y. Aït-Sahalia and L. P. Hansen. Amsterdam: North-Holland, 479–552. [1619]

GHYSELS, E., A. HARVEY, AND E. RENAULT (1996): "Stochastic Volatility," in *Handbook of Statistics*, Vol. 14, ed. by G. Maddala and C. Rao. Amsterdam: North-Holland, Chapter 5, 119–191. [1614]

HANSEN, B. E. (1996): "Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays," *Econometric Theory*, 12 (2), 347–359. [1624]

——— (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [1613,1627]

HANSEN, L. P., AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50 (5), 1269–1286. [1618]

HÄRDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, 21, 1926–1947. [1629]

JACOD, J., AND P. PROTTER (2012): *Discretization of Processes*. Heidelberg: Springer. [1617,1622]

JACOD, J., AND M. ROSENBAUM (2013): "Quarticity and Other Functionals of Volatility: Efficient Estimation," *The Annals of Statistics*, 41, 1462–1484. [1615,1621,1624]

JACOD, J., AND A. N. SHIRYAEV (2003): *Limit Theorems for Stochastic Processes* (Second Ed.). New York: Springer-Verlag. [1624]

JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Their Applications*, 119, 2249–2276. [1618]

KRISTENSEN, D. (2010): "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach," *Econometric Theory*, 26 (1), 60–93. [1622]

LI, J., AND D. XIU (2016): "Supplement to 'Generalized Method of Integrated Moments for High-Frequency Data'," *Econometrica Supplemental Material*, 84, http://dx.doi.org/10.3982/ECTA12306. [1615]

MANCINI, C. (2001): "Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion," *Giornale Dell'Istituto Italiano Degli Attuari*, LXIV, 19–47. [1622]

NEWEY, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62 (6), 1349–1382. [1614]

NEWEY, W. K., AND D. L. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4. Amsterdam: Elsevier. [1614]

PEARSON, K. (1894): "Contribution to the Mathematical Theory of Evolution," *Philosophical Transactions of the Royal Society of London, Series A*, 185, 71–110. [1613]

RENAULT, E., C. SARISOY, AND B. J. WERKER (2014): "Efficient Estimation of Integrated Volatility and Related Processes," Discussion Paper, Brown University. [1621]

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM With Weak Identification," *Econometrica*, 68, 1055–1096. [1615,1628]

XIU, D. (2010): "Quasi-Maximum Likelihood Estimation of Volatility With High Frequency Data," *Journal of Econometrics*, 159, 235–250. [1618]

ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411. [1617]

Duke University, Box 90097, Durham, NC 27708, U.S.A.; jl410@duke.edu
and
University of Chicago Booth School of Business, 5807 S. Woodlawn Avenue, Chicago, IL 60637, U.S.A.; dacheng.xiu@chicagobooth.edu.

Co-editor Elie Tamer handled this manuscript.

Manuscript received February, 2014; final revision received February, 2016.