

# Generalized Methods of Integrated Moments for High-Frequency Data\*

Jia Li<sup>†</sup>                      Dacheng Xiu<sup>‡</sup>  
Duke University              Chicago Booth

This Version: February 14, 2014

## Abstract

We study the asymptotic inference for a conditional moment equality model using high-frequency data sampled within a fixed time span. The model involves the latent spot variance of an asset as a covariate. We propose a two-step semiparametric inference procedure by first nonparametrically recovering the volatility path from asset returns and then conducting inference by matching integrated moment conditions. We show that, due to the first-step estimation error, a bias-correction is needed for the sample moment condition to achieve asymptotic (mixed) normality. We provide feasible inference procedures for the model parameter and establish their asymptotic validity. Empirical applications on VIX pricing and the volatility-volume relationship are provided to illustrate the use of the proposed method.

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

JEL CODES: C22.

---

\*This paper supersedes our working paper previously circulated under the title “Spot Variance Regressions,” with substantially more general results. We are grateful to Yacine Aït-Sahalia, Torben Andersen, Federico Bandi, Alan Bester, Tim Bollerslev, Federico Bugni, Chris Hansen, Oliver Linton, Per Mykland, Andrew Patton, Eric Renault, Jeff Russell, Elie Tamer, George Tauchen, Viktor Todorov, Lan Zhang, as well as many seminar and conference participants at the University of Chicago, Brown University, the 2012 Triangle Econometrics Conference, the 2013 Financial Econometrics Conference at Toulouse School of Economics, the 6th Annual SoFiE Conference, and the 2013 workshop on “Measuring and Modeling Financial Risk with High Frequency Data” at EUI 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.

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

<sup>‡</sup>University of Chicago Booth School of Business, 5807 S. Woodlawn Avenue, Chicago, IL 60637. Email: dacheng.xiu@chicagobooth.edu.

# 1 Introduction

Inference methods based on moment equalities have been a powerful tool in empirical economists' arsenal since the invention of the generalized method of moments (GMM) (Hansen (1982), Hansen and Singleton (1982)). In their application, moment conditions often arise from conditional moment equalities as orthogonality conditions between instruments and random disturbances. Asymptotic properties of these methods are determined by the properties of sample moments, which are well known (White (2001)) in the classical "large  $T$ " setting with an asymptotically expanding time span. In this paper, we study a novel variant of the GMM for estimating conditional moment equality models using high-frequency (intraday) data that are sampled within a relatively short sample period. We derive an asymptotic theory in a setting where data are sampled at asymptotically increasing frequencies within a fixed time span, allowing for general forms of dependence and heterogeneity in the data. Our study is mainly motivated by financial applications such as the estimation of certain types of option pricing models and market microstructure models, where high-frequency data are rapidly becoming more readily available.

An important aspect of financial models is that they often involve volatility processes of financial time series. This is not surprising since volatility is the primary measure of risk in modern finance (Engle (2004)). Since volatility is unobservable, its appearance in the model poses a substantial challenge for inference. The common solution to the latent volatility problem is to impose auxiliary parametric restrictions on volatility dynamics; see Bollerslev, Engle, and Nelson (1994), Ghysels, Harvey, and Renault (1995) and Shephard (2005) for reviews. Since an incorrect parametric specification of the auxiliary model may affect the inference of the primary model, it is prudent to consider a nonparametric approach as a complement.<sup>1</sup> Indeed, a large literature on nonparametric inference for volatility has emerged during the past decade by harnessing the rich information in high-frequency data; see Jacod and Protter (2012), Hautsch (2012) and Andersen, Bollerslev, Christoffersen, and Diebold (2013) for recent reviews.

This paper proposes a simple, yet general, two-step semiparametric procedure for estimating conditional moment equality models that include volatility as a latent variable. In the first step, we nonparametrically recover the volatility process from high-frequency asset returns via a spot realized variance estimator (Foster and Nelson (1996), Comte and Renault (1998)) with truncation for price jumps (Mancini (2001), Jacod and Protter (2012)). In the second step, we construct sample versions of instrumented conditional moment equalities. Unlike the classical GMM, the population moment condition here takes form of an integrated stochastic process that involves the

---

<sup>1</sup>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. Pseudo-true parameters (White (1982)) for misspecified parametric models may be worth considering in practice as well.

spot variance and other state variables over a fixed time span, instead of an unconditional moment. We thus refer to the proposed framework as the *generalized method of integrated moments* (GMIM). The GMIM estimator for a finite-dimensional model parameter is constructed as the minimizer of a sample criterion function of the quadratic form. Our analysis also extends the scope of the high-frequency literature on volatility estimation: while prior work focused on the inference of the volatility itself, we treat its estimation only as a preliminary step and mainly consider the subsequent inference of parameters in economic models.

Since we treat the volatility process in a nonparametric manner, our method is semiparametric in this particular aspect. The key distinctive feature of our semiparametric procedure is that the nonparametric object here (i.e., the volatility process) is a nonsmooth stochastic process rather than a smooth deterministic function. Indeed, the sample path of the volatility process in a typical stochastic volatility model (Heston (1993), Duffie, Pan, and Singleton (2000)) is nowhere differentiable because of Brownian volatility shocks and is often discontinuous due to volatility jumps. This feature gives rise to an interesting theoretical result: the first-step volatility estimation leads to a “large” bias in the sample moment function, in the sense that the bias cannot be made asymptotically negligible in the derivation of central limit theorems by just restricting the asymptotic behavior of tuning parameters. We hence consider an explicit bias-correction to the sample moment function and show that the bias-corrected sample moment function enjoys a central limit theorem. This result extends the theory of Jacod and Rosenbaum (2013) and is one of our main technical contributions. In contrast, in typical kernel- or sieve-based methods, the bias from the nonparametric estimation can be “tuned” to be asymptotically small by undersmoothing (or overfitting) the unknown function, under the assumption that the function is sufficiently smooth; see, for example, Newey (1994) and Gagliardini, Gouriéroux, and Renault (2011).

The GMIM estimator is constructed using the bias-corrected sample moment function. We show that the GMIM estimator is consistent and has a mixed Gaussian asymptotic distribution. The asymptotic covariance matrix is random and consists of two additive components. The first component is due to the random disturbances (e.g., pricing errors in an option pricing model) that implicitly define the conditional moment equalities. We allow the random disturbance to be serially weakly dependent and propose a heteroskedasticity and autocorrelation consistent (HAC) estimator for it. The HAC estimator is nonstandard (cf. Newey and West (1987)) due to its involvement with discretized processes including, in particular, the latent volatility process, in an in-fill asymptotic setting. The second component is contributed by the first-step estimation error, for which new consistent estimators are also provided in closed form. Overidentification tests (Hansen (1982)) and Anderson–Rubin–type confidence sets (Anderson and Rubin (1949), Stock and Wright (2000), Andrews and Soares (2010)) are also discussed as by-products.

We illustrate the proposed method with two empirical applications. The first application

concerns the pricing of the CBOE volatility index (VIX). We exploit a simple idea: a large (but far from exhaustive) class of structural models for the risk-neutral volatility dynamics with linear mean-reversion implies that the squared VIX is linear in the spot variance of the S&P 500 index. We test the specification of this class of models via the GMIM overidentification test and find that these models are rejected in 14 out of 23 quarters (2007Q1–2012Q3) at the 5% significance level. In the second application, we investigate the relationship between return variance and trading volume for stock data. Using daily data, Andersen (1996) found that a conditional Poisson model for trading volume is broadly consistent with data and outperforms early models considered by Tauchen and Pitts (1983) and Harris (1986). We estimate and conduct specification tests for these models using high-frequency data under the GMIM framework and find further support for the findings of Andersen (1996).

This paper is organized as follows. Section 2 presents the setting. Section 3 presents the main theory. Section 4 shows simulation results, followed by two empirical applications in Section 5. We discuss related literature in Section 6. Section 7 concludes. The appendix contains all proofs.

## 2 Generalized method of integrated moments

### 2.1 The setting

We observe a data sequence  $(X_t, Z_t, Y_t)$  at discrete times  $t = 0, \Delta_n, 2\Delta_n, \dots$  within a *fixed* time span  $[0, T]$ , with the sampling interval  $\Delta_n \rightarrow 0$  asymptotically. In applications,  $X_t$  typically denotes the (logarithmic) asset price,  $Z_t$  denotes observable state variables and  $Y_t$  denotes dependent variables such as prices of derivative contracts, trading volumes, etc. In this subsection, we formalize the probabilistic setting underlying our analysis, with concrete empirical examples given in Section 2.2.

Let  $(\Omega^{(0)}, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}^{(0)})$  be a filtered probability space. Without further mention, we assume that all processes defined on this space are càdlàg (i.e., right continuous with left limit) adapted and take values in some finite-dimensional real space. We endow this probability space with the processes  $X_t$ ,  $Z_t$  and  $\beta_t$  that, respectively, take values in  $\mathcal{X}$ ,  $\mathcal{Z}$  and  $\mathcal{B}$ . The process  $\beta_t$  is not observable; instead, we observe

$$Y_{i\Delta_n} = \mathcal{Y}(\beta_{i\Delta_n}, \chi_i), \quad i = 0, \dots, [T/\Delta_n], \quad (2.1)$$

where  $\chi_i$  is a random disturbance,  $\mathcal{Y}(\cdot)$  is a deterministic transform taking values in a finite-dimensional real space  $\mathcal{Y}$  and  $[T/\Delta_n]$  is the integer part of  $T/\Delta_n$ .

We shall assume the random disturbances  $(\chi_i)_{i \geq 0}$  to be  $\mathcal{F}$ -conditionally stationary and weakly dependent. To be precise, we describe the formal setting as follows. We consider another probabil-

ity 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  $\chi_i$  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  $(\chi_i)_{i \geq 0}$  to be serially independent. Let  $\Omega = \Omega^{(0)} \times \Omega^{(1)}$  and  $\mathbb{P} = \mathbb{P}^{(0)} \otimes \mathbb{P}^{(1)}$ . Processes defined on 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  $\sigma$ -fields  $\mathcal{F}$  and  $\mathcal{F}_t$  with their trivial extensions  $\mathcal{F} \otimes \{\emptyset, \Omega^{(1)}\}$  and  $\mathcal{F}_t \otimes \{\emptyset, \Omega^{(1)}\}$  on the product space. By construction, the sequence  $(\chi_i)_{i \in \mathbb{Z}}$  is independent of  $\mathcal{F}$ .

We note that the variable  $Y_{i\Delta_n}$  is a noisy transform of  $\beta_{i\Delta_n}$  with  $\chi_i$  being the confounding random disturbance. In its simplest form, (2.1) may have a signal-plus-noise appearance:  $Y_{i\Delta_n} = \beta_{i\Delta_n} + \chi_i$ . That noted, (2.1) often takes more complicated forms in many applications, as illustrated by the examples in Section 2.2. Heuristically, the formulation (2.1) highlights two distinct model components for the sequence  $(Y_{i\Delta_n})_{i \geq 0}$ : information “inside” the information set  $\mathcal{F}$  (e.g.,  $\mathcal{F}$ -conditional temporal heterogeneity) is captured by the process  $\beta_t$  and information “outside”  $\mathcal{F}$  is captured by  $(\chi_i)_{i \geq 0}$ .<sup>2</sup>

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

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

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

where the process  $b_t$  is locally bounded; the process  $\sigma_t$  is strictly positive;  $W_t$  is a standard Brownian motion;  $\delta : \Omega \times \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}$  is a predictable function and  $\mu$  is a Poisson random measure with compensator  $\nu$  of the form  $\nu(dt, dz) = dt \otimes \lambda(dz)$  for some  $\sigma$ -finite measure  $\lambda$  on  $\mathbb{R}$ . Moreover, for some constant  $r \in (0, 1)$ , a sequence of stopping times  $(T_m)_{m \geq 1}$  and  $\lambda$ -integrable deterministic functions  $(J_m)_{m \geq 1}$ , we have  $|\delta(\omega^{(0)}, t, z)|^r \wedge 1 \leq J_m(z)$  for all  $\omega^{(0)} \in \Omega^{(0)}$ ,  $t \leq T_m$  and  $z \in \mathbb{R}$ .

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

$$\begin{aligned} \tilde{Z}_t &= \tilde{Z}_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s d\tilde{W}_s \\ &\quad + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s, z) 1_{\{\|\tilde{\delta}(s, z)\| \leq 1\}} (\mu - \nu)(ds, dz) \\ &\quad + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s, z) 1_{\{\|\tilde{\delta}(s, z)\| > 1\}} \mu(ds, dz), \end{aligned}$$

---

<sup>2</sup>This formal setting for introducing weakly dependent random disturbances into high-frequency data has been considered by, for example, Jacod, Li, and Zheng (2013), who consider  $\mathcal{B}(\cdot)$  with a location-scale form.

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

The key condition in Assumption H is that the process  $X_t$  is an Itô semimartingale. In applications,  $X_t$  is typically the (logarithmic) price of an asset and  $\sigma_t$  is its stochastic volatility process. We set  $V_t \equiv \sigma_t^2$  and refer to it as the *spot variance process*; it takes values in  $\mathcal{V} \equiv (0, \infty)$ .

Assumption H accommodates many models in finance and is commonly used for deriving in-fill asymptotic results for high-frequency data; see, for example, Jacod and Protter (2012) and the references therein. There is no stationarity requirement on the processes  $X_t$ ,  $\beta_t$ ,  $Z_t$  and  $\sigma_t$ . Although the sequence  $\chi_i$  is stationary, the sequence  $Y_{i\Delta_n}$  is allowed to be highly nonstationary through its dependence on  $\beta_{i\Delta_n}$ . Assumption H also 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  $r$  in Assumption H(i) serves as an upper bound for the generalized Blumenthal–Gettoor index, or the “activity,” of jumps. Assumption H(ii) also restricts the processes  $\beta_t$ ,  $Z_t$  and  $\sigma_t$  to be Itô semimartingales. We note that this assumption accommodates stochastic volatility models with multiple factors (see, e.g., Chernov, Gallant, Ghysels, and Tauchen (2003)), provided that each factor is an Itô semimartingale. This assumption also allows general forms for volatility-of-volatility and volatility jumps, where the latter may have infinite activity and even infinite variation. While Assumption H(ii) 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 a focused research on its own and is left to future study.

## 2.2 The conditional moment equality model and examples

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

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

where  $\psi : \mathcal{Y} \times \mathcal{Z} \times \mathcal{V} \mapsto \mathbb{R}^{q_1}$ ,  $q_1 \geq 1$ , is a measurable function with a known functional form up to the unknown parameter  $\theta^*$ , and the conditional expectation integrates out the random disturbance  $\chi_i$ . We suppose that the true parameter  $\theta^*$  is deterministic and takes value in a compact parameter space  $\Theta \subset \mathbb{R}^{\dim(\theta^*)}$ . In the sequel, we use  $\theta$  to denote a generic element in  $\Theta$ . The transpose of a matrix  $A$  is denoted by  $A^\top$ .

To motivate model (2.2), we consider a few empirical examples.

EXAMPLE 1 (LINEAR REGRESSION MODEL): Let  $X_t$  denote the logarithm of the S&P 500 index and let  $VIX_t$  denote the CBOE volatility index. We set  $Y_t \equiv VIX_t^2$ . For a large (but far from exhaustive) class of risk-neutral dynamics for the spot variance process  $V_t$ , the theoretical value of the squared VIX has a linear form  $\theta_1^* + \theta_2^* V_t$ ; see Section 5.1 for details. Empirically, we can model the observed  $Y_{i\Delta_n}$  as the theoretical price plus a pricing error  $a_{i\Delta_n} \chi_i$ , that is,

$$Y_{i\Delta_n} = \theta_1^* + \theta_2^* V_{i\Delta_n} + a_{i\Delta_n} \chi_i, \quad \mathbb{E}[\chi_i | \mathcal{F}] = 0, \quad \mathbb{E}[\chi_i^2 | \mathcal{F}] = 1, \quad (2.3)$$

where we allow the scaling factor  $a_t$  of the pricing error to be stochastic with the condition  $\mathbb{E}[\chi_i^2 | \mathcal{F}] = 1$  being a normalization. Note that (2.3) can be written in the form of (2.1) with  $\beta_t \equiv (\theta_1^* + \theta_2^* V_t, a_t)$ , where  $\mathcal{Y}(\cdot)$  takes a location-scale form. The pricing error  $a_{i\Delta_n} \chi_i$  is introduced to capture price components that standard risk-neutral pricing models do not intend to capture. The pricing errors can be serially dependent as we allow the process  $a_t$  and the sequence  $(\chi_i)_{i \geq 0}$  both to be serially dependent in a nonparametric manner; allowing for general statistical structure on the pricing errors is important, as emphasized by Bates (2000). By setting  $\psi(Y_t, V_t; \theta) = Y_t - \theta_1 - \theta_2 V_t$ , we verify (2.2).

EXAMPLE 2 (NONLINEAR REGRESSION MODEL): Let  $X_t$  be the price process of an underlying asset and  $Y_t$  be the price vector of  $q_1$  options 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,<sup>3</sup> then the theoretical prices of the collection of  $q_1$  options can be written as a  $\mathbb{R}^{q_1}$ -valued function  $f(Z_t, V_t; \theta^*)$ , where  $\theta^*$  arises from the risk-neutral model for the dynamics of the state variables. Empirically, it is common to model the observed option price vector  $Y_t$  as the theoretical price plus a pricing error, that is,

$$Y_{i\Delta_n} = f(Z_{i\Delta_n}, V_{i\Delta_n}; \theta^*) + a_{i\Delta_n} \chi_i, \quad \mathbb{E}[\chi_i | \mathcal{F}] = 0, \quad \mathbb{E}[\chi_i \chi_i^\top | \mathcal{F}] = I_{q_1}, \quad (2.4)$$

where  $a_t$  is a  $q_1 \times q_1$  matrix-valued process that denotes the stochastic covolatility of the pricing errors with the condition  $\mathbb{E}[\chi_i \chi_i^\top | \mathcal{F}] = I_{q_1}$  being a normalization. Note that (2.4) can be written in the form of (2.1) with  $\beta_t \equiv (\beta_{1,t}^\top, \beta_{2,t}^\top)^\top$ ,  $\beta_{1,t} \equiv f(Z_t, V_t; \theta^*)$  and  $\beta_{2,t} \equiv \text{vec}(a_t)$ , where  $\text{vec}(\cdot)$  denotes the vectorization operator. Setting  $\psi(Y_t, Z_t, V_t; \theta) = Y_t - f(Z_t, V_t; \theta)$ , we verify (2.2).

---

<sup>3</sup>Assuming that  $V_t$  is the only unobservable Markov state variable excludes derivative pricing models with multiple volatility factors under the risk-neutral measure, which have been considered by, for example, Christoffersen, Heston, and Jacobs (2009), Bates (2012) and Andersen, Fusari, and Todorov (2013). Note that this assumption does not imply  $(Z_t, V_t)$  is Markov under the physical measure (i.e.  $\mathbb{P}$ ), as 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 H) under the physical measure even if one imposes additional restrictions under the risk-neutral measure.

EXAMPLE 3 (PARAMETRIZED CONDITIONAL HETEROSKEDASTICITY): Consider the same setting as Example 2. The process  $A_t = \text{vec}(a_t a_t^\top)$  is an economically relevant quantity as it can be interpreted as a summary measure of market quality (Hasbrouck (1993), Ait-Sahalia and Yu (2009)). To investigate whether  $A_t$  depends on other state variables, one may further model  $A_t$  as  $A_t = h(Z_t, V_t; \theta^*)$  for some deterministic function  $h(\cdot)$ .<sup>4</sup> Then we can verify (2.2) by setting

$$\psi(Y_t, Z_t, V_t; \theta) = \begin{pmatrix} Y_t - f(Z_t, V_t; \theta) \\ \text{vec}((Y_t - f(Z_t, V_t; \theta))(Y_t - f(Z_t, V_t; \theta))^\top) - h(Z_t, V_t; \theta) \end{pmatrix}.$$

EXAMPLE 4 (SCALED POISSON REGRESSION MODEL): Andersen (1996) proposes a Poisson model for the volatility–volume 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 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  $\beta$  and then set  $\beta_t \equiv \theta_2^* + \theta_3^* V_t$  and  $Y_{i\Delta_n} = \theta_1^* \chi_i(\beta_{i\Delta_n})$ . In Section 5.2, we estimate this model by using the first two conditional moments of  $Y_t$ . This amounts to setting

$$\psi(Y_t, V_t; \theta) = \begin{pmatrix} Y_t - \theta_1(\theta_2 + \theta_3 V_t) \\ Y_t^2 - \theta_1^2(\theta_2 + \theta_3 V_t)^2 - \theta_1^2(\theta_2 + \theta_3 V_t) \end{pmatrix}, \quad (2.5)$$

which readily verifies (2.2).

As shown in the above examples, the conditional moment equality model (2.2) arises in a variety of empirical settings. These settings naturally involve the spot variance process  $V_t$ , but are agnostic regarding the precise form of its dynamics (under the physical measure). 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, while it may be reasonable to assume that the stochastic volatility process is stationary in the classical large- $T$  setting for daily or weakly data, the stationarity assumption is more restrictive for high-frequency data due to intradaily seasonalities.

Finally, we note that while  $X_{i\Delta_n}$  is assumed to be observed without microstructure noise, we do allow  $Y_{i\Delta_n}$  to be noisy in a quite general fashion. In particular, in option pricing settings such as Examples 1–3,  $Y_{i\Delta_n}$  has the form of a semimartingale plus a noise (i.e., pricing error) term,

---

<sup>4</sup>Upon a reparametrization, we can assume that  $f(\cdot)$  and  $h(\cdot)$  share the same parameter without loss of generality.



which is commonly used in the study of noise-robust estimations of integrated volatility.<sup>5</sup> Our “asymmetric” treatment for microstructure noise in  $X_{i\Delta_n}$  and  $Y_{i\Delta_n}$  is reasonably realistic as the option market is less liquid than the stock market, so microstructure effects play a less important role for the latter than the former.<sup>6</sup>

### 2.3 Integrated moment equalities and the GMIM estimator

Our inference is based on matching a set of integrated moment equalities that are implied by (2.2). To construct these integrated moment conditions, we consider a measurable function  $\varphi : \mathcal{Z} \times \mathcal{V} \times \Theta \mapsto \mathbb{R}^{q_2}$  for some  $q_2 \geq 1$ . Below, we refer to  $\varphi(\cdot)$  as the instrument. We set  $q = q_1 q_2$  and consider a  $\mathbb{R}^q$ -valued function

$$g(y, z, v; \theta) \equiv \psi(y, z, v; \theta) \otimes \varphi(z, v; \theta), \quad (2.6)$$

with which we associate

$$\bar{g}(\beta, z, v; \theta) \equiv \int g(\mathcal{Y}(\beta, \chi), z, v; \theta) \mathbb{P}_\chi(d\chi). \quad (2.7)$$

Since  $Z_{i\Delta_n}$  and  $V_{i\Delta_n}$  are  $\mathcal{F}$ -measurable, (2.2) implies that  $\mathbb{E}[g(Y_{i\Delta_n}, Z_{i\Delta_n}, V_{i\Delta_n}; \theta^*) | \mathcal{F}] = 0$  or, equivalently,

$$\bar{g}(\beta_{i\Delta_n}, Z_{i\Delta_n}, V_{i\Delta_n}; \theta^*) = 0, \quad i = 0, \dots, [T/\Delta_n]. \quad (2.8)$$

If  $\bar{g}(\beta, z, v; \theta)$  is continuous in  $(\beta, z, v)$ , then the process  $(\bar{g}(\beta_t, Z_t, V_t; \theta))_{t \geq 0}$  is càdlàg, so we can define

$$G(\theta) \equiv \int_0^T \bar{g}(\beta_s, Z_s, V_s; \theta) ds, \quad \theta \in \Theta. \quad (2.9)$$

By (2.8) and a Riemann approximation, we obtain a vector of integrated moment equalities given by

$$G(\theta^*) = 0. \quad (2.10)$$

In Section 3.2, we construct an estimator  $G_n(\cdot)$  for the random function  $G(\cdot)$  and show that  $G_n(\cdot)$  converges in probability toward  $G(\cdot)$  uniformly. Following Sargan (1958) and Hansen (1982),

<sup>5</sup>See, for example, Zhang, Mykland, and Ait-Sahalia (2005), Hansen and Lunde (2006), Bandi and Russell (2008), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), Jacod, Li, Mykland, Podolskij, and Vetter (2009) and Xiu (2010).

<sup>6</sup>We note that our analysis is based on general integrated volatility functionals, for which little is known in noisy settings in the current literature. To the best of our knowledge, the most general class of estimators is the pre-averaging method of Jacod, Podolskij, and Vetter (2010), which can be used to estimate integrated volatility functionals of the form  $\int_0^T V_s^j ds$  for positive integer  $j$ . This class of integrated volatility polynomials, however, is quite restrictive for our purpose of estimating general nonlinear models. Since estimating general integrated volatility functionals in the noisy setting is a very challenging task by itself, we leave the extension with noisy  $X$  to future research, so as to focus on the main idea of the current paper.

we estimate  $\theta^*$  by making  $G_n(\theta)$  as close to zero as possible according to some metric. More precisely, we consider a sequence  $\Xi_n$  of weighting matrices and define the GMIM estimator  $\hat{\theta}_n$  as

$$\hat{\theta}_n \equiv \underset{\theta \in \Theta}{\operatorname{argmin}} Q_n(\theta), \quad \text{where} \quad Q_n(\theta) \equiv G_n(\theta)^\top \Xi_n G_n(\theta). \quad (2.11)$$

The GMIM estimator clearly resembles the classical GMM estimator. Moreover, transforming the conditional moment equality (2.2) into the integrated moment equality (2.10) is analogous to the common practice of estimating conditional moment equality models by forming unconditional moment conditions.

That being said, there are fundamental differences between the two settings. The classical GMM setting requires a large sample with an expanding time span in order to recover the invariant distribution of the studied processes. In the in-fill setting here, we do not require the existence of an invariant distribution. In the continuous-time limit, the integrated moment function  $G(\cdot)$ , rather than being an unconditional moment, arises naturally as the limiting, or “population,” version of the sample moment condition. The phenomenon that stochastic limits take the form of temporally integrated quantities is common in the econometrics for high-frequency data; see Andersen, Bollerslev, Diebold, and Labys (2003), Barndorff-Nielsen and Shephard (2004a), Jacod and Protter (2012) and references therein. As is typical in the high-frequency literature, our in-fill asymptotic results require only mild conditions on the sample-path regularity of the processes  $\beta_t$ ,  $Z_t$ ,  $X_t$  and  $V_t$  (see Assumption H), while allowing for general forms of nonstationarity and dependence; the current setting is actually non-ergodic, as the integrated moment function  $G(\cdot)$  is itself a random function.

### 3 Asymptotic theory

In Section 3.1, we discuss regularity conditions. In Sections 3.2 and 3.3, we present the key theoretical results of the current paper, that is, the asymptotic properties of the bias-corrected sample moment function (Section 3.2) and consistent estimators of its asymptotic covariance matrix (Section 3.3). Asymptotic results for the GMIM estimator then follow straightforwardly and are presented in Section 3.4.

#### 3.1 Assumptions

In this subsection, we collect and discuss some regularity conditions that are used repeatedly in the sequel. This subsection is technical in nature and may be skipped by readers interested in our main results during their first reading.

ASSUMPTION MIX: The sequence  $(\chi_i)_{i \in \mathbb{Z}}$  is stationary and  $\alpha$ -mixing with mixing coefficient  $\alpha_{\text{mix}}(\cdot)$  of size  $-k/(k-2)$  for some  $k > 2$ .<sup>7</sup>

Assumption MIX imposes a mixing condition on the sequence  $(\chi_i)_{i \in \mathbb{Z}}$  so that, conditional on  $\mathcal{F}$ , the sequence  $(Y_{i\Delta_n})_{i \geq 0}$  is also  $\alpha$ -mixing with mixing coefficients bounded by  $\alpha_{\text{mix}}(\cdot)$ . Note that Assumption MIX only concerns  $(\chi_i)_{i \in \mathbb{Z}}$ . We do not need processes defined on  $(\Omega^{(0)}, \mathcal{F}, \mathbb{P}^{(0)})$  to be mixing. Our use of  $\alpha$ -mixing coefficients is only for concreteness; other types of mixing concepts can also be used. The degree of dependence is controlled by the constant  $k$ . A larger value of  $k$  makes Assumption MIX weaker, but demands stronger dominance conditions as shown below (see Assumption D).

We need some notation for introducing additional assumptions. Let  $\|\cdot\|$  denote the Euclidean norm. For  $j \geq 0$ ,  $p \geq 1$ ,  $\theta \in \Theta$ ,  $\beta, \beta' \in \mathcal{B}$ ,  $z, z' \in \mathcal{Z}$  and  $v, v' \in \mathcal{V}$ , we set

$$\left\{ \begin{array}{l} \bar{g}_{j,p}(\beta, z, v; \theta) \equiv \left( \int \|\partial_v^j g(\mathcal{Y}(\beta, \chi), z, v; \theta)\|^p \mathbb{P}_\chi(d\chi) \right)^{1/p}, \\ \rho_p((\beta, z, v), (\beta', z', v')) \\ \equiv \left( \int \|g(\mathcal{Y}(\beta, \chi), z, v; \theta^*) - g(\mathcal{Y}(\beta', \chi), z', v'; \theta^*)\|^p \mathbb{P}_\chi(d\chi) \right)^{1/p}, \end{array} \right. \quad (3.1)$$

provided that the  $j$ th partial derivative  $\partial_v^j g$  exists. The functions  $\bar{g}_{j,p}(\cdot)$  compute the  $L_p$ -norms of  $g(\mathcal{Y}(\cdot, \chi_i), \cdot, \cdot)$  and its partial derivatives. The function  $\rho_p(\cdot, \cdot)$  computes the  $L_p$ -distance between  $g(\mathcal{Y}(\beta, \chi_i), z, v; \theta^*)$  and  $g(\mathcal{Y}(\beta', \chi_i), z', v'; \theta^*)$  under the probability measure  $\mathbb{P}^{(1)}$ . This semimetric is useful for considering the smoothness of the  $\mathcal{F}$ -conditional moments (such as the covariance and autocovariance) of the sequence  $(g(\mathcal{Y}(\beta, \chi_i), z, v; \theta^*))_{i \geq 0}$  as functions of  $(\beta, z, v)$ .

It is also convenient to introduce a few classes of functions. Let  $\mathcal{A}$  be the collection of all measurable functions that are defined on  $\mathcal{B} \times \mathcal{Z} \times \mathcal{V}$  and take values in some finite-dimensional real space. For  $p \geq 0$ , we set

$$\mathcal{P}(p) \equiv \left\{ \begin{array}{l} f \in \mathcal{A} : \text{for each bounded set } \mathcal{K} \subseteq \mathcal{B} \times \mathcal{Z}, \text{ there exists a constant } K > 0, \\ \text{such that } \|f(\beta, z, v)\| \leq K(1 + v^p) \text{ for all } (\beta, z) \in \mathcal{K} \text{ and } v \in \mathcal{V} \end{array} \right\}$$

and  $\mathcal{C}(p) \equiv \{f \in \mathcal{P}(p) : f \text{ is continuous}\}$ . We denote by  $\mathcal{C}^{2,3}$  the subclass of functions in  $\mathcal{A}$  that are twice continuously differentiable in  $(\beta, z) \in \mathcal{B} \times \mathcal{Z}$  and three times continuously differentiable in  $v \in \mathcal{V}$ . We then set, for  $p \geq 3$ ,

$$\mathcal{C}^{2,3}(p) \equiv \left\{ \begin{array}{l} f \in \mathcal{C}^{2,3} : \text{for each bounded set } \mathcal{K} \subseteq \mathcal{B} \times \mathcal{Z}, \text{ there exists a constant } K > 0, \text{ such} \\ \text{that } \|\partial_v^j f(\beta, z, v)\| \leq K(1 + v^{p-j}) \text{ for all } (\beta, z) \in \mathcal{K}, v \in \mathcal{V} \text{ and } j = 0, 1, 2, 3 \end{array} \right\}.$$

<sup>7</sup>The mixing coefficients are of size  $-a$ ,  $a > 0$ , if they decay at polynomial rate  $a + \varepsilon$  for some  $\varepsilon > 0$ . See Definition 3.45 in White (2001).

The constant  $K$  in the definitions of  $\mathcal{P}(p)$  and  $\mathcal{C}^{2,3}(p)$  is uniform with respect to  $\beta$  and  $z$ , but this requirement is not strong, because we only need the uniformity to hold over a bounded set  $\mathcal{K}$  and we allow  $K$  to depend on  $\mathcal{K}$ .<sup>8</sup> The key restriction on  $\mathcal{P}(p)$ ,  $\mathcal{C}(p)$  and  $\mathcal{C}^{2,3}(p)$  is that their member functions, as well as the derivatives of these functions with respect to  $v$  for the third, have at most polynomial growth in  $v$ . In our analysis, the argument  $v$  often takes value at some estimate of the spot variance, and the polynomial growth condition is used for controlling the effect of approximation error between the spot variance and its estimate.

Our main regularity conditions on  $g(\cdot)$  are given by Assumptions S, D and LIP below.

ASSUMPTION S: (i) The function  $g(y, z, v; \theta)$  is continuously differentiable in  $\theta$  and twice continuously differentiable in  $v$ ; (ii) for some  $p \geq 3$  and each  $\theta \in \Theta$ , we have  $\bar{g}(\cdot; \theta) \in \mathcal{C}^{2,3}(p)$ ,  $\partial_\theta \bar{g}(\cdot; \theta) \in \mathcal{C}(p)$  and  $\partial_\theta \partial_v^2 \bar{g}(\cdot; \theta) \in \mathcal{C}(p-2)$ ; (iii) for each  $\theta \in \Theta$  and  $(\beta, z, v) \in \mathcal{B} \times \mathcal{Z} \times \mathcal{V}$ , we have  $\partial_v^j \bar{g}(\beta, z, v; \theta) = \int \partial_v^j g(\mathcal{Y}(\beta, \chi), z, v; \theta) \mathbb{P}_\chi(d\chi)$  and  $\partial_\theta \partial_v^j \bar{g}(\beta, z, v; \theta) = \int \partial_\theta \partial_v^j g(\mathcal{Y}(\beta, \chi), z, v; \theta) \mathbb{P}_\chi(d\chi)$  for  $j = 0, 1, 2$ .

Assumption S mainly concerns smoothness. Assumption S(i) specifies the basic smoothness requirement on the function  $g(\cdot)$ . Assumption S(ii) imposes additional smoothness conditions on  $\bar{g}(\cdot; \theta)$ . We consider  $\bar{g}(\cdot)$  directly because, as an integrated version of  $g(\cdot)$  (recall (2.7)), it is often smooth even if the latter is not. Assumption S(iii) is a mild condition that allows us to change the order between differentiation and integration. We do not elaborate primitive conditions for it, because they are well known.

In Assumption D below, the function  $\overline{\partial_\theta g}_{j,k}(\cdot, \theta)$  is defined by (3.1) with  $g(\cdot)$  replaced by  $\partial_\theta g(\cdot)$ .

ASSUMPTION D: For some  $k > 2$ ,  $p \geq 3$  and  $\kappa \in (0, 1]$ , we have (i)  $\bar{g}_{0,k}(\cdot; \theta) \in \mathcal{P}((p/2) \vee (2p/k))$ ,  $\overline{\partial_\theta g}_{0,k}(\cdot, \theta) \in \mathcal{P}(p)$  and  $\bar{g}_{2,k}(\cdot; \theta)$ ,  $\overline{\partial_\theta g}_{2,k}(\cdot; \theta) \in \mathcal{P}(p-2)$  for each  $\theta \in \Theta$ ; (ii) for any bounded set  $\mathcal{K} \subseteq \mathcal{B} \times \mathcal{Z}$ , there exists a finite constant  $K > 0$  such that,  $\rho_k(\tilde{z}, \tilde{z}') \leq K(1 + |v|^{p/2-1} + |v'|^{p/2-1}) \|\tilde{z} - \tilde{z}'\|^\kappa$  for all  $\tilde{z}, \tilde{z}' \in \mathcal{K} \times \mathcal{V}$  with  $\|\tilde{z} - \tilde{z}'\| \leq 1$ , where  $\tilde{z} \equiv (\beta, z, v)$  and  $\tilde{z}' \equiv (\beta', z', v')$ .

Assumption D is of the dominance type. Assumption D(i) restricts the  $k$ th  $\mathcal{F}$ -conditional absolute moments to have at most polynomial growth in the spot variance and is mainly needed for using mixing inequalities. Assumption D(ii) is a local dominance condition for the semimetric  $\rho_k(\cdot, \cdot)$ . This condition is weaker when the Hölder exponent  $\kappa$  is closer to zero. The multiplicative factor  $K(1 + |v|^{p/2-1} + |v'|^{p/2-1})$  is uniform in  $(\beta, z, \beta', z')$  on bounded sets and has at most polynomial growth in the arguments that correspond to the spot variance.

DEFINITION 1 (CLASS LIP): Let  $j, p$  be integers such that  $0 \leq j \leq p$ . A function  $(y, z, v, \theta) \mapsto g(y, z, v; \theta)$  on  $\mathcal{Y} \times \mathcal{Z} \times \mathcal{V} \times \Theta$  is said to be in the class LIP( $p, j$ ) if, for each  $0 \leq i \leq j$ , there exists a

<sup>8</sup>Our theory does not need the processes  $\beta_t$  and  $Z_t$  to be bounded. However, by a localization argument, we can assume these processes to be bounded without loss of generality when deriving limit theorems.

function  $B_i(y, z, v)$  such that  $\|\partial_v^i g(y, z, v; \theta) - \partial_v^i g(y, z, v; \theta')\| \leq B_i(y, z, v) \|\theta - \theta'\|$  for all  $\theta, \theta' \in \Theta$  and  $(y, z, v) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{V}$ , and the function  $(\beta, z, v) \mapsto \bar{B}_i(\beta, z, v) \equiv \sqrt{\int B_i(\mathcal{Y}(\beta, \chi), z, v)^2 \mathbb{P}_\chi(d\chi)}$  belongs to  $\mathcal{P}(p - i)$ .

ASSUMPTION LIP: (i)  $g(\cdot) \in \text{LIP}(p, 2)$ ; (ii)  $\partial_\theta g(\cdot) \in \text{LIP}(p, 2)$ .

Assumption LIP imposes a type of Lipschitz condition for  $g(\cdot; \theta)$  and its partial derivatives. This condition is used for establishing uniform (w.r.t.  $\theta$ ) convergence in probability of various sample moment functions. It is also used to show that the effect of replacing the true parameter value with its estimate is asymptotically negligible in the HAC estimation.

For concreteness, we illustrate how to verify the above regularity conditions in the setting of Example 4, which is the main focal point of our numerical work in Sections 4 and 5. Focusing on this example is instructive because it illustrates the key technical argument which is common to many applications.

EXAMPLE 4—CONTINUED: To simplify the discussion, we take the constant  $k$  in Assumptions MIX and D as an integer. We use  $K$  to denote a positive constant which may vary from line to line. We consider an instrument of the form  $\varphi(v) = v^\iota$  for some integer  $\iota \geq 0$ , while noting that setting  $\varphi(\cdot)$  to be scalar-valued is without loss of generality for the purpose of verifying Assumptions S, D and LIP. It is easy to see

$$\begin{aligned} g(y, v; \theta) &= \left( \begin{array}{c} y - \theta_1(\theta_2 + \theta_3 v) \\ y^2 - \theta_1^2(\theta_2 + \theta_3 v)^2 - \theta_1^2(\theta_2 + \theta_3 v) \end{array} \right) v^\iota, \\ \bar{g}(\beta, v; \theta) &= \left( \begin{array}{c} \theta_1^* \beta - \theta_1(\theta_2 + \theta_3 v) \\ \theta_1^{*2}(\beta + \beta^2) - \theta_1^2(\theta_2 + \theta_3 v)^2 - \theta_1^2(\theta_2 + \theta_3 v) \end{array} \right) v^\iota. \end{aligned}$$

Assumption S is verified for any  $p \geq \max\{3, \iota + 2\}$  by direct inspection. By properties of the Poisson distribution,  $\mathbb{E}[|Y_t|^k | \mathcal{F}] \leq K(|\beta_t| + |\beta_t|^k)$ . It is then easy to see that  $\bar{g}_{j,k}(\cdot; \theta) \in \mathcal{P}(\iota + 2 - j)$  for  $j \in \{0, 1, 2\}$ , so Assumption D(i) is verified for  $p \geq \max\{2, k/2\}(\iota + 2)$ . In addition, for  $\beta$  and  $\beta'$  in a bounded set with  $|\beta - \beta'| \leq 1$ , we have  $\mathbb{E}|\chi_i(\beta) - \chi_i(\beta')|^{2k} \leq K|\beta - \beta'|$ . By the Cauchy–Schwarz inequality,  $\mathbb{E}|\chi_i(\beta)^2 - \chi_i(\beta')^2|^k \leq K|\beta - \beta'|^{1/2}$ . It is then easy to see that  $\rho_k((\beta, v), (\beta', v')) \leq K(1 + |v|^{\iota+1} + |v'|^{\iota+1})(|\beta - \beta'|^{1/2k} + |v - v'|)$ . Hence, Assumption D(ii) is verified for  $\kappa = 1/2k$  and  $p \geq 2(\iota + 2)$ . Turning to Assumption LIP, we note that  $\|\partial_v^j g(y, v; \theta) - \partial_v^j g(y, v; \theta')\| \leq K(1 + v^{\iota+2-j})\|\theta - \theta'\|$  for  $\theta, \theta'$  in the compact set  $\Theta$ . Assumption LIP(i) is verified for  $p \geq \iota + 2$ . Assumption LIP(ii) can be verified similarly. To sum up, for any  $k > 2$ , Assumptions S, D and LIP are verified for  $p \geq \max\{2, k/2\}(\iota + 2)$  and  $\kappa = 1/2k$ .

### 3.2 The bias-corrected sample moment function and its asymptotic properties

In this subsection, we construct a sample moment function  $G_n(\cdot)$  for estimating the integrated moment function  $G(\cdot)$  in (2.9). We then present the asymptotic properties of  $G_n(\cdot)$ .

We first nonparametrically recover the spot variance  $V_{i\Delta_n}$  by using a spot truncated realized variation estimator. To this end, we consider a sequence  $k_n$  of integers with  $k_n \rightarrow \infty$  and  $k_n\Delta_n \rightarrow 0$ , which plays the role of the local window for spot variance estimation. The spot variance estimate is given as follows: for each  $i = 0, \dots, [T/\Delta_n] - k_n$ ,

$$\widehat{V}_{i\Delta_n} \equiv \frac{1}{k_n\Delta_n} \sum_{j=1}^{k_n} (\Delta_{i+j}^n X)^2 1_{\{|\Delta_{i+j}^n X| \leq \bar{\alpha}\Delta_n^{\bar{\varpi}}\}}, \quad \text{where } \Delta_{i+j}^n X \equiv X_{(i+j)\Delta_n} - X_{(i+j-1)\Delta_n},$$

and  $\bar{\alpha} > 0$ ,  $\bar{\varpi} \in (0, 1/2)$  are constants that specify the truncation threshold. This estimator is a localized version of the estimator proposed by Mancini (2001), where the truncation is needed so that the spot variance estimate  $\widehat{V}_{i\Delta_n}$  is robust to jumps in  $X$ .<sup>9</sup> Below, we denote  $N_n \equiv [T/\Delta_n] - k_n$ .

We start with a (seemingly) natural sample-analogue estimator for  $G(\theta)$ , which is given by

$$\widehat{G}_n(\theta) \equiv \Delta_n \sum_{i=0}^{N_n} g\left(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta\right), \quad \theta \in \Theta.$$

Theorem 1 shows that  $\widehat{G}_n(\cdot)$  is a consistent estimator for  $G(\cdot)$  under the uniform metric.

**Theorem 1.** *Suppose (i) Assumptions H and MIX hold for some  $r \in (0, 1)$  and  $k > 2$ ; (ii) for some  $p \geq 0$  and each  $\theta \in \Theta$ ,  $\bar{g}(\cdot; \theta) \in \mathcal{C}(p)$  and  $\bar{g}_{0,k}(\cdot; \theta) \in \mathcal{P}(p)$ ; (iii) if  $p > 1$ , we further assume that  $\bar{\varpi} \geq (p-1)/(2p-r)$ ; (iv)  $g(\cdot) \in LIP(p, 0)$ ; (v)  $k_n \rightarrow \infty$  and  $k_n\Delta_n \rightarrow 0$ . Then  $\widehat{G}_n(\cdot) \xrightarrow{\mathbb{P}} G(\cdot)$  uniformly on compact sets.*

We also need a central limit theorem for the sample moment function (evaluated at  $\theta^*$ ), which is useful for conducting asymptotic inference. It turns out that the “raw” sample analogue  $\widehat{G}_n(\theta)$  does not admit a central limit theorem due to a high-order bias; see Corollary 1 below for a formal statement. Nevertheless, Theorem 1 is useful for establishing the consistency of various estimators, such as that of the asymptotic variance.

We hence consider a *bias-corrected sample moment function* given by

$$G_n(\theta) \equiv \widehat{G}_n(\theta) - \frac{1}{k_n} \widehat{B}_n(\theta), \quad \text{where } \widehat{B}_n(\theta) \equiv \Delta_n \sum_{i=0}^{N_n} \partial_v^2 g\left(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta\right) \widehat{V}_{i\Delta_n}^2. \quad (3.2)$$

This sample moment function is used for defining the GMIM estimator in (2.11). As shown in

<sup>9</sup>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 Renò (2008), Kristensen (2010), and references therein.

Theorem 2 below,  $\Delta_n^{-1/2}G_n(\theta^*)$  enjoys a central limit theorem with a mixed Gaussian asymptotic distribution.

To describe the asymptotic covariance matrix, we need more notation. For each  $l \geq 0$ , we denote the joint distribution of  $(\chi_i, \chi_{i-l})$  by  $\mathbb{P}_{\chi, l}$  and set, for  $(\beta, z, v) \in \mathcal{B} \times \mathcal{Z} \times \mathcal{V}$ ,

$$\gamma_l(\beta, z, v) \equiv \int g(\mathcal{Y}(\beta, \chi), z, v; \theta^*) g(\mathcal{Y}(\beta, \chi'), z, v; \theta^*)^\top \mathbb{P}_{\chi, l}(d\chi, d\chi'). \quad (3.3)$$

We then set

$$\begin{cases} \bar{\gamma}(\beta, z, v) \equiv \gamma_0(\beta, z, v) + \sum_{l=1}^{\infty} (\gamma_l(\beta, z, v) + \gamma_l(\beta, z, v)^\top), \\ \bar{\Gamma} \equiv \int_0^T \bar{\gamma}(\beta_s, Z_s, V_s) ds. \end{cases} \quad (3.4)$$

Here,  $\gamma_l(\beta, z, v)$  is the  $\mathcal{F}$ -conditional autocovariance of the sequence  $g(\mathcal{Y}(\beta, \chi_i), z, v; \theta^*)$  at lag  $l$ , and  $\bar{\gamma}(\beta, z, v)$  is the corresponding “long-run” covariance matrix.<sup>10</sup> Finally, we set

$$\bar{S} \equiv 2 \int_0^T \partial_v \bar{g}(\beta_s, Z_s, V_s; \theta^*) \partial_v \bar{g}(\beta_s, Z_s, V_s; \theta^*)^\top V_s^2 ds. \quad (3.5)$$

The ( $\mathcal{F}$ -conditional) asymptotic covariance matrix of  $\Delta_n^{-1/2}G_n(\theta^*)$  is given by

$$\Sigma_g \equiv \bar{\Gamma} + \bar{S}, \quad (3.6)$$

where  $\bar{\Gamma}$  arises from the serially dependent random disturbances  $(\chi_i)_{i \geq 0}$  and  $\bar{S}$  arises from the first-step sampling error in  $\widehat{V}_{i\Delta_n}$ .

We are now ready to state the asymptotic properties of  $G_n(\cdot)$ . We then characterize the aforementioned high-order bias of the raw estimator  $\widehat{G}_n(\theta^*)$  as a direct corollary (Corollary 1). In the sequel, we use  $\xrightarrow{\mathcal{L}\text{-}\mathfrak{s}}$  to denote  $\mathcal{F}$ -stable convergence in law<sup>11</sup> and, for a generic  $\mathcal{F}$ -measurable positive semidefinite matrix  $\Sigma$ , we use  $\mathcal{MN}(0, \Sigma)$  to denote the centered mixed Gaussian distribution with  $\mathcal{F}$ -conditional covariance matrix  $\Sigma$ . We shall assume the following for the local window  $k_n$ .

ASSUMPTION LW:  $k_n^2 \Delta_n \rightarrow 0$  and  $k_n^3 \Delta_n \rightarrow \infty$ .

<sup>10</sup>The process  $\bar{\gamma}(\beta_t, Z_t, V_t)$  may be more properly referred to as the *local long-run covariance matrix*, as it is evaluated locally at time  $t$ . It arises from a large number of adjacent observations that are serially dependent (through  $\chi_i$ ), but all these observations are sampled from an asymptotically shrinking time window. In other words,  $\bar{\gamma}(\beta_t, Z_t, V_t)$  is long-run in tick time, but local in calendar time. The series in (3.4) is absolutely convergent. Indeed, under Assumption MIX, by the mixing inequality,  $\|\gamma_0(\beta, z, v)\| + \sum_{l=1}^{\infty} \|\gamma_l(\beta, z, v)\| \leq K \bar{g}_{0,k}(\beta, z, v; \theta^*)^2$ . Therefore,  $\bar{\gamma}(\beta, z, v)$  is finite whenever  $\bar{g}_{0,k}(\beta, z, v; \theta^*)$  is finite, for which Assumption D suffices.

<sup>11</sup>Stable convergence in law is slightly 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 variable and stable convergence allows for feasible inference using a consistent estimator for its  $\mathcal{F}$ -conditional variance. See Jacod and Shiryaev (2003) for further details on stable convergence.

**Theorem 2.** Suppose (i) Assumptions *H*, *MIX*, *S*, *D* and *LW* hold for some  $r \in (0, 1)$ ,  $k > 2$  and  $p \geq 3$ ; (ii)  $\varpi \geq (2p - 1)/2(2p - r)$ . Then

(a) under Assumption *LIP*(i),  $\widehat{B}_n(\theta) \xrightarrow{\mathbb{P}} \int_0^T \partial_v^2 \bar{g}(\beta_s, Z_s, V_s; \theta) V_s^2 ds$  and  $G_n(\theta) \xrightarrow{\mathbb{P}} G(\theta)$ , uniformly in  $\theta$  on compact sets;

(b)  $\Delta_n^{-1/2} G_n(\theta^*) \xrightarrow{\mathcal{L}\text{-s}} \mathcal{MN}(0, \Sigma_g)$ .

**Corollary 1.** Under the conditions in Theorem 2,  $k_n \widehat{G}_n(\theta^*) \xrightarrow{\mathbb{P}} \int_0^T \partial_v^2 \bar{g}(\beta_s, Z_s, V_s; \theta^*) V_s^2 ds$ .

COMMENTS. (i) Theorem 2(a) shows the uniform consistency of  $G_n(\cdot)$ . This result is a simple consequence of Theorem 1 and is used for establishing the consistency of the GMIM estimator.

(ii) Theorem 2(b) characterizes the stable convergence of  $\Delta_n^{-1/2} G_n(\theta^*)$ . The rate of convergence is parametric, as is typical in semiparametric problems. Note that  $G_n(\theta^*)$  is centered at zero because of (2.10). We only consider  $G_n(\cdot)$  evaluated at the true value  $\theta^*$  because this is enough for conducting asymptotic inference on the basis of (2.10).

(iii) In the special case where  $g(y, z, v; \theta^*)$  does not depend on  $y$  and  $z$ , Theorem 2(b) coincides with Theorem 3.2 of Jacod and Rosenbaum (2013), which concerns the estimation of integrated volatility functionals of the form  $\int_0^T g(V_s) ds$ . For the same technical reasons as here, Jacod and Rosenbaum (2013) (see (3.6) there) also adopt Assumption *LW* to restrict the range of rates at which  $k_n$  grows to infinity. Jacod and Rosenbaum (2013) show<sup>12</sup> that  $\Delta_n^{-1/2} \widehat{G}_n(\theta^*)$  contains several bias terms of order  $O_p(k_n \sqrt{\Delta_n})$  which arise from border effects, diffusive movement of the spot variance process, and volatility jumps, with the latter two being very difficult (if possible) to correct. As a consequence, the condition  $k_n^2 \Delta_n \rightarrow 0$  is needed to make these bias terms asymptotically negligible. However, an additional bias term (which is characterized by Corollary 1) remains in  $\Delta_n^{-1/2} \widehat{G}_n(\theta^*)$ , which is of the order  $O_p(1/k_n \sqrt{\Delta_n})$  and is explosive when  $k_n^2 \Delta_n \rightarrow 0$ . This bias term has to be explicitly corrected for the purpose of deriving a well-behaved limit theorem; the correction term  $k_n^{-1} \widehat{B}_n(\cdot)$  in (3.2) exactly fulfills this task.

### 3.3 Estimation of asymptotic covariance matrices

In this subsection, we describe estimators for the asymptotic covariance matrix  $\Sigma_g$ . These estimators are essential for conducting feasible inference. We start with the estimation of  $\bar{\Gamma}$  (recall (3.4)). Let  $\hat{\theta}_n$  be a preliminary estimator of  $\theta^*$ . We consider the sample analogue of  $\int_0^T \gamma_l(\beta_s, Z_s, V_s; \theta^*) ds$ ,  $l \geq 0$ , given by

$$\widehat{\Gamma}_{l,n}(\hat{\theta}_n) \equiv \Delta_n \sum_{i=l}^{N_n} g\left(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \hat{\theta}_n\right) g\left(Y_{(i-l)\Delta_n}, Z_{(i-l)\Delta_n}, \widehat{V}_{(i-l)\Delta_n}; \hat{\theta}_n\right)^\top. \quad (3.7)$$

<sup>12</sup>See Theorem 3.1 in Jacod and Rosenbaum (2013).



Following Newey and West (1987), we consider a kernel function  $w(j, m_n)$  and a bandwidth sequence  $m_n$  of integers. The estimator for  $\bar{\Gamma}$  is then given by

$$\widehat{\Gamma}_n(\hat{\theta}_n) \equiv \widehat{\Gamma}_{0,n}(\hat{\theta}_n) + \sum_{j=1}^{m_n} w(j, m_n) \left( \widehat{\Gamma}_{j,n}(\hat{\theta}_n) + \widehat{\Gamma}_{j,n}(\hat{\theta}_n)^\top \right). \quad (3.8)$$

We need the following condition for studying the asymptotics of  $\widehat{\Gamma}_n(\hat{\theta}_n)$ .

ASSUMPTION HAC: (i) The kernel function  $w(\cdot, \cdot)$  is uniformly bound and for each  $j \geq 1$ ,  $\lim_{m \rightarrow \infty} w(j, m) = 1$ ; (ii)  $m_n \rightarrow \infty$  and  $m_n k_n^{-\kappa/2} \rightarrow 0$ , where  $\kappa \in (0, 1]$  is the constant given in Assumption D; (iii) the function  $\bar{g}_{0,2k}(\cdot; \theta^*)$  is bounded on bounded sets.

As in Newey and West (1987), when the kernel function  $w(\cdot, \cdot)$  is chosen properly,  $\widehat{\Gamma}_n(\hat{\theta}_n)$  is positive semidefinite in finite samples; one example is to take  $w(j, m) = 1 - j/(m + 1)$ , that is, the Bartlett kernel. In this paper, we restrict attention to kernels with bounded support. It is possible to consider estimators with more general forms as considered by Andrews (1991). Since the efficient estimation of the asymptotic covariance matrix is not the primary focus of the current paper, we leave this complication to a future study.

We consider two estimators for  $\bar{S}$ . The first estimator is applicable in a general setting. We choose a sequence of integers  $k'_n$  and set

$$\hat{\eta}_i^n(\hat{\theta}_n) \equiv \frac{1}{k'_n} \sum_{j=0}^{k'_n-1} \partial_v g \left( Y_{(i+j)\Delta_n}, Z_{(i+j)\Delta_n}, \widehat{V}_{i\Delta_n}; \hat{\theta}_n \right), \quad i \geq 0.$$

The variable  $\hat{\eta}_i^n(\hat{\theta}_n)$  serves as an approximation of  $\partial_v \bar{g}(\beta_{i\Delta_n}, Z_{i\Delta_n}, V_{i\Delta_n}; \theta^*)$ . We then set

$$\widehat{S}_{1,n}(\hat{\theta}_n) \equiv 2\Delta_n \sum_{i=0}^{N_n \wedge ([T/\Delta_n] - k'_n + 1)} \hat{\eta}_i^n(\hat{\theta}_n) \hat{\eta}_i^n(\hat{\theta}_n)^\top \widehat{V}_{i\Delta_n}^2.$$

We need Assumption AVAR1 below for the consistency of  $\widehat{S}_{1,n}(\hat{\theta}_n)$  toward  $\bar{S}$ .

ASSUMPTION AVAR1: (i)  $\bar{g}_{1,k}(\cdot; \theta^*)$ ,  $\partial_\beta \partial_v \bar{g}(\cdot; \theta^*)$  and  $\partial_z \partial_v \bar{g}(\cdot; \theta^*)$  belong to  $\mathcal{P}(p/2 - 1)$ ; (ii)  $k'_n \rightarrow \infty$  and  $k'_n \Delta_n \rightarrow 0$ .

Assumption AVAR1(i) imposes dominance conditions for the moments, as well as their derivatives with respect to  $\beta$  and  $z$ , of  $\partial_v g(\mathcal{Y}(\cdot, \chi), \cdot, \cdot; \theta^*)$ . Assumption AVAR1(ii) imposes mild conditions on the sequence  $k'_n$ . While  $k'_n$  is allowed to be different from  $k_n$ , setting  $k'_n = k_n$  is a convenient choice.

The second estimator for  $\bar{S}$  is designed to exploit a special structure of regression models, which is formalized by the following assumption.

ASSUMPTION AVAR2: There exists a function  $(z, v; \theta) \mapsto \tilde{\varphi}(z, v; \theta)$  with a known functional form such that for some  $p \geq 1$ , (i)  $\partial_v \bar{g}(\beta_t, Z_t, V_t; \theta^*) = \tilde{\varphi}(Z_t, V_t; \theta^*)$  for all  $t \in [0, T]$ ; (ii)  $\tilde{\varphi}(\cdot; \theta^*) \in \mathcal{C}(p-1)$ ; (iii)  $\tilde{\varphi}(\cdot) \in \text{LIP}(p-1, 0)$ .

Assumption AVAR2(i) posits that the value of  $\partial_v \bar{g}(\beta_t, Z_t, V_t; \theta^*)$  can be computed from the realizations of  $Z_t$  and  $V_t$ , provided that  $\theta^*$  is known. Assumption AVAR2(ii) imposes some mild smoothness requirements on  $\tilde{\varphi}(\cdot; \theta^*)$ . Assumption AVAR2(iii) says that  $\tilde{\varphi}(\cdot)$  is smooth in  $\theta$ , so that replacing  $\theta^*$  with its preliminary estimator results in an asymptotically negligible effect. The example below shows that, in a nonlinear regression setting such as Example 2, Assumption AVAR2 imposes essentially no additional restrictions beyond Assumptions S(ii) and LIP(i).

EXAMPLE 2—CONTINUED: Under the setting of Example 2, it is easy to see that  $\partial_v \bar{g}(\beta_t, Z_t, V_t; \theta) = (\beta_{1,t} - f(Z_t, V_t; \theta)) \otimes \partial_v \varphi(Z_t, V_t; \theta) - \partial_v f(Z_t, V_t; \theta) \otimes \varphi(Z_t, V_t; \theta)$ . We set  $\tilde{\varphi}(z, v; \theta) \equiv -\partial_v f(z, v; \theta) \otimes \varphi(z, v; \theta)$  and note that Assumption AVAR2(i) readily follows because  $\beta_{1,t} \equiv f(Z_t, V_t; \theta^*)$ . Assumptions AVAR2(ii) and AVAR2(iii) are related to and are somewhat weaker than Assumptions S(ii) and LIP(i), respectively. To see the connection, we note that

$$\begin{aligned} \partial_v \bar{g}(\beta, z, v; \theta^*) &= (\beta_1 - f(z, v; \theta^*)) \otimes \partial_v \varphi(z, v; \theta^*) + \tilde{\varphi}(z, v; \theta^*), \\ \partial_v g(y, z, v; \theta) &= (y - f(z, v; \theta)) \otimes \partial_v \varphi(z, v; \theta) + \tilde{\varphi}(z, v; \theta). \end{aligned}$$

While Assumptions S(ii) and LIP(i) imply that  $\partial_v \bar{g}(\cdot; \theta^*) \in \mathcal{C}(p-1)$  and  $\partial_v g(\cdot) \in \text{LIP}(p-1, 0)$ , Assumptions AVAR2(ii) and AVAR2(iii) only require the second component in each of the two displayed decompositions above to satisfy the same regularity condition.

The second estimator for  $\bar{S}$  is given by

$$\widehat{S}_{2,n}(\hat{\theta}_n) \equiv 2\Delta_n \sum_{i=0}^{N_n} \tilde{\varphi}(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \hat{\theta}_n) \tilde{\varphi}(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \hat{\theta}_n)^\top \widehat{V}_{i\Delta_n}^2.$$

**Theorem 3.** Suppose (i) the conditions in Theorem 2 (ii)  $\Delta_n^{-1/2}(\hat{\theta}_n - \theta^*) = O_p(1)$ . Then

- (a) under Assumption HAC,  $\widehat{\Gamma}_n(\hat{\theta}_n) \xrightarrow{\mathbb{P}} \bar{\Gamma}$ ;
- (b) under Assumption AVAR1,  $\widehat{S}_{1,n}(\hat{\theta}_n) \xrightarrow{\mathbb{P}} \bar{S}$ ;
- (c) under Assumption AVAR2,  $\widehat{S}_{2,n}(\hat{\theta}_n) \xrightarrow{\mathbb{P}} \bar{S}$ .

Consequently, under Assumptions HAC and AVAR1 (resp. AVAR2),  $\widehat{\Sigma}_{g,n}(\hat{\theta}_n) \equiv \widehat{\Gamma}_n(\hat{\theta}_n) + \widehat{S}_{1,n}(\hat{\theta}_n)$  (resp.  $\widehat{\Sigma}_{g,n}(\hat{\theta}_n) \equiv \widehat{\Gamma}_n(\hat{\theta}_n) + \widehat{S}_{2,n}(\hat{\theta}_n)$ ) is a consistent estimator of  $\Sigma_g$ .

COMMENTS. (i) The preliminary estimator  $\hat{\theta}_n$  is assumed to be  $\Delta_n^{-1/2}$ -consistent. The GMIM estimator satisfies this condition; see Proposition 1 below.

(ii) The HAC estimator  $\widehat{\Gamma}_n(\hat{\theta}_n)$  is valid under the assumption that  $(\chi_i)_{i \geq 0}$  is weakly dependent. If it is known *a priori* that  $(\chi_i)_{i \geq 0}$  forms an independent sequence, then  $\bar{\Gamma} = \int_0^T \gamma_0(\beta_s, Z_s, V_s) ds$ ,

which can be consistently estimated by  $\widehat{\Gamma}_{0,n}(\widehat{\theta}_n)$ . Indeed, an intermediate step of the proof of Theorem 3(a) is to show that  $\widehat{\Gamma}_{l,n}(\widehat{\theta}_n) \xrightarrow{\mathbb{P}} \int_0^T \gamma_l(\beta_s, Z_s, V_s) ds$  for each  $l \geq 0$ .

As a direct consequence of Theorems 2 and 3, we can construct Anderson–Rubin–type confidence sets for  $\theta^*$  by inverting tests. To this end, we consider a function  $L(\cdot, \cdot) : \mathbb{R}^q \times \mathbb{R}^{q \times q} \mapsto \mathbb{R}$  and a test statistic of the form  $L_n(\theta) \equiv L(\Delta_n^{-1/2} G_n(\theta), \widehat{\Sigma}_{g,n}(\theta))$ , where  $\widehat{\Sigma}_{g,n}(\cdot)$  is given by Theorem 3. We let  $\alpha \in (0, 1)$  denote the significance level.

**Corollary 2.** *Suppose (i) the conditions in Theorem 3 hold; (ii) the function  $(u, A) \mapsto L(u, A)$  is continuous at  $(u, A)$  for all  $u \in \mathbb{R}^q$  and for almost every  $A$  under the distribution of  $\Sigma_g$ . Then*

(a)  $L_n(\theta^*) \xrightarrow{\mathcal{L}\text{-s}} L(\xi, \Sigma_g)$ , where the variable  $\xi$  is defined on an extension of the space  $(\Omega, \mathcal{F} \otimes \mathcal{G}, \mathbb{P})$  and, conditional on  $\mathcal{F}$ , is centered Gaussian with covariance matrix  $\Sigma_g$ .

(b) Let  $U$  be a generic  $q$ -dimensional standard normal variable that is independent of  $\mathcal{F} \otimes \mathcal{G}$ . If, in addition, the  $\mathcal{F}$ -conditional distribution of  $L(\xi, \Sigma_g)$  is continuous and strictly increasing at its  $1 - \alpha$  quantile  $cv_{1-\alpha}$ , then the  $1 - \alpha$  quantile of the  $\mathcal{F} \otimes \mathcal{G}$ -conditional distribution of  $L(\widehat{\Sigma}_{g,n}(\theta^*)^{1/2} U, \widehat{\Sigma}_{g,n}(\theta^*))$ , denoted by  $cv_{n,1-\alpha}(\theta^*)$ , converges in probability to  $cv_{1-\alpha}$ . Consequently,  $\mathbb{P}(L_n(\theta^*) \leq cv_{n,1-\alpha}(\theta^*)) \rightarrow 1 - \alpha$ .

Corollary 2(a) establishes the asymptotic distribution of the test statistic  $L_n(\theta^*)$ . Corollary 2(b) further shows that  $cv_{n,1-\alpha}$  forms an asymptotically valid sequence of critical values, as it consistently estimates the  $1 - \alpha$   $\mathcal{F}$ -conditional quantile of the limit variable  $L(\xi, \Sigma_g)$ . We can then construct a sequence of confidence sets  $CS_n \equiv \{\theta \in \Theta : L_n(\theta) \leq cv_{n,1-\alpha}(\theta)\}$ . Since  $\mathbb{P}(L_n(\theta^*) \leq cv_{n,1-\alpha}(\theta^*)) \rightarrow 1 - \alpha$ , we have  $\mathbb{P}(\theta^* \in CS_n) \rightarrow 1 - \alpha$ . That is,  $CS_n$  forms a sequence of confidence sets for  $\theta^*$  with asymptotic level  $1 - \alpha$ .

The confidence set  $CS_n$  is similar to that proposed by Stock and Wright (2000) when the test statistic takes a quadratic form (i.e.  $L(u, A) = u^\top A^{-1} u$ ). In this case, the distribution of the limit variable  $L(\xi, \Sigma_g)$  is chi-square with degree of freedom  $q$  and, hence, the critical value can be chosen as a constant. Since  $CS_n$  is of the Anderson–Rubin type, it is asymptotically valid even if  $\theta^*$  is only “weakly identified,” with the lack of identification considered as an extreme form of weak identification. The test statistic may also take other forms, such as the maximum of t-statistics (i.e.,  $L(u, A) = \max_{1 \leq j \leq q} |u_j|/A_{jj}^{1/2}$ ), as considered by Andrews and Soares (2010). In general, the critical value  $cv_{n,1-\alpha}(\theta)$  depends on  $\theta$  and does not have a closed-form expression, but it can be easily computed by simulation.

### 3.4 Asymptotic properties of the GMIM estimator

We now describe the asymptotic behavior of the GMIM estimator  $\widehat{\theta}_n$  defined by (2.11). With the limit theorems for sample moment functions (Theorems 1 and 2) in hand, we can derive

the asymptotics of  $\hat{\theta}_n$  by using standard techniques from the classical GMM literature (see, e.g., Hansen (1982), Newey and McFadden (1994) and Hall (2005)). Below, we collect a standard set of assumptions, with some slight modifications made so as to accommodate the current setting.

ASSUMPTION GMIM: (i)  $\Theta$  is compact; (ii)  $\theta^*$  is in the interior of  $\Theta$ ; (iii)  $\Xi_n \xrightarrow{\mathbb{P}} \Xi$ , where  $\Xi$  is an  $\mathcal{F}$ -measurable (random) matrix that is positive semidefinite a.s.; (iv)  $\Xi G(\theta) = 0$  a.s. only if  $\theta = \theta^*$ ; (v) for  $H \equiv \int_0^T \partial_{\theta} \bar{g}(\beta_s, Z_s, V_s; \theta^*) ds$ , the random matrix  $H^\top \Xi H$  is nonsingular a.s.

Assumption GMIM(i) imposes compactness on the parameter space. This condition is used to establish the consistency of the GMIM estimator. Assumption GMIM(ii) allows us to derive a linear representation for the GMIM estimator through a Taylor expansion for the first-order condition of the minimization problem (2.11). Assumption GMIM(iii) specifies the limiting behavior of the weighting matrix  $\Xi_n$ . Unlike in the standard GMM setting, the limit  $\Xi$  may be random, which is important because the limiting optimal weighting matrix is typically random in the current setting. Assumption GMIM(iv) is an identification condition, which guarantees the uniqueness of  $\theta^*$  as a minimizer of the population GMIM criterion function  $Q(\theta) \equiv G(\theta)^\top \Xi G(\theta)$ , up to a  $\mathbb{P}$ -null set. This condition is a joint restriction on the population moment function  $G(\cdot)$  and the weighting matrix  $\Xi$ . In particular, when  $\Xi$  has full rank, Assumption GMIM(iv) amounts to saying that  $\theta^*$  is the unique solution to  $G(\theta) = 0$ . This condition is commonly used to specify identification in a GMM setting, but it takes a somewhat nonstandard form here because the population moment function  $G(\cdot)$  is itself a random function. It is instructive to further illustrate the nature of this condition in the simple setting of Example 1: if we set the instrument  $\varphi(v)$  to be  $(1, v)^\top$  as for ordinary least squares, then

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

We see that  $\theta^*$  is the unique solution to  $G(\theta)$  if and only if  $T \int_0^T V_s^2 ds - (\int_0^T V_s ds)^2 \neq 0$ . By the Cauchy–Schwarz inequality,  $T \int_0^T V_s^2 ds - (\int_0^T V_s ds)^2 \geq 0$  and the inequality is strict unless the process  $V_t$  is time-invariant over  $[0, T]$ . In other words, the identification is achieved as soon as the process  $V_t$  is not colinear, in a pathwise sense, with the constant term. Finally, Assumption GMIM(v) is used to derive an asymptotic linear representation of the GMIM estimator.

The asymptotic behavior of the GMIM estimator  $\hat{\theta}_n$  is summarized by Proposition 1 below.

**Proposition 1.** *Suppose (i) Assumptions H, MIX, S, D, LIP, LW and GMIM hold for some  $r \in (0, 1)$ ,  $k > 2$  and  $p \geq 3$ ; (ii)  $\varpi \geq (2p - 1)/2(2p - r)$ . Then*

(a)  $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta^*$ .

(b)  $\Delta_n^{-1/2}(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}\text{-s}} \mathcal{MN}(0, \Sigma)$ , where  $\Sigma \equiv (H^\top \Xi H)^{-1} H^\top \Xi \Sigma_g \Xi H (H^\top \Xi H)^{-1}$ .

(c) Suppose, in addition, Assumptions HAC and AVAR1 (resp. AVAR2) and let  $\widehat{S}_n(\widehat{\theta}_n) \equiv \widehat{S}_{1,n}(\widehat{\theta}_n)$  (resp.  $\widehat{S}_{2,n}(\widehat{\theta}_n)$ ). Denote  $H_n \equiv \partial_{\theta} G_n(\widehat{\theta}_n)$ ,  $\widehat{\Sigma}_{g,n}(\widehat{\theta}_n) \equiv \widehat{\Gamma}_n(\widehat{\theta}_n) + \widehat{S}_n(\widehat{\theta}_n)$  and  $\widehat{\Sigma}_n \equiv (H_n^{\top} \Xi_n H_n)^{-1} H_n^{\top} \Xi_n \widehat{\Sigma}_{g,n}(\widehat{\theta}_n) \Xi_n H_n (H_n^{\top} \Xi_n H_n)^{-1}$ . We have  $\widehat{\Sigma}_n \xrightarrow{\mathbb{P}} \Sigma$ .

COMMENT. Proposition 1(a) shows the consistency of the GMIM estimator  $\widehat{\theta}_n$ . Part (b) shows the associated stable convergence in law, where the asymptotic distribution is centered mixed Gaussian with ( $\mathcal{F}$ -conditional) asymptotic covariance matrix  $\Sigma$ . The asymptotic covariance matrix can be consistently estimated by  $\widehat{\Sigma}_n$  as shown by part (c).

The asymptotic covariance matrix  $\Sigma$  has a familiar form as in the classical GMM setting (Hansen (1982)), although  $\Sigma$  is a random matrix here. Similar to the well-known result in the GMM literature, the asymptotic covariance matrix  $\Sigma$  is minimized in the matrix sense when  $\Xi = \Sigma_g^{-1}$ , provided that  $\Sigma_g$  is nonsingular a.s. A feasible *efficient GMIM estimator* can be obtained by first computing a preliminary GMIM estimator, say  $\tilde{\theta}_n$ , with the identity weighting matrix and then conduct the GMIM estimation with the weighting matrix  $\Xi_n = \widehat{\Sigma}_{g,n}(\tilde{\theta}_n)^{-1}$ . Here, the efficiency is with respect to the choice of weighting matrix while taking the instrument  $\varphi(\cdot)$  as given. The choice of optimal instrument and, as a matter of fact, the characterization of the semi-parametric efficiency bound in the current in-fill setting for nonstationary dependent data remain open questions. Efficient estimation of integrated volatility functionals of the form  $\int_0^T g(V_s) ds$  has been recently tackled by Clément, Delattre, and Gloter (2013), Jacod and Rosenbaum (2013) and Renault, Sarisoy, and Werker (2013). Efficiency may also be improved by considering a continuum of instruments as in Carrasco, Chernov, Florens, and Ghysels (2007). Extending these results to the analysis of GMIM appears to be very challenging and is left to future research.

Hansen's (1982) overidentification test can be adapted to the current setting with the familiar chi-square distribution, as shown by the following proposition.

**Proposition 2.** *Suppose (i) conditions in Proposition 1; (ii)  $\Sigma_g$  is non-singular a.s. and  $\Xi = \Sigma_g^{-1}$ . Then  $\Delta_n^{-1} Q_n(\widehat{\theta}_n) \xrightarrow{\mathcal{L}-s} \chi_{q-\dim(\theta)}^2$ .*

## 4 Simulation results

In this section, we examine the asymptotic theory above in a simulation setting that mimics the setup of our empirical application in Section 5.2. Throughout the simulations, we fix  $T = 21$  days and consider two sampling frequencies:  $\Delta = 1$  or 5 minutes. The window size  $k_n$  in the spot variance estimation is taken to be 120, 150, and 180 for the 1-minute sample, and 40, 45, and 50 for the 5-minute sample. The perturbation on  $k_n$  is reasonably large for checking robustness. We set  $k'_n = k_n$ . For each day, the truncation parameters are taken as  $\varpi = 0.49$  and  $\bar{\alpha} = 3\sqrt{BV}$

where  $BV$  is the daily bipower variation (Barndorff-Nielsen and Shephard (2004b)). There are 2,000 Monte Carlo trials in total.

We simulate  $X_t$  and  $V_t$  according to

$$\begin{cases} dX_t &= (0.05 + 0.5V_t)dt + \sqrt{V_t}dW_t + J_X dN_t + 0.02\lambda_N dt, \\ V_t &= \exp(-2.8 + 6F_t), \quad dF_t = -4F_t dt + 0.8d\widetilde{W}_t + J_F dN_t - 0.02\lambda_N dt, \end{cases} \quad (4.1)$$

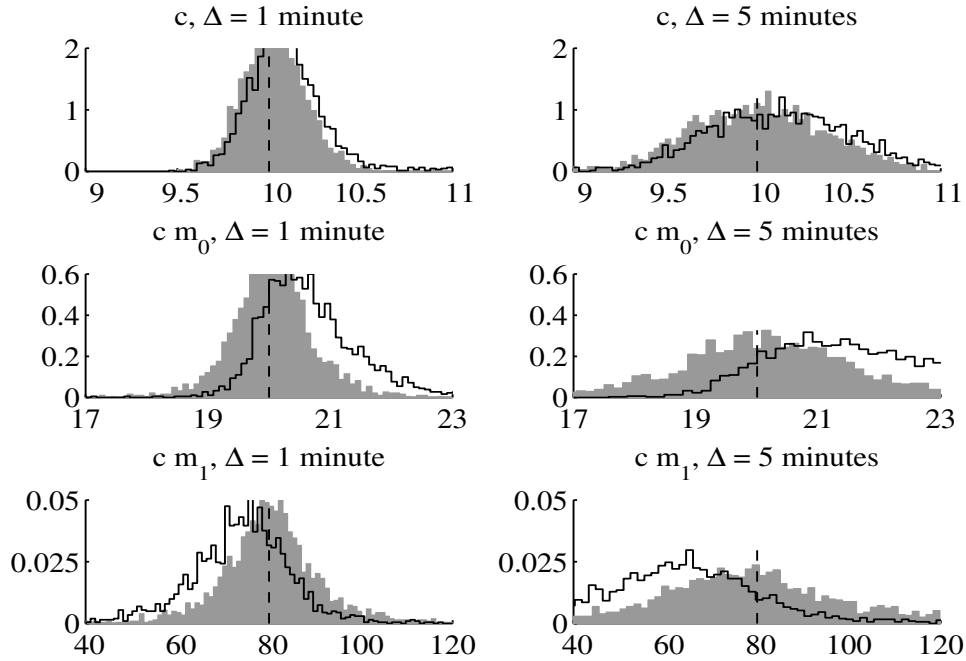
with  $\mathbb{E}[dW_t d\widetilde{W}_t] = -0.75dt$ ,  $J_X \sim \mathcal{N}(-0.02, 0.05^2)$ ,  $J_F \sim \mathcal{N}(0.02, 0.02^2)$ , and  $N_t$  being a Poisson process with intensity  $\lambda_N = 25$ . Given the path of  $V_t$ , the sequence  $(Y_{i\Delta_n})_{i \geq 0}$  is simulated independently with the marginal conditional distribution  $c\text{-Poisson}(m_0 + m_1 V_{i\Delta_n})$ , where  $c = 10$ ,  $cm_0 = 20$ , and  $cm_1 = 80$  are calibrated to data used in Section 5.2. The parameter of interest is  $\theta^* = (c, cm_0, cm_1)$ ; this reparametrization is also used in the empirical study in Section 5.2 as in Andersen (1996).

We conduct this estimation using the first two conditional moments of  $Y_t$  as described by (2.5) in Example 4. We set the instrument as  $\varphi(V_t) = (1, V_t)^\top$ , which results in four integrated moment conditions, leaving one degree of freedom for overidentification. Our goal in this exercise is to examine the finite-sample properties of the GMIM estimator, as well as the rejection rates of the overidentification test. The estimator for the asymptotic covariance matrix  $\Sigma_g$  is taken to be  $\widehat{\Gamma}_{0,n}(\widehat{\theta}_n) + \widehat{S}_{1,n}(\widehat{\theta}_n)$ ; see comment (ii) of Theorem 3. For comparison, we also report results for the uncorrected procedure, which is implemented according to the classical GMM theory but with the spot variance estimate  $\widehat{V}_{i\Delta_n}$  treated as if it were the true spot variance  $V_{i\Delta_n}$ . Note that, for the uncorrected procedure, the asymptotic covariance matrix  $\Sigma_g$  only contains the component  $\bar{\Gamma}$ .

Figure 1 presents finite-sample distributions of the efficient GMIM estimator and the “efficient” uncorrected estimator. We see that the uncorrected estimators exhibit evident biases, while the GMIM estimators are properly centered at the true values. This finding is further confirmed by Table 1, from which we see that, in all Monte Carlo scenarios, the bias of the GMIM estimator is much smaller than that of the uncorrected estimator and fairly close to zero. Moreover, we find that the bias-correction also reduces the root mean squared error (RMSE) of the estimates in most cases.

Figure 2 compares finite-sample distributions of the standardized GMIM and the standardized uncorrected estimators with the asymptotic  $\mathcal{N}(0, 1)$  distribution; the standardization is feasible (i.e., estimators of asymptotic variances are used). As predicted by the asymptotic theory, the distribution of the standardized GMIM estimator is well approximated by the asymptotic  $\mathcal{N}(0, 1)$  distribution for both 1-minute and 5-minute sampling, although some distortion can be seen for the latter. On the contrary, the distribution of the standardized uncorrected estimator differs substantially from  $\mathcal{N}(0, 1)$ .

Figure 1: Histograms of Non-standardized Estimators



Note: This figure compares the finite-sample distributions of the “efficient” uncorrected estimators (solid) and the efficient GMIM estimators (shaded area). The dashed lines highlight the true parameter values. The sampling interval is  $\Delta = 1$  (left) and 5 minutes (right). We set  $T = 21$  days and  $k_n = 150$  and 45 for 1-minute and 5-minute sampling, respectively. There are 2,000 Monte Carlo trials.

Finally, Table 2 reports the finite-sample rejection rates of overidentification tests using the GMIM procedure, along with results from the uncorrected procedure as a comparison. We find that tests based on the uncorrected procedure almost always falsely reject the null hypothesis, which is not surprising in view of the findings above. The rejection rates of the GMIM overidentification test are fairly close to, although slightly lower than, the nominal level for 1-minute sampling. For 5-minute sampling, the GMIM overidentification tests become more undersized. This evidence suggests that in small samples, the GMIM overidentification test tends to be conservative, at least for the Monte Carlo setting considered here.

Table 1: Summary of Monte Carlo Estimation Results

		Panel A. $\Delta = 1$ minute					
		Uncorrected			GMIM		
		$k_n = 120$	$k_n = 150$	$k_n = 180$	$k_n = 120$	$k_n = 150$	$k_n = 180$
$c$	Bias	0.077	0.079	0.072	0.020	0.032	0.033
	RMSE	0.280	0.286	0.283	0.186	0.204	0.213
$c \cdot m_0$	Bias	0.706	0.611	0.560	0.029	0.049	0.083
	RMSE	1.035	0.977	0.905	0.660	0.684	0.661
$c \cdot m_1$	Bias	-7.864	-6.598	-6.398	0.744	0.566	-0.286
	RMSE	12.969	12.564	12.217	10.749	11.236	10.853
		Panel B. $\Delta = 5$ minutes					
		$k_n = 40$	$k_n = 45$	$k_n = 50$	$k_n = 40$	$k_n = 45$	$k_n = 50$
$c$	Bias	0.143	0.140	0.123	0.015	0.029	0.018
	RMSE	0.569	0.555	0.544	0.518	0.498	0.502
$c \cdot m_0$	Bias	1.712	1.656	1.510	0.034	0.145	0.100
	RMSE	2.348	2.283	2.165	1.845	1.824	1.740
$c \cdot m_1$	Bias	-19.717	-19.484	-17.768	1.808	0.011	0.355
	RMSE	28.202	28.282	27.218	30.774	30.078	28.878

Note: We report the bias and the root mean squared error (RMSE) of the “efficient” uncorrected and the efficient GMIM estimators in the simulation for various  $k_n$  values. We set  $T = 21$  days. The sampling interval is  $\Delta = 1$  or 5 minutes. The true parameter values are  $c = 10$ ,  $c \cdot m_0 = 20$  and  $c \cdot m_1 = 80$ . There are 2,000 Monte Carlo trials.

## 5 Empirical applications

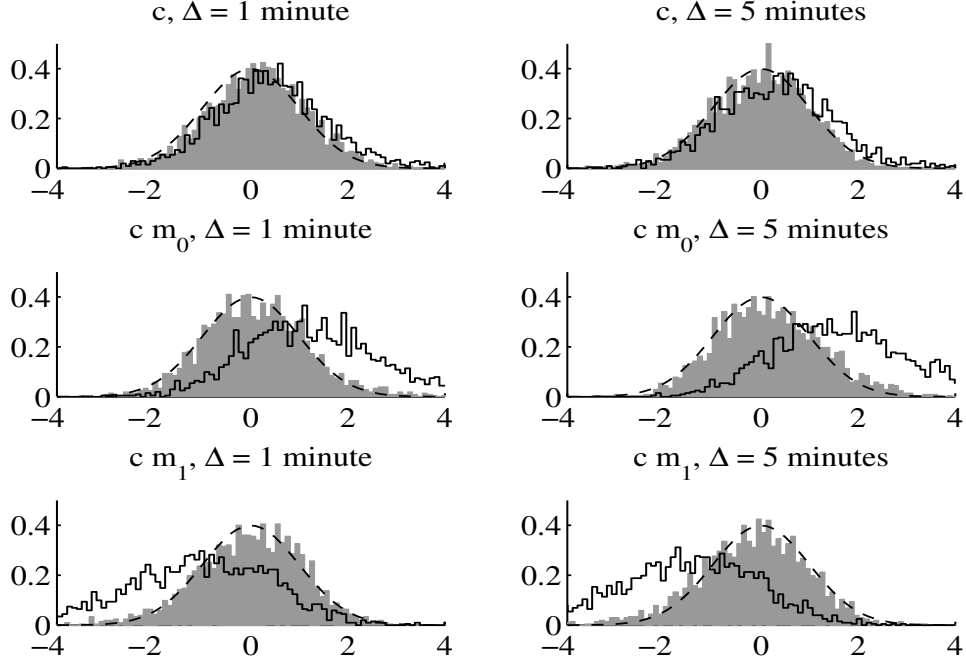
### 5.1 Application 1: VIX pricing models

To illustrate the use of the proposed method, we first apply it to study 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. Starting with the setup, we suppose that the dynamics of the logarithm of the S&P 500 index  $X_t$  under the risk-neutral measure, henceforth the  $\mathbb{Q}$ -measure, follows

$$X_t = X_0 + \int_0^t b_s^{\mathbb{Q}} ds + \int_0^t \sqrt{V_s} dW_s^{\mathbb{Q}} + \int_0^t \int_{\mathbb{R}} z \left( N(ds, dz) - \nu^{\mathbb{Q}}(V_s, dz) ds \right), \quad (5.1)$$



Figure 2: Histograms of Standardized Estimators



Note: This figure compares the finite-sample distributions of the standardized “efficient” uncorrected estimators (solid) and efficient GMIM estimators (shaded area). The  $\mathcal{N}(0, 1)$  density function is plotted for comparison (dashed). The sampling interval is  $\Delta = 1$  (left) and 5 minutes (right). We set  $T = 21$  days and  $k_n = 150$  and 45 for 1-minute and 5-minute sampling, respectively. There are 2,000 Monte Carlo trials.

where the drift  $b_t^{\mathbb{Q}}$  is determined by the no-arbitrage condition,  $W_t^{\mathbb{Q}}$  is a Brownian motion under the  $\mathbb{Q}$ -measure, and  $N(dt, dz)$  is the jump measure of  $X$  with compensator  $\nu^{\mathbb{Q}}(V_t, dz)dt$  which is allowed to depend on the spot variance.<sup>13</sup> We assume that the predictable compensator of the jump quadratic variation is an affine function in the spot variance, that is,  $\int_{\mathbb{R}} z^2 \nu^{\mathbb{Q}}(V_t, dz) = \eta_0 + \eta_1 V_t$ , where  $\eta_0$  and  $\eta_1$  are constants. While this assumption is commonly adopted in empirical work,<sup>14</sup> the discussion below does rely on its validity.

The main focus of this empirical exercise is on the risk-neutral dynamics of the stochastic variance process, which is given by:

$$V_t = V_0 + \int_0^t \kappa^{\mathbb{Q}}(\bar{v}^{\mathbb{Q}} - V_s) ds + M_t^{\mathbb{Q}}, \quad (5.2)$$

<sup>13</sup>See Duffie (2001), Singleton (2006) and Garcia, Ghysels, and Renault (2010) for comprehensive reviews of the no-arbitrage pricing theory and related econometric methods.

<sup>14</sup>This assumption is trivially satisfied if the compensator  $\nu^{\mathbb{Q}}(\cdot)$  does not depend on  $V_t$ . It is also satisfied if  $X_t$  has compound Poisson jumps with its stochastic arrival rate for jumps being a linear function in  $V_t$ . See Chapter 15 of Singleton (2006) and many references therein for detailed examples.

Table 2: Comparison of Monte Carlo Null Rejection Rates

		Uncorrected Procedure			GMIM Procedure		
$\Delta$	Level	$k_n = 120$	$k_n = 150$	$k_n = 180$	$k_n = 120$	$k_n = 150$	$k_n = 180$
	1 %	84.35	84.75	84.10	0.55	0.85	0.80
1 min	5 %	87.85	88.40	87.90	3.10	3.65	4.40
	10%	90.00	90.15	89.95	7.70	7.60	8.25
	Level	$k_n = 40$	$k_n = 45$	$k_n = 50$	$k_n = 40$	$k_n = 45$	$k_n = 50$
	1%	85.70	85.60	83.80	0.20	0.15	0.15
5 min	5%	88.80	89.15	87.55	1.05	1.40	1.95
	10%	90.65	91.25	89.35	3.90	3.90	4.50

Note: We report the finite-sample rejection rates (%) of the overidentification tests for the uncorrected procedure (left) and the GMIM procedure (right) at significance levels 1%, 5% and 10% for various  $k_n$  values. We set  $T = 21$  days. The sampling interval is  $\Delta = 1$  or 5 minutes. There are 2,000 Monte Carlo trials.

where  $\kappa^{\mathbb{Q}}$  and  $\bar{v}^{\mathbb{Q}}$  are model parameters and  $M^{\mathbb{Q}}$  is a martingale under the  $\mathbb{Q}$ -measure that captures both Brownian movements and (compensated) jumps of  $V_t$ . We note that (5.2) only imposes a mean-reverting parametric restriction on the drift term while leaving the martingale component  $M^{\mathbb{Q}}$  completely nonparametric. In particular, we allow for general forms of volatility-of-volatility and volatility jumps. This setting allows us to focus on the specification of the risk-neutral drift of the spot variance. We also note that, since (5.2) only parametrizes the drift term under the  $\mathbb{Q}$  measure, the equivalence between  $\mathbb{P}$  and  $\mathbb{Q}$ , which is implied by no-arbitrage, does not further restrict the dynamics of  $V_t$  under the  $\mathbb{P}$ -measure. This class of risk-neutral volatility models has been widely studied in empirical option pricing and financial econometrics.<sup>15</sup> While it has proven very useful for modeling option prices at the daily or weekly frequency, whether it fulfills the more challenging task of providing a satisfactory pricing specification for intraday data is an open and important question.

We investigate this empirical question by examining the pricing of the VIX. Below, we refer to the squared VIX as the *implied variance*. As shown by Jiang and Tian (2005) and Carr and Wu

<sup>15</sup>Examples include those studied by Bakshi, Cao, and Chen (1997), Bates (2000), Chernov and Ghysels (2000), Pan (2002), Eraker, Johannes, and Polson (2003), Eraker (2004), Ait-Sahalia and Kimmel (2007) and Broadie, Chernov, and Johannes (2007), 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)). This class also include non-Gaussian OU processes considered by Barndorff-Nielsen and Shephard (2001); see also Shephard (2005) for a collection of similar models.

(2009), the theoretical value of the implied variance is given by

$$Y_t^* \equiv \mathbb{E}^{\mathbb{Q}} \left[ \int_t^{t+\tau} V_s ds + \int_t^{t+\tau} \int_{\mathbb{R}} z^2 \nu^{\mathbb{Q}}(V_s, dz) ds \middle| \mathcal{F}_t \right], \quad (5.3)$$

where  $\tau = 21$  trading days and  $\mathbb{E}^{\mathbb{Q}}$  is the expectation operator under  $\mathbb{Q}$ . To derive an analytical expression for  $Y_t^*$ , we first observe that (5.2) implies that  $d\mathbb{E}^{\mathbb{Q}}[V_s|\mathcal{F}_t]/ds = \kappa^{\mathbb{Q}}(\bar{v}^{\mathbb{Q}} - \mathbb{E}^{\mathbb{Q}}[V_s|\mathcal{F}_t])$ . We solve this differential equation for  $\mathbb{E}^{\mathbb{Q}}[V_s|\mathcal{F}_t]$ , which is then plugged into (5.3), yielding

$$Y_t^* = \theta_1^* + \theta_2^* V_t, \quad (5.4)$$

where  $\theta_1^* \equiv \eta_0 + \bar{v}^{\mathbb{Q}}(1 + \eta_1)(1 - (1 - e^{-\kappa^{\mathbb{Q}}\tau})/\kappa^{\mathbb{Q}}\tau)$  and  $\theta_2^* \equiv (1 + \eta_1)(1 - e^{-\kappa^{\mathbb{Q}}\tau})/\kappa^{\mathbb{Q}}\tau$ .

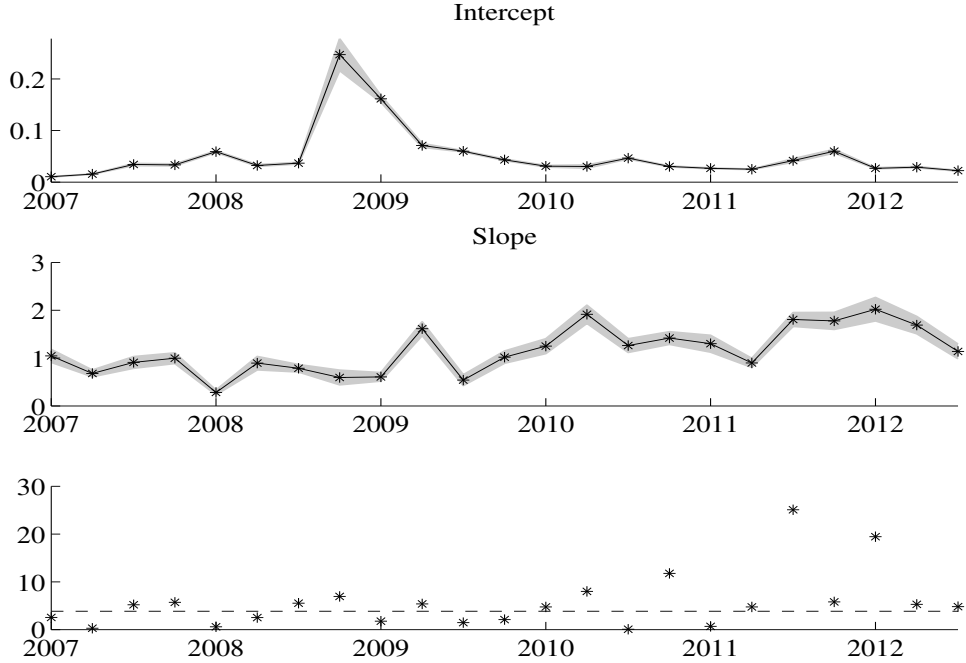
Equation (5.4) highlights the key aspect for using the VIX to study the risk-neutral volatility dynamics. The aforementioned large class of models, although potentially very different from each other with distinct pricing implications for individual options, all imply a linear pricing function for the implied variance. Consequently, specification tests for this class of structural models can be carried out by examining the linear specification (5.4). We do so by conducting the overidentification test (see Proposition 2). As described in Example 1, we suppose that the observed implied variance  $Y_t \equiv \text{VIX}_t^2$  is the theoretical price plus a pricing error such that  $\mathbb{E}[Y_t - \theta_1^* - \theta_2^* V_t | \mathcal{F}] = 0$ . We implement the GMIM procedure with the instrument  $\varphi(V_t) = (1, V_t, 1/(V_t + c))^{\top}$  for  $c = 0.001$ .<sup>16</sup> Note that the third instrument  $1/(V_t + c)$  gives more weight to low volatility levels, while the second instrument  $V_t$  does the opposite.

Our sample period ranges from January 2007 to September 2012, as constrained by data availability; the data source is TickData Inc. The VIX is updated by the CBOE roughly every 15 seconds. The S&P 500 index data is updated more frequently. In order to reduce the asynchronicity between the two time series, we resample the data at every minute. At this frequency, microstructure effects on the S&P 500 index are negligible in our sample. We remind the reader that we allow  $Y_t$  to contain general forms of noise (i.e. pricing error), so microstructure noise in the VIX data is readily accommodated. Days with irregular trading hours are eliminated, resulting in a sample of 1,457 days spanning 23 quarters. Tuning parameters are chosen as follows: the truncation parameters  $\bar{\alpha}$  and  $\varpi$  are set as in the simulation,  $k_n = 150$ , the estimator of asymptotic covariance matrix is given by Proposition 1(c) with  $\widehat{S}_n(\hat{\theta}_n) = \widehat{S}_{1,n}(\hat{\theta}_n)$ , using  $m_n = 12$  and  $k'_n = 150$ .

For each quarter, we estimate parameters in (5.4) via the efficient GMIM estimator. In the upper two panels of Figure 3, we plot the parameter estimates of  $\theta_1^*$  and  $\theta_2^*$  and their confidence

<sup>16</sup>Setting  $c > 0$ , instead of  $c = 0$ , facilitates the verification of regularity conditions in Proposition 2. In particular, notice that  $1/(V_t + c)$  is three-times continuously differentiable in  $V_t$  with bounded derivatives.

Figure 3: Results for the VIX Pricing Model



Note: We plot the time series of quarterly estimates (solid) of the intercept ( $\theta_1$ ) and slope ( $\theta_2$ ) in the linear VIX pricing model, along with their 90% two-sided pointwise confidence bands (shaded area). The lower bound of each confidence band is the 95% lower confidence bound. The bottom panel plots the scaled overidentification test statistic (i.e.  $\Delta_n^{-1}Q_n(\hat{\theta}_n)$ ) (asterisk), and the dashed line indicates the 95% critical value. We fix  $k_n = k'_n = 150$  and  $m_n = 12$ .

intervals. We see that these quarterly estimates exhibit substantial temporal variation. We also conduct an overidentification test for each quarter and plot the value of the test statistic on the bottom panel of Figure 3. We find that the linear specification (5.4) is rejected at the 5% level for 14 out of 23 quarters. The evidence here points away from the linear specification of VIX pricing and, hence, the linear mean-reversion specification of the risk-neutral volatility dynamics given by (5.2), even in sample periods as short as a quarter. This finding suggests that some form of nonlinearity (e.g., an exponential Ornstein–Uhlenbeck specification) needs to be incorporated in the risk-neutral drift term of the spot variance process.

## 5.2 Application 2: Volatility-volume relationship

In the second application, we use high-frequency equity data to investigate the Mixture of Distribution Hypothesis (MDH). The MDH posits a joint dependence between returns and volume on

a latent information flow variable and has spurred a sizable literature in financial economics.<sup>17</sup> A key implication of the MDH is the relationship between return volatility and trading volume. In its classical form, the volatility-volume relationship predicts that the conditional distribution of the trading volume  $Y_t$  given the return variance  $V_t$  is  $\mathcal{N}(\mu_{MDH}V_t, \sigma_{MDH}^2V_t)$ , where  $\mu_{MDH}$  and  $\sigma_{MDH}^2$  are parameters and the normal distribution is motivated by an asymptotic argument; see Tauchen and Pitts (1983). We refer to this model as the *standard MDH*. Andersen (1996) proposes a *modified MDH* on the basis of the Glosten and Milgrom (1985) model. The modified MDH features the random arrival of uninformed and informed traders and predicts that the conditional distribution of trading volume given the spot variance is scaled Poisson, that is,  $Y_t|V_t \sim c \cdot \text{Poisson}(m_0 + m_1V_t)$ . Using Hansen’s (1982) overidentification test on daily data, Andersen (1996) (p. 201) finds that the modified MDH is broadly consistent with the data and performs vastly better than the standard MDH.

Motivated by the fact that the trading activity has increased substantially over the past decade, we take Andersen’s model one step further to address intraday data. As described in Example 4, we set  $Y_{i\Delta_n}$  to be the trading volume within the time interval  $[i\Delta_n, (i+1)\Delta_n)$  and suppose that  $Y_{i\Delta_n}|V_{i\Delta_n} \sim c \cdot \text{Poisson}(m_0 + m_1V_{i\Delta_n})$ . We implement the efficient GMIM estimation procedure and, subsequently, the overidentification test on the basis of the first two conditional moment conditions given by (2.5). We conduct the same exercise for the standard MDH, for which the first two conditional moment conditions are given by (2.2) with

$$\psi(Y_t, V_t; (\mu_{MDH}, \sigma_{MDH}^2)) = (Y_t - \mu_{MDH}V_t, Y_t^2 - \mu_{MDH}^2V_t^2 - \sigma_{MDH}^2V_t)^{\top}.$$

The same instrument  $\varphi(V_t) = (1, V_t)^{\top}$  is used for both the standard MDH and the modified MDH, so as to maintain a fair comparison.

Our sample comprises transaction price and volume data for five tickers: GE, IBM, JPM, MMM, and PG; the data source is the TAQ database. The sample contains 20 quarters from January 2008 to December 2012. In our analysis, each quarter is treated on its own. This sample period includes some of the most volatile periods in modern financial history. Data preprocessing takes a few steps. First, we keep transactions from major exchanges at which most of the trading of these tickers take place.<sup>18</sup> Second, we sample transaction price with a sampling interval  $\Delta = 5$  minutes using the previous-tick approach. The volume  $Y_{i\Delta}$  is the total volume within  $[i\Delta, (i+1)\Delta)$  across all exchanges. To mitigate the impact of block trades, we omit before aggregation all

<sup>17</sup>See, for example, Clark (1973), Epps and Epps (1976), Tauchen and Pitts (1983), Harris (1986), Harris (1987), Richardson and Smith (1994), Andersen (1996) and references therein.

<sup>18</sup>These exchanges include National Association of Securities (ADF), NYSE, NYSE Arca, NASDAQ, Direct Edge A and X, BATS, and BATS Y-Exchanges. NYSE is the exchange where the studied companies are listed, whereas the other exchanges are electronic communication networks. Our results do not change qualitatively when using transactions only from NYSE.

transactions with volumes exceeding 10,000 shares. Next, we delete the U.S. holidays and half-trading days, as well as May 6, 2010 when the “Flash Crash” occurred. As the economic mechanism around opening and closing auctions is very different from the regular intraday trading, we remove data during the first and the last 5 minutes of regular trading hours. Overnight returns and volumes are also eliminated. We do not de-trend the trading volume series, as the trend (if there is any) is unlikely to be important for the quarterly horizon. The unit of the volume series is 10,000 shares. Tuning parameters are chosen as follows: the truncation parameters  $\bar{\alpha}$  and  $\varpi$  are set as in the simulation,  $k_n = 45$ , the estimator of asymptotic covariance matrix is given by Proposition 1(c) with  $\widehat{S}_n(\hat{\theta}_n) = \widehat{S}_{1,n}(\hat{\theta}_n)$ , using  $m_n = 12$  and  $k'_n = 45$ .

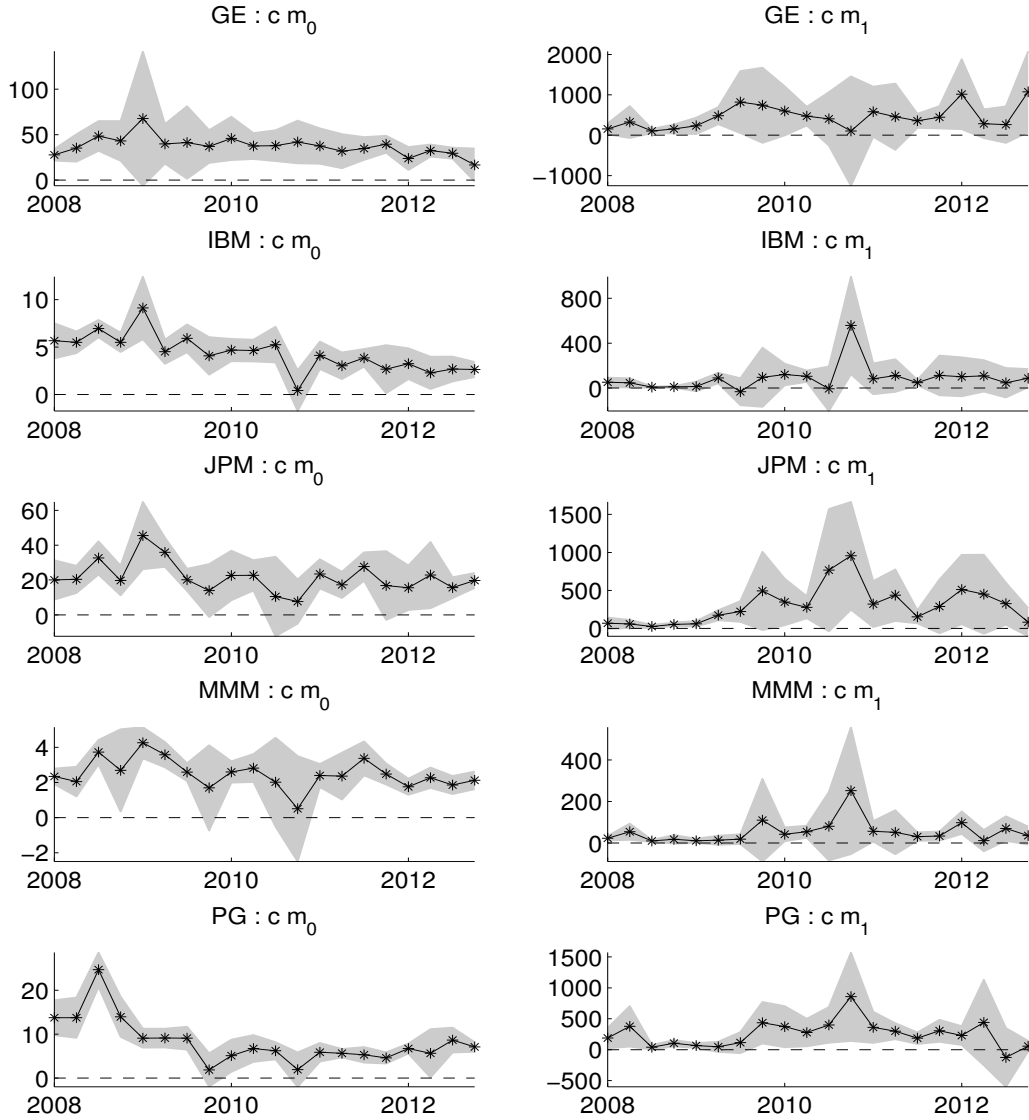
Figure 4 plots the quarterly time series of the efficient GMIM estimates for the modified MDH model. To save space, we only plot the estimates for  $cm_0$  and  $cm_1$ , in that these two parameters determine the conditional mean of volume given the spot variance, that is,  $\mathbb{E}[Y_t|V_t] = cm_0 + cm_1 V_t$ . The findings are summarized as follows. First, consistent with Andersen (1996), the point estimates of  $cm_0$  and  $cm_1$  are almost always positive. We further report in Panels A and B of Table 3 the numbers of quarters with statistically significant estimates, which show that the estimates are indeed significant in most cases. Second, while we observe some temporal variation of the parameter estimates, the parameter instability is not very severe, in the sense that estimates in many adjacent quarters appear to be statistically indifferent. Third, we find that for all tickers, the estimates of  $cm_1$  have quite small values during the financial crisis in 2008. This finding suggests that the underlying information flow has higher price impact during the crisis period than normal periods, which is likely due to the increased level of information asymmetry in the marketplace during the crisis.

We further examine the specification of the modified and the standard MDH for each ticker-quarter using the GMIM overidentification test. In Panel C of Table 3, we report the number of quarters for which the modified MDH is rejected by the test. We see that the modified MDH is rarely rejected and the number of rejections is in line with the Type-I error of our test. By contrast, we see from Panel D of Table 3 that the standard MDH is rejected by the GMIM overidentification test for a majority of quarters for all tickers. These findings are consistent with those of Andersen (1996), and provide further support to the prior findings in a high-frequency setting.

## 6 Related literature

This paper is related to several strands of literature. First, it is closely related to prior work on nonparametric inference for integrated volatility functionals; see Andersen, Bollerslev, Diebold, and Labys (2003), Barndorff-Nielsen and Shephard (2004a), Jacod and Protter (2012) and many references therein. The most closely related paper is the recent work of Jacod and Rosenbaum

Figure 4: Quarterly Parameter Estimates for the Modified MDH Model



Note: For each ticker, we plot the time series of quarterly estimates (solid) of  $cm_0$  (left) and  $cm_1$  (right) in the modified MDH model, along with their 90% two-sided pointwise confidence bands (shaded area). The lower bound of each confidence band is the 95% lower confidence bound. From top to bottom, the tickers are GE, IBM, JPM, MMM and PG. We fix  $k_n = k'_n = 45$  and  $m_n = 12$ .

Table 3: Summary of Testing Results for MDH Models

Sig. Level	GE	IBM	JPM	MMM	PG
Panel A. $H_0 : cm_0 = 0$ vs. $H_1 : cm_0 > 0$ .					
5%	19	19	18	18	18
10%	19	19	18	18	18
Panel B. $H_0 : cm_1 = 0$ vs. $H_1 : cm_1 > 0$ .					
5%	11	10	14	11	15
10%	16	10	19	15	15
Panel C. $H_0$ : Modified MDH is correctly specified.					
5%	2	1	0	1	1
10%	2	3	0	2	2
Panel D. $H_0$ : Standard MDH is correctly specified.					
5%	17	15	14	16	14
10%	17	16	14	18	16

Note: For each ticker, we report the number of quarters (out of 20 quarters in total) in which the null hypothesis of interest is rejected at the 5% or 10% significance level. Panel A (resp. Panel B) reports one-sided testing results for the null hypothesis with  $cm_0 = 0$  (resp.  $cm_1 = 0$ ). Panel C (resp. Panel D) reports overidentification testing results for the modified MDH (resp. standard MDH).

(2013), who use the spot volatility estimates to construct estimators for a large class of integrated volatility functionals of the form  $\int_0^T g(V_s)ds$ . The use of spot volatility estimates can be dated back to early work such as Foster and Nelson (1996) and Comte and Renault (1998), to the best of our knowledge. Jacod and Rosenbaum (2013) provide a detailed analysis of the bias from the first-step estimation and propose a bias-correction that is similar to ours. The integrated moment condition  $G(\cdot)$  in the current paper has a more general form because it not only depends on  $V_t$ , but also depends on the observable process  $Z_t$  and the unobservable process  $\beta_t$ ; moreover, the functional form of  $\bar{g}(\cdot)$  is in general unknown as it is partially determined by the unknown distribution  $\mathbb{P}_\chi$ . These complications make our analysis notably different from Jacod and Rosenbaum (2013). Conceptually, the scope of our analysis is very different from the existing literature: while prior work focused on the inference of the volatility itself, we treat its estimation only as a preliminary step and mainly consider the subsequent inference of parameters in economic models.

Second, our semiparametric method can be considered as one with nonparametrically generated regressors.<sup>19</sup> From this viewpoint, the method can be further compared with the literature

<sup>19</sup>Although the first-step spot variance estimation can be considered as a “noisy measurement” of the true spot



on estimating stochastic volatility models using joint in-fill and long-span asymptotics, see, for example, Bollerslev and Zhou (2002), Barndorff-Nielsen and Shephard (2002), Bandi and Phillips (2003), Corradi and Distaso (2006), Gloter (2007), Kanaya and Kristensen (2010), Bandi and Renò (2012) and Todorov and Tauchen (2012). These papers use realized volatility measures formed from high-frequency data to proxy volatility functionals defined in continuous time. The realized measures can then be used to perform parametric or nonparametric estimation with an appeal to the “large  $T$ ” asymptotics. These methods rely crucially on the in-fill approximation error being dominated by the sampling variability in the long-span asymptotics, so that the former can be considered negligible for asymptotic inference. In contrast, the “fixed  $T$ ” setting here allows us to explicitly characterize the asymptotic bias induced by the in-fill approximation error, construct bias-correction, and incorporate the effect of the approximation error into the asymptotic variance of the GMIM estimator. That being said, the role of the current paper for this literature is completely complementary, because inference concerning certain quantities, such as the drift term (and hence the law) of a stochastic volatility model, demands an asymptotic setting with a long span.

Finally, when specialized in an option pricing setting, the current paper can be compared with Andersen, Fusari, and Todorov (2013). These authors consider a setting where the pricing errors of a large number of option contracts are weakly dependent so that they can be averaged out by virtue of the central limit theorem. This “large cross section” setting simplifies the analysis of option pricing models with multiple latent factors, because the risk factors in each day can be identified from the large cross section as random parameters. Our method is limited to pricing models with one volatility factor, but does not require a large panel of options with cross-sectionally independent or weakly dependent pricing errors. Indeed, we allow the errors in pricing equations to be arbitrarily correlated across option contracts, as is typical in GMM.

## 7 Conclusion

The proposed GMIM framework extends the classical GMM for estimating conditional moment equality models using high-frequency data. Such data have become increasingly available in financial markets during the past decade and provide rich information for studying econometric models. Our asymptotic framework is in-fill with a fixed time span and allows for general forms of nonstationarity and dependence. Since the method can be applied to relatively short samples, it conveniently allows for time-varying parameters across short (e.g., quarterly) subsamples.

The key to our analysis is the derivation of the asymptotic properties of the bias-corrected

---

variance, the nature of our econometric analysis is very different from the literature on errors-in-variables models (see, e.g., Hausman, Newey, Ichimura, and Powell (1991), Schennach (2004, 2007)).

sample moment function, which depends on the noisy process  $Y$ , the observable semimartingale  $Z$  and the spot variance estimate  $\hat{V}$ . Our analysis on the estimator of its asymptotic covariance matrix is also new. Given these technical innovations, inference methods in the classical GMM literature, such as overidentification tests and Anderson–Rubin–type confidence sets, can be adapted to the GMIM setting. The theory is derived under a reasonably general setting, as we allow for complications such as price and volatility jumps, the leverage effect and serially dependent noise in the  $Y$  variable. The usefulness of the proposed method is demonstrated with two distinct empirical examples.

## References

- AÏT-SAHALIA, Y., AND R. KIMMEL (2007): “Maximum Likelihood Estimation of Stochastic Volatility Models,” *Journal of Financial Economics*, 83, 413–452.
- AÏT-SAHALIA, Y., AND J. YU (2009): “High Frequency Market Microstructure Noise Estimates and Liquidity Measures,” *Annals of Applied Statistics*, 3(422-457), 422–457.
- ANDERSEN, T. (1996): “Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility,” *Journal of Finance*, 51, 169–204.
- ANDERSEN, T. G., T. BOLLERSLEV, P. F. CHRISTOFFERSEN, AND F. X. DIEBOLD (2013): “Financial Risk Measurement for Financial Risk Management,” in *Handbook of the Economics of Finance, Vol.II*, ed. by G. Constantinides, M. Harris, and R. Stulz, chap. 17, pp. 1127–1220. Amsterdam: Elsevier Science B.V.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71, 579–625.
- ANDERSEN, T. G., N. FUSARI, AND V. TODOROV (2013): “Parametric Inference and Dynamic State Recovery from Option Panels,” Discussion paper, Northwestern University.
- ANDERSON, T. W., AND H. RUBIN (1949): “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations,” *Annals of Mathematical Statistics*, 20, 46–63.
- ANDREWS, D. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817–858.
- ANDREWS, D. W. K., AND G. SOARES (2010): “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 78(1), 119–157.

- BAKSHI, G., C. CAO, AND Z. CHEN (1997): “Empirical Performance of Alternative Option Pricing Models,” *The Journal of Finance*, 52, 2003–2049.
- BANDI, F., AND R. RENÒ (2012): “Time-varying Leverage Effects,” *Journal of Econometrics*, 169, 94–113.
- BANDI, F. M., AND P. C. B. PHILLIPS (2003): “Fully Nonparametric Estimation of Scalar Diffusion Models,” *Econometrica*, 71, 241–283.
- BANDI, F. M., AND J. R. RUSSELL (2008): “Microstructure Noise, Realized Volatility and Optimal Sampling,” *Review of Economic Studies*, 75, 339–369.
- BARNDORFF-NIELSEN, O., AND N. SHEPHARD (2002): “Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models,” *Journal of the Royal Statistical Society, Series B*, 64.
- 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.
- BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2001): “Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics,” *Journal of the Royal Statistical Society, B*, 63, 167–241.
- (2004a): “Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics,” *Econometrica*, 72(3), pp. 885–925.
- (2004b): “Power and Bipower Variation with Stochastic Volatility and Jumps (with Discussion),” *Journal of Financial Econometrics*, 2, 1–48.
- BATES, D. (2000): “Post-’87 Crash Fears in the S&P 500 Futures Option Market,” *Journal of Econometrics*, 94, 181–238.
- BATES, D. (2012): “U.S. Stock Market Crash Risk, 1926-2010.,” *Journal of Financial Economics*, 105, 229–259.
- BLACK, F. (1976): “Studies of Stock Price Volatility Changes,” in *Proceedings of the 1976 Meetings of the American Statistical Association*, pp. 171–181.
- BOLLERSLEV, T., R. ENGLE, AND D. NELSON (1994): “ARCH Models,” in *Handbook of Econometrics*, vol. 4. Elsevier.

- BOLLERSLEV, T., AND H. ZHOU (2002): “Estimating Stochastic Volatility Diffusion using Conditional Moments of Integrated Volatility,” *Journal of Econometrics*, 109, 33–65.
- BROADIE, M., M. CHERNOV, AND M. JOHANNES (2007): “Model Specification and Risk Premia: Evidence from Futures Options,” *The Journal of Finance*, 62, 1453–1490.
- CARR, P., H. GEMAN, D. B. MADAN, AND M. YOR (2003): “Stochastic Volatility for Lévy Processes,” *Mathematical Finance*, 13, 345–342.
- CARR, P., AND L. WU (2009): “Variance Risk Premiums,” *Review of Financial Studies*, 22, 1311–1341.
- CARRASCO, M., M. CHERNOV, J.-P. FLORENS, AND E. GHYSELS (2007): “Efficient Estimation of General Dynamic Models with a Continuum of Moment Conditions,” *Journal of Econometrics*, 140, 529–573.
- CHERNOV, M., A. R. GALLANT, E. GHYSELS, AND G. T. TAUCHEN (2003): “Alternative Models of Stock Price Dynamics,” *Journal of Econometrics*, 116, 225 – 257.
- CHERNOV, M., AND E. GHYSELS (2000): “A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation,” *Journal of Financial Economics*, 57, 407–458.
- CHRISTOFFERSEN, P., S. HESTON, AND K. JACOBS (2009): “The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well.,” *Management Science*, 55, 1914–1932.
- CLARK, P. K. (1973): “A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices,” *Econometrica*, 41, 135–155.
- 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.
- COMTE, F., AND E. RENAULT (1996): “Long Memory Continuous Time Models,” *Journal of Econometrics*, 73, 101–149.
- (1998): “Long Memory in Continuous-time Stochastic Volatility Models,” *Mathematical Finance*, 8, 291–323.
- CORRADI, V., AND W. DISTASO (2006): “Semiparametric Comparison of Stochastic Volatility Models Using Realized Measures,” *Review of Economic Studies*, 73, 635–667.

- DUFFIE, D. (2001): *Dynamic Asset Pricing Theory*. Princeton University Press, third edn.
- DUFFIE, D., J. PAN, AND K. SINGLETON (2000): “Transform Analysis and Asset Pricing for Affine Jump-Diffusions,” *Econometrica*, 68(6), pp. 1343–1376.
- ENGLE, R. F. (2004): “Risk and Volatility: Econometric Models and Financial Practice,” *American Economic Review*, pp. 405–420.
- EPPS, T. W., AND M. L. EPPS (1976): “The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis,” *Econometrica*, 44(2), 305–321.
- ERAKER, B. (2004): “Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices,” *The Journal of Finance*, 59, 1367–1404.
- ERAKER, B., M. S. JOHANNES, AND N. POLSON (2003): “The Impact of Jumps in Equity Index Volatility and Returns,” *The Journal of Finance*, 58, 1269–1300.
- FOSTER, D., AND D. B. NELSON (1996): “Continuous record asymptotics for rolling sample variance estimators,” *Econometrica*, 64, 139–174.
- GAGLIARDINI, P., C. GOURIÉROUX, AND E. RENAULT (2011): “Efficient Derivative Pricing by the Extended Method of Moments,” *Econometrica*, 79(4), 1181–1232.
- 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, pp. 479–552. North Holland, Amsterdam, The Netherlands.
- GHYSELS, E., A. HARVEY, AND E. RENAULT (1995): “Stochastic Volatility,” in *Handbook of Statistics 14, Statistical Methods in Finance*, ed. by G. Maddala, and C. Rao, chap. 5, pp. 119–191. North Holland, Amsterdam.
- GLOSTEN, L. R., AND P. R. MILGROM (1985): “Bid, Ask, and Transaction Prices in A Specialist Market with Heterogeneously Informed Traders,” *Journal of Financial Economics*, 14, 71–100.
- GLOTER, A. (2007): “Efficient Estimation of Drift Parameters in Stochastic Volatility Models,” *Finance and Stochastics*, 11, 495–519.
- HALL, A. (2005): *Generalized Method of Moments*. Oxford University Press.
- HANSEN, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029–1054.

- HANSEN, L. P., AND K. J. SINGLETON (1982): “Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models,” *Econometrica*, 50(5), pp. 1269–1286.
- HANSEN, P. R., AND A. LUNDE (2006): “Realized Variance and Market Microstructure Noise,” *Journal of Business and Economic Statistics*, 24, 127–161.
- HARRIS, L. (1986): “Cross-Security Tests of the Mixture of Distribution Hypothesis,” *Journal of Financial and Quantitative Analysis*, 21, 39–46.
- (1987): “Transaction Data Tests of the Mixture of Distribution Hypothesis,” *Journal of Financial and Quantitative Analysis*, 22, 127–141.
- HASBROUCK, J. (1993): “Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement,” *Review of Financial Studies*, 6, 191–212.
- HAUSMAN, J., W. NEWEY, H. ICHIMURA, AND J. POWELL (1991): “Identification and Estimation of Polynomial Errors-in-Variables Models,” *Journal of Econometrics*, 50, 273–295.
- HAUTSCH, N. (2012): *Econometrics of Financial High-Frequency Data*. Springer.
- HESTON, S. (1993): “A Closed-form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options,” *Review of Financial Studies*, 6, 327–343.
- JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): “Microstructure Noise in the Continuous Case: The Pre-Averaging Approach,” *Stochastic Processes and Their Applications*, 119, 2249–2276.
- JACOD, J., Y. LI, AND X. ZHENG (2013): “Statistical Properties of Microstructure Noise,” Discussion paper, Hong Kong University of Science and Technology.
- JACOD, J., M. PODOLSKIJ, AND M. VETTER (2010): “Limit Theorems for Moving Averages of Discretized Processes Plus Noise,” *The Annals of Statistics*, 38(3), pp. 1478–1545.
- JACOD, J., AND P. PROTTER (2012): *Discretization of Processes*. Springer.
- JACOD, J., AND M. ROSENBAUM (2013): “Quarticity and Other Functionals of Volatility: Efficient Estimation,” *Annals of Statistics*, 118, 1462–1484.
- JACOD, J., AND A. N. SHIRYAEV (2003): *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York, second edn.
- JIANG, G. J., AND Y. S. TIAN (2005): “The Model-Free Implied Volatility and its Information Content,” *Review of Financial Studies*, 18, 1305–1342.

- KANAYA, S., AND D. KRISTENSEN (2010): “Estimation of Stochastic Volatility Models by Nonparametric Filtering,” Discussion paper, University of Oxford.
- KRISTENSEN, D. (2010): “Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach,” *Econometric Theory*, 26(1), pp. 60–93.
- MANCINI, C. (2001): “Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion,” *Giornale dell’Istituto Italiano degli Attuari*, LXIV, 19–47.
- NEWBY, W. K. (1994): “The Asymptotic Variance of Semiparametric Estimators,” *Econometrica*, 62(6), pp. 1349–1382.
- NEWBY, W. K., AND D. L. MCFADDEN (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, vol. 4. Elsevier.
- NEWBY, W. K., AND K. D. WEST (1987): “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–708.
- PAN, J. (2002): “The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study,” *Journal of Financial Economics*, 63, 3–50.
- RENAULT, E., C. SARISOY, AND B. J. WERKER (2013): “Efficient Estimation of Integrated Volatility and Related Processes,” Discussion paper, Brown University.
- RENÒ, R. (2008): “Nonparametric Estimation of the Diffusion Coefficient of Stochastic Volatility Models,” *Econometric Theory*, 24, 1174–1206.
- RICHARDSON, M., AND T. SMITH (1994): “A Direct Test of the Mixture of Distributions Hypothesis: Measuring the Daily Flow of Information,” *The Journal of Financial and Quantitative Analysis*, 29(1), pp. 101–116.
- SARGAN, J. D. (1958): “The Estimation of Economic Relationships using Instrumental Variables,” *Econometrica*, 26(3), 393–415.
- SCHENNACH, S. M. (2004): “Estimation of Nonlinear Models with Measurement Error,” *Econometrica*, 72(1), pp. 33–75.
- (2007): “Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models,” *Econometrica*, 75(1), pp. 201–239.
- SHEPHARD, N. (2005): *Stochastic Volatility*. Oxford University Press.

- SINGLETON, K. J. (2006): *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment*. Princeton University Press.
- STOCK, J. H., AND J. H. WRIGHT (2000): “GMM with Weak Identification,” *Econometrica*, 68, 1055–1096.
- TAUCHEN, G., AND M. PITTS (1983): “The Price Variability-Volume Relationship on Speculative Markets,” *Econometrica*, 51, 485–505.
- TODOROV, V., AND G. TAUCHEN (2012): “The Realized Laplace Transform of Volatility,” *Econometrica*, 80, 1105–1127.
- WHITE, H. (1982): “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50, 1–25.
- (2001): *Asymptotic Theory for Econometricians*. Academic Press.
- XIU, D. (2010): “Quasi-Maximum Likelihood Estimation of Volatility with High Frequency Data,” *Journal of Econometrics*, 159, 235–250.
- YOSHIHARA, K. (1978): “Moment Inequalities for Mixing Sequences,” *Kodai Mathematical Journal*, 1, 316–328.
- 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.



## A Proofs

The following notations are used throughout the proofs below. We denote the conditional expectation operator  $\mathbb{E}[\cdot|\mathcal{F}]$  by  $\mathbb{E}_{\mathcal{F}}$ . For a random variable  $\xi$  and  $p \geq 1$ , we write  $\|\xi\|_{\mathcal{F},p} = (\mathbb{E}_{\mathcal{F}} \|\xi\|^p)^{1/p}$ . Recall that  $N_n \equiv \lceil T/\Delta_n \rceil - k_n$ . We write  $\sum_i$  for  $\sum_{i=0}^{N_n}$  and write  $\sum_{i,j}$  for  $\sum_{i,j=0}^{N_n}$ . We use  $K$  to denote a generic positive constant that may vary from line to line; we sometimes write  $K_u$  to indicate its dependence on some constant  $u$ . As is typical in this type of problems, by a classical localization argument, we can replace Assumption H with the following assumption without loss of generality.

**ASSUMPTION SH:** We have Assumption H. Moreover, the processes  $\beta_t$ ,  $Z_t$ ,  $\sigma_t$ ,  $\tilde{b}_t$  and  $\tilde{\sigma}_t$  are bounded and, for some  $\lambda$ -integrable function  $J : \mathbb{R} \mapsto \mathbb{R}$ , we have  $|\delta(\omega, t, z)| \leq J(z)$  and  $\|\tilde{\delta}(\omega, t, z)\|^2 \leq J(z)$ , for all  $\omega^{(0)} \in \Omega^{(0)}$ ,  $t \geq 0$  and  $z \in \mathbb{R}$ .

We recall some known, but nontrivial, estimates that are repeatedly used below. Consider a continuous process  $X'_t$  given by

$$X'_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s.$$

We then set, for each  $i = 0, \dots, N_n$ ,

$$\widehat{V}'_{i\Delta_n} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} (\Delta_{i+j}^n X')^2, \quad \tilde{v}_i^n = \widehat{V}'_{i\Delta_n} - V_{i\Delta_n}. \quad (\text{A.1})$$

**Lemma A.1.** *Suppose that Assumption SH holds for some  $r \in (0, 1)$ . Let  $u \geq 1$  be a constant. Then for some deterministic sequence  $a_n \rightarrow 0$ , we have*

$$\begin{cases} \mathbb{E}|\widehat{V}_{i\Delta_n} - \widehat{V}'_{i\Delta_n}|^u \leq K_u a_n \Delta_n^{(2u-r)\varpi+1-u}, \\ \mathbb{E}|\tilde{v}_i^n|^u \leq K_u (k_n^{-u/2} + (k_n \Delta_n)^{(u/2)\wedge 1}), \\ \mathbb{E}|\widehat{V}_{i\Delta_n}|^u \leq K_u + K_u \Delta_n^{(2u-r)\varpi+1-u}. \end{cases} \quad (\text{A.2})$$

**PROOF:** The first inequality is by (4.8) in Jacod and Rosenbaum (2013). The second inequality is by (4.11) in Jacod and Rosenbaum (2013) and Jensen's inequality. The third inequality readily follows from the first two inequalities and the boundedness of  $V_t$ . *Q.E.D.*

### A.1 Proof of Theorem 1

We start with a technical lemma.

**Lemma A.2.** Let  $\tilde{\beta}_t = (\beta_t^\top, Z_t^\top)^\top$ . Suppose (i) Assumption H; (ii) for some  $p \geq 0$ , we have  $f \in \mathcal{C}(p)$  and  $h \in \mathcal{P}(2p)$ ; (iii) if  $p > 1$ , we further assume  $\varpi \in [(p-1)/(2p-r), 1/2)$ ; (iv)  $k_n \rightarrow \infty$  and  $k_n \Delta_n \rightarrow 0$ . Then (a)  $\Delta_n \sum_i f(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \xrightarrow{\mathbb{P}} \int_0^T f(\tilde{\beta}_s, V_s) ds$ ; (b)  $\Delta_n^2 \sum_i h(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \xrightarrow{\mathbb{P}} 0$ .

PROOF: (a) By a localization argument, we suppose Assumption SH without loss of generality. We first prove the assertion under the assumption that  $f$  is bounded. Construct two processes,  $\tilde{\beta}_t^+$  and  $\widehat{V}_t^+$ , as follows: for each  $i \geq 1$  and  $t \in [(i-1)\Delta_n, i\Delta_n)$ , we set  $\tilde{\beta}_t^+ \equiv \tilde{\beta}_{i\Delta_n}$  and  $\widehat{V}_t^+ \equiv \widehat{V}_{i\Delta_n}$ . Observe that

$$\begin{aligned} & \mathbb{E} \left| \Delta_n \sum_i f(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) - \int_0^T f(\tilde{\beta}_s, V_s) ds \right| \\ & \leq K k_n \Delta_n + \int_0^{N_n \Delta_n} \mathbb{E} |f(\tilde{\beta}_s^+, \widehat{V}_s^+) - f(\tilde{\beta}_s, V_s)| ds. \end{aligned}$$

By Theorem 9.3.2 in Jacod and Protter (2012), we have  $\widehat{V}_s^+ \xrightarrow{\mathbb{P}} V_s$  for each  $s \geq 0$ . By the right continuity of the process  $\tilde{\beta}$ , we have  $\tilde{\beta}_s^+ \rightarrow \tilde{\beta}_s$  for each  $s \geq 0$ , which further implies  $(\tilde{\beta}_s^+, \widehat{V}_s^+) \xrightarrow{\mathbb{P}} (\tilde{\beta}_s, V_s)$ . By the continuity of  $f(\cdot)$ ,  $f(\tilde{\beta}_s^+, \widehat{V}_s^+) \xrightarrow{\mathbb{P}} f(\tilde{\beta}_s, V_s)$ . By the bounded convergence theorem,  $\int_0^{N_n \Delta_n} \mathbb{E} |f(\tilde{\beta}_s^+, \widehat{V}_s^+) - f(\tilde{\beta}_s, V_s)| ds \rightarrow 0$ , yielding

$$\Delta_n \sum_i f(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \xrightarrow{\mathbb{P}} \int_0^T f(\tilde{\beta}_s, V_s) ds \quad \text{for bounded } f. \quad (\text{A.3})$$

We now prove the assertion of part (a) with the boundedness condition on  $f$  relaxed. Let  $\phi(\cdot)$  be a  $C^\infty$  function  $\mathbb{R}_+ \mapsto [0, 1]$ , with  $1_{[1, \infty)}(x) \leq \phi(x) \leq 1_{[1/2, \infty)}(x)$ , and for  $m \geq 1$ , we set  $\phi_m(v) \equiv \phi(|v|/m)$ ,  $\phi'_m(v) \equiv 1 - \phi_m(v)$ . We define  $f_m(\tilde{\beta}, v) \equiv f(\tilde{\beta}, v)\phi_m(v)$  and  $f'_m(\tilde{\beta}, v) \equiv f(\tilde{\beta}, v)\phi'_m(v)$ , so  $f(\tilde{\beta}, v) = f_m(\tilde{\beta}, v) + f'_m(\tilde{\beta}, v)$ . Since  $f'_m$  is bounded and continuous, (A.3) yields  $\Delta_n \sum_i f'_m(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) = \int_0^T f'_m(\tilde{\beta}_s, V_s) ds + o_p(1)$ . Since the process  $V_t$  is bounded,  $\int_0^T f'_m(\tilde{\beta}_s, V_s) ds = \int_0^T f(\tilde{\beta}_s, V_s) ds$  for  $m$  large enough. By Proposition 2.2.1 in Jacod and Protter (2012) and Markov's inequality, it remains to show that

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left| \Delta_n \sum_i f_m(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \right| = 0. \quad (\text{A.4})$$

By condition (ii), for all  $m \geq 1$ ,  $|f_m(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n})| \leq K(1 + |\widehat{V}_{i\Delta_n}|^p)\phi_m(\widehat{V}_{i\Delta_n}) \leq K|\widehat{V}_{i\Delta_n}|^p 1_{\{\widehat{V}_{i\Delta_n} \geq m/2\}}$ . Under the same condition on  $X$ , Jacod and Protter (2012) show that (see (9.4.7)), for some deterministic sequence  $a_n \rightarrow 0$ ,  $\mathbb{E}[|\widehat{V}_{i\Delta_n}|^p 1_{\{\widehat{V}_{i\Delta_n} \geq m/2\}}] \leq Km^{-p} + K\Delta_n^{1-p+\varpi(2p-r)} a_n$ . Under condition (iii), we have  $1 - p + \varpi(2p-r) \geq 0$  and, hence,  $\mathbb{E}|\Delta_n \sum_i f_m(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n})| \leq Km^{-p} + O(a_n)$ . From here, (A.4) follows. The proof of part (a) is now complete.

(b) By Lemma A.1 and  $h \in \mathcal{P}(2p)$ ,

$$\mathbb{E} \left[ \Delta_n^2 \sum_i h(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \right] \leq \mathbb{E} \left[ \Delta_n^2 \sum_i (1 + |\widehat{V}_{i\Delta_n}|^{2p}) \right] \leq K\Delta_n + K\Delta_n^{2-2p+\varpi(4p-r)} a_n.$$

Note that condition (iii) implies that  $2 - 2p + \varpi(4p - r) \geq 0$  and, hence, the majorant side of the above inequality is  $o(1)$ . The assertion in part (b) readily follows. *Q.E.D.*

**PROOF OF THEOREM 1:** By a componentwise argument, we can assume that  $g(\cdot)$  is  $\mathbb{R}$ -valued without loss of generality. We first show that  $\widehat{G}_n(\theta) \xrightarrow{\mathbb{P}} G(\theta)$  for each  $\theta \in \Theta$ . We decompose  $\widehat{G}_n(\theta) = \widehat{G}_{1,n}(\theta) + \widehat{G}_{2,n}(\theta)$ , where

$$\begin{aligned} \widehat{G}_{1,n}(\theta) &\equiv \Delta_n \sum_i \zeta_i^n, \quad \zeta_i^n \equiv g(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta) - \bar{g}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta), \\ \widehat{G}_{2,n}(\theta) &\equiv \Delta_n \sum_i \bar{g}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta). \end{aligned}$$

Since  $(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n})$  is  $\mathcal{F}$ -measurable,  $\bar{g}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta) = \mathbb{E}_{\mathcal{F}}[g(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta)]$  by the definition of  $\bar{g}(\cdot)$ . Hence, the variables  $(\zeta_i^n)_{i \geq 0}$  have zero  $\mathcal{F}$ -conditional mean. Under the transition probability  $\mathbb{P}^{(1)}$ , the  $\alpha$ -mixing coefficient of these variables is bounded by  $\alpha_{\text{mix}}(\cdot)$ . Setting  $a_i^n \equiv \bar{g}_{0,k}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta) + 1$  and  $\bar{\zeta}_i^n \equiv \zeta_i^n / a_i^n$ , we have  $\mathbb{E}_{\mathcal{F}}|\bar{\zeta}_i^n|^k \leq 1$  for all  $i$ . Since  $\sum_{l \geq 0} \alpha_{\text{mix}}(l)^{(k-2)/k} < \infty$ , by the mixing inequality (see, e.g., Theorem 3 in Yoshihara (1978)), we have

$$\mathbb{E}_{\mathcal{F}} \left[ \left| \widehat{G}_{1,n}(\theta) \right|^2 \right] \leq K\Delta_n^2 \sum_i (a_i^n)^2 \leq K\Delta_n + K\Delta_n^2 \sum_i \bar{g}_{0,k}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta)^2.$$

Since  $\bar{g}_{0,k}(\cdot; \theta) \in \mathcal{P}(p)$ , we can apply Lemma A.2 (with  $h(\cdot) = \bar{g}_{0,k}(\cdot; \theta)^2 \in \mathcal{P}(2p)$ ) to deduce that the majorant side of the above display is  $o_p(1)$ . From here, it follows that  $\widehat{G}_{1,n}(\theta) = o_p(1)$ . In addition, since  $\bar{g}(\cdot; \theta) \in \mathcal{C}(p)$ , we can apply Lemma A.2 again (with  $f(\cdot) = \bar{g}(\cdot; \theta)$ ) to derive

$$\widehat{G}_{2,n}(\theta) \xrightarrow{\mathbb{P}} \int_0^T \bar{g}(\beta_s, Z_s, V_s; \theta) ds \equiv G(\theta).$$

Hence,  $\widehat{G}_n(\theta) \xrightarrow{\mathbb{P}} G(\theta)$  for each  $\theta \in \Theta$ .

To show the asserted uniform convergence in probability, it remains to show that  $\widehat{G}_n(\cdot)$  is stochastically equicontinuous. Since  $g(\cdot) \in \text{LIP}(p, 0)$ , we see that for  $\theta, \theta' \in \Theta$ ,

$$\left| \widehat{G}_n(\theta) - \widehat{G}_n(\theta') \right| \leq B_{0,n} \|\theta - \theta'\|, \quad \text{where} \quad B_{0,n} \equiv \Delta_n \sum_i B_0(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n})$$

for some function  $B_0(\cdot)$  as described in Definition 1, which satisfies  $\bar{B}_0 \in \mathcal{P}(p)$ . Hence,  $\mathbb{E}_{\mathcal{F}}[B_{0,n}] \leq$

$\Delta_n \sum_i \bar{B}_0(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \leq K \Delta_n \sum_i (1 + |\widehat{V}_{i\Delta_n}|^p)$ . By Lemma A.1,  $\mathbb{E}_{\mathcal{F}}[B_{0,n}] = O_p(1)$ . Since  $B_{0,n}$  is positive, we further derive  $B_{0,n} = O_p(1)$ . From here, it follows that  $\widehat{G}_n(\cdot)$  is stochastically equicontinuous. Q.E.D.

## A.2 Proof of Theorem 2

We need two lemmas. Lemma A.3 is used to combine stable convergence and convergence in conditional law (see Definition 2 below). Lemma A.4 generalizes Theorem 3.2 in Jacod and Rosenbaum (2013).

**DEFINITION 2 (CONVERGENCE IN CONDITIONAL LAW):** Let  $\zeta_n$  be a sequence of  $\mathbb{R}^d$ -valued random variables defined on the space  $(\Omega, \mathcal{F} \otimes \mathcal{G}, \mathbb{P})$  and  $\mathbb{L}$  be a transition probability from  $(\Omega, \mathcal{F} \otimes \{\emptyset, \Omega^{(1)}\})$  to an extension of  $(\Omega, \mathcal{F} \otimes \mathcal{G}, \mathbb{P})$ . We write  $\zeta_n \xrightarrow{\mathcal{L}|\mathcal{F}} \mathbb{L}$  if and only if the  $\mathcal{F}$ -conditional characteristic function of  $\zeta_n$  converges in probability to the  $\mathcal{F}$ -conditional characteristic function of  $\mathbb{L}$ . If a variable  $\zeta$  defined on the extension has  $\mathcal{F}$ -conditional law  $\mathbb{L}$ , we also write  $\zeta_n \xrightarrow{\mathcal{L}|\mathcal{F}} \zeta$ .

**Lemma A.3.** *Let  $\xi_n$  and  $\zeta_n$  be two sequences of random vectors defined on  $(\Omega, \mathcal{F} \otimes \mathcal{G}, \mathbb{P})$  and let  $\xi$  and  $\zeta$  be variables defined on an extension of  $(\Omega, \mathcal{F} \otimes \mathcal{G}, \mathbb{P})$ . Suppose that  $\xi_n$  is  $\mathcal{F}$ -measurable,  $\xi_n \xrightarrow{\mathcal{L}-s} \xi$  and  $\zeta_n \xrightarrow{\mathcal{L}|\mathcal{F}} \zeta$ . Then  $(\xi_n, \zeta_n) \xrightarrow{\mathcal{L}-s} (\xi, \zeta)$ , with  $\xi$  and  $\zeta$  being  $\mathcal{F}$ -conditionally independent.*

**PROOF:** The joint convergence  $(\xi_n, \zeta_n) \xrightarrow{\mathcal{L}-s} (\xi, \zeta)$  is by Proposition 5 in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). It remains to show that  $\xi$  and  $\zeta$  are  $\mathcal{F}$ -conditionally independent. Let  $f(\cdot)$  and  $g(\cdot)$  be bounded continuous functions and  $U$  be a bounded  $\mathcal{F}$ -measurable variable. It remains to verify

$$\mathbb{E}[f(\xi)g(\zeta)U] = \mathbb{E}[\mathbb{E}_{\mathcal{F}}[f(\xi)]\mathbb{E}_{\mathcal{F}}[g(\zeta)]U]. \quad (\text{A.5})$$

Since  $\xi_n \xrightarrow{\mathcal{L}-s} \xi$ ,  $\mathbb{E}[f(\xi_n)\mathbb{E}_{\mathcal{F}}[g(\zeta)]U] \rightarrow \mathbb{E}[f(\xi)\mathbb{E}_{\mathcal{F}}[g(\zeta)]U]$ . By repeated conditioning, we see that the limit coincides with the right-hand side of (A.5). By the assumption on  $\zeta_n$ , we have  $\mathbb{E}_{\mathcal{F}}[g(\zeta_n)] \xrightarrow{\mathbb{P}} \mathbb{E}_{\mathcal{F}}[g(\zeta)]$ . Then, by the bounded convergence theorem,  $\mathbb{E}[f(\xi_n)\mathbb{E}_{\mathcal{F}}[g(\zeta_n)]U] - \mathbb{E}[f(\xi_n)\mathbb{E}_{\mathcal{F}}[g(\zeta)]U] \rightarrow 0$ . Since  $\xi_n$  is  $\mathcal{F}$ -measurable,  $\mathbb{E}[f(\xi_n)\mathbb{E}_{\mathcal{F}}[g(\zeta_n)]U] = \mathbb{E}[f(\xi_n)g(\zeta_n)U]$ . Therefore, the right-hand side of (A.5) is also the limit of  $\mathbb{E}[f(\xi_n)g(\zeta_n)U]$ . But, since  $(\xi_n, \zeta_n) \xrightarrow{\mathcal{L}-s} (\xi, \zeta)$ , we see that  $\mathbb{E}[f(\xi_n)g(\zeta_n)U]$  also converges to the left-hand side of (A.5). Hence, (A.5) must hold. Q.E.D.

**Lemma A.4.** *Let  $\tilde{\beta}_t = (\beta_t^\top, Z_t^\top)^\top$  and let  $f$  be a  $\mathbb{R}^d$ -valued function for some  $d \geq 1$ . Suppose that (i) Assumption H holds for some  $r \in (0, 1)$ ; (ii)  $f \in \mathcal{C}^{2,3}(p)$  for some  $p \geq 3$ ; (iii)  $\varpi \geq$*

$(2p - 1) / 2(2p - r)$ ; (iv) Assumption LW. Then the sequence of variables

$$\Delta_n^{-1/2} \left( \Delta_n \sum_i \left( f(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) - \frac{1}{k_n} \partial_v^2 f(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \widehat{V}_{i\Delta_n}^2 \right) - \int_0^T f(\tilde{\beta}_s, V_s) ds \right) \quad (\text{A.6})$$

converges  $\mathcal{F}$ -stably in law to  $\mathcal{MN}(0, \Sigma_f)$ , where  $\Sigma_f \equiv 2 \int_0^T \partial_v f(\tilde{\beta}_s, V_s) \partial_v f(\tilde{\beta}_s, V_s)^\top V_s^2 ds$ .

PROOF: This lemma generalizes Theorem 3.2 in Jacod and Rosenbaum (2013) by allowing  $f(\cdot)$  to depend on the additional process  $\tilde{\beta}$ . The proof is adapted from Jacod and Rosenbaum (2013). To avoid repetition, we only emphasize the modifications. By localization, we suppose that Assumption SH holds without loss of generality. For notational simplicity, we set, for  $(\beta, z, v) \in \mathcal{B} \times \mathcal{Z} \times \mathcal{V}$ ,

$$h(\beta, z, v) = \partial_v^2 f(\beta, z, v) v^2. \quad (\text{A.7})$$

Recall (A.1). The variable in (A.6) can be decomposed as  $\sum_{j=1}^5 F_{j,n}$ , where

$$\begin{aligned} F_{1,n} &\equiv \Delta_n^{1/2} \sum_i \left( f(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) - f(\tilde{\beta}_{i\Delta_n}, \widehat{V}'_{i\Delta_n}) \right) \\ &\quad - \Delta_n^{1/2} k_n^{-1} \sum_i \left( h(\tilde{\beta}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) - h(\tilde{\beta}_{i\Delta_n}, \widehat{V}'_{i\Delta_n}) \right), \\ F_{2,n} &\equiv \Delta_n^{-1/2} \sum_i \int_{i\Delta_n}^{(i+1)\Delta_n} \left( f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) - f(\tilde{\beta}_s, V_s) \right) ds - \Delta_n^{-1/2} \int_{(N_n+1)\Delta_n}^T f(\tilde{\beta}_s, V_s) ds, \\ F_{3,n} &\equiv \Delta_n^{1/2} \sum_i \partial_v f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) k_n^{-1} \sum_{u=1}^{k_n} (V_{(i+u-1)\Delta_n} - V_{i\Delta_n}), \\ F_{4,n} &\equiv \Delta_n^{1/2} \sum_i \left( f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n} + \tilde{v}_i^n) - f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) \right) \\ &\quad - \partial_v f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) \tilde{v}_i^n - k_n^{-1} h(\tilde{\beta}_{i\Delta_n}, \widehat{V}'_{i\Delta_n}), \\ F_{5,n} &\equiv \Delta_n^{-1/2} k_n^{-1} \sum_i \left( \partial_v f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) \sum_{u=1}^{k_n} ((\Delta_{i+u}^n X')^2 - V_{(i+u-1)\Delta_n} \Delta_n) \right). \end{aligned}$$

The proof will be completed by showing the following claims:

$$\begin{cases} F_{j,n} = o_p(1), & \text{for } j = 1, 2, 3, 4, \\ F_{5,n} \xrightarrow{\mathcal{L}\text{-s}} \mathcal{MN}(0, \Sigma_f). \end{cases} \quad (\text{A.8})$$

We first consider (A.8) for the case with  $j = 1$ . Since  $f \in \mathcal{C}^{2,3}(p)$  and  $\tilde{\beta}_{i\Delta_n}$  is bounded by Assumption SH, we have for all  $i \geq 0$  and  $v \in \mathcal{V}$ ,  $\|\partial_v f(\tilde{\beta}_{i\Delta_n}, v)\| \leq K(1 + v^{p-1})$  and

$\|\partial_v h(\tilde{\beta}_{i\Delta_n}, v)\| \leq K(v + v^{p-1})$ . Hence, by a mean-value expansion,

$$\mathbb{E} \|F_{1,n}\| \leq K\Delta_n^{1/2} \sum_i \mathbb{E} \left[ \left(1 + |\widehat{V}'_{i\Delta_n}|^{p-1} + |\widehat{V}_{i\Delta_n} - \widehat{V}'_{i\Delta_n}|^{p-1}\right) |\widehat{V}_{i\Delta_n} - \widehat{V}'_{i\Delta_n}| \right].$$

As shown in Lemma 4.4 in Jacod and Rosenbaum (2013) (see case  $v = 1$ ), the majorant side of the above inequality can be bounded by  $Ka_n\Delta_n^{(2p-r)\varpi+1/2-p}$  for some deterministic sequence  $a_n \rightarrow 0$ . Since  $\varpi \geq (2p-1)/2(2p-r)$ , we derive (A.8) for  $j = 1$ .

Now, consider (A.8) with  $j = 2$ . Since  $f(\tilde{\beta}_s, V_s)$  is uniformly bounded, it is easy to see that

$$\left\| \Delta_n^{-1/2} \int_{(N_n+1)\Delta_n}^T f(\tilde{\beta}_s, V_s) ds \right\| \leq Kk_n\Delta_n^{1/2} \rightarrow 0,$$

where the convergence is due to Assumption LW. Moreover, the first term in  $F_{2,n}$  is also  $o_p(1)$  due to a standard estimate (see, e.g., p. 153–154 in Jacod and Protter (2012)) for the Riemann approximation error of Itô semimartingales.

Next, consider (A.8) with  $j = 3$ . We set  $\zeta_{3,i}^n \equiv k_n^{-1} \sum_{u=1}^{k_n} (V_{(i+u-1)\Delta_n} - V_{i\Delta_n})$ ,  $\zeta_{3,i}^m \equiv \mathbb{E}[\zeta_{3,i}^n | \mathcal{F}_{i\Delta_n}]$  and  $\zeta_{3,i}^{m,m} \equiv \zeta_{3,i}^n - \zeta_{3,i}^m$ . We then decompose  $F_{3,n} = F'_{3,n} + F''_{3,n}$ , where

$$F'_{3,n} \equiv \Delta_n^{1/2} \sum_i \partial_v f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) \zeta_{3,i}^m, \quad F''_{3,n} \equiv \Delta_n^{1/2} \sum_i \partial_v f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) \zeta_{3,i}^{m,m}.$$

Under Assumption SH, it is easy to see  $\mathbb{E}|\zeta_{3,i}^m| \leq Kk_n\Delta_n$ . Since  $\|\partial_v f(\tilde{\beta}_t, V_t)\|$  is bounded, we further have  $\mathbb{E}\|F'_{3,n}\| \leq Kk_n\Delta_n^{1/2} \rightarrow 0$ . Hence,  $F'_{3,n} = o_p(1)$ . Moreover, by a standard estimate for Itô semimartingales, we have, for any  $u = 1, \dots, k_n$ ,  $\mathbb{E}|V_{(i+u-1)\Delta_n} - V_{i\Delta_n}|^2 \leq Kk_n\Delta_n$ . From here, a use of the Cauchy–Schwarz inequality yields  $\mathbb{E}|\zeta_{3,i}^{m,m}|^2 \leq K\mathbb{E}|\zeta_{3,i}^n|^2 \leq Kk_n\Delta_n$ . By construction,  $\mathbb{E}[\zeta_{3,i}^{m,m} | \mathcal{F}_{i\Delta_n}] = 0$  and  $\zeta_{3,i}^{m,m}$  is  $\mathcal{F}_{(i+k_n-1)\Delta_n}$  measurable. Hence,  $\partial_v f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) \zeta_{3,i}^{m,m}$  and  $\partial_v f(\tilde{\beta}_{l\Delta_n}, V_{l\Delta_n}) \zeta_{3,l}^{m,m}$  are uncorrelated whenever  $|i-l| \geq k_n$ . We then use the Cauchy–Schwarz inequality to derive

$$\mathbb{E}\|F''_{3,n}\|^2 \leq Kk_n\Delta_n \sum_i \mathbb{E} \left\| \partial_v f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) \zeta_{3,i}^{m,m} \right\|^2 \leq Kk_n^2\Delta_n \rightarrow 0,$$

which further implies  $F''_{3,n} = o_p(1)$ . From here, (A.8) with  $j = 3$  readily follows.

To prove (A.8) with  $j = 4$ , we set

$$\begin{aligned} \zeta_{4,i}^m &= \frac{1}{2} \partial_v^2 f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) \left( (\tilde{v}_i^n)^2 - 2k_n^{-1} V_{i\Delta_n}^2 \right), \\ \zeta_{4,i}^{m,m} &= f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n} + \tilde{v}_i^n) - f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) \\ &\quad - \partial_v f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) \tilde{v}_i^n - \frac{1}{2} \partial_v^2 f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) (\tilde{v}_i^n)^2 \end{aligned}$$

$$+k_n^{-1}h(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n}) - k_n^{-1}h(\tilde{\beta}_{i\Delta_n}, \widehat{V}'_{i\Delta_n}).$$

We can then decompose  $F_{4,n} = F'_{4,n} + F''_{4,n}$ , where

$$F'_{4,n} \equiv \Delta_n^{1/2} \sum_i (\mathbb{E} [\zeta_{4,i}^m | \mathcal{F}_{i\Delta_n}] + \zeta_{4,i}^m), \quad F''_{4,n} \equiv \Delta_n^{1/2} \sum_i (\zeta_{4,i}^m - \mathbb{E} [\zeta_{4,i}^m | \mathcal{F}_{i\Delta_n}]).$$

Since  $f \in \mathcal{C}^{2,3}(p)$  and  $\tilde{\beta}_{i\Delta_n}$  is uniformly bounded, we have  $\|\partial_v^2 f(\tilde{\beta}_{i\Delta_n}, v)\| \leq K(1 + v^{p-2})$  and  $\|\partial_v^3 f(\tilde{\beta}_{i\Delta_n}, v)\| \leq K(1 + v^{p-3})$  for all  $v \in \mathcal{V}$ . We can then use the mean-value theorem to derive  $\|\zeta_{4,i}^m\| \leq K(1 + |\tilde{v}_i^n|^{p-3})|\tilde{v}_i^n|^3 + Kk_n^{-1}(1 + |\tilde{v}_i^n|^{p-1})|\tilde{v}_i^n|$ . Now, we can use the same argument in the proof of Lemma 4.4 in Jacod and Rosenbaum (2013) (see case  $v = 4$  there) to derive  $F'_{4,n} = o_p(1)$ . Moreover, note that  $\zeta_{4,i}^m - \mathbb{E}[\zeta_{4,i}^m | \mathcal{F}_{i\Delta_n}]$  and  $\zeta_{4,l}^m - \mathbb{E}[\zeta_{4,l}^m | \mathcal{F}_{l\Delta_n}]$  are uncorrelated whenever  $|i - l| \geq k_n$ , we have  $\mathbb{E}\|F''_{4,n}\|^2 \leq Kk_n\Delta_n \sum_i \mathbb{E}|\zeta_{4,i}^m|^2 \leq Kk_n(k_n^{-2} + k_n\Delta_n)$ , where the first inequality is by the Cauchy-Schwarz inequality and the second inequality is by the second line of (A.2). From here, it follows that  $\mathbb{E}\|F''_{4,n}\|^2 \rightarrow 0$ . Hence,  $F_{4,n} = F'_{4,n} + F''_{4,n} = o_p(1)$ , as claimed in (A.8).

Finally, we notice that the stable convergence in (A.8) follows essentially the same proof as that of Lemma 4.5 in Jacod and Rosenbaum (2013). (To be precise, the only modification needed is to replace the weight  $\partial_{lm}g(c_i^n)$  in their definition of  $V_t^{n,5}$  by  $\partial_v f(\tilde{\beta}_{i\Delta_n}, V_{i\Delta_n})$ .) The proof is now complete. *Q.E.D.*

Now, we are ready to prove Theorem 2.

**PROOF OF THEOREM 2:** (a) We first verify that the conditions in Theorem 1 hold when replacing the function  $g(y, z, v; \theta)$  with  $h(y, z, v; \theta) \equiv \partial_v^2 g(y, z, v; \theta)v^2$ . We define  $\bar{h}(\cdot)$  and  $\bar{h}_{j,p}(\cdot)$  via (2.7) and (3.1) but with  $h$  replacing  $g$ . By Assumption S(iii),  $\bar{h}(\beta, z, v; \theta) = \partial_v^2 \bar{g}(\beta, z, v; \theta)v^2$ . Since  $\bar{g}(\cdot; \theta) \in \mathcal{C}^{2,3}(p)$ , we see that  $\partial_v^2 \bar{g}(\cdot; \theta) \in \mathcal{C}(p-2)$  and, hence,  $\bar{h}(\cdot; \theta) \in \mathcal{C}(p)$ . Further observe that  $\bar{h}_{0,k}(\cdot; \theta) = \bar{g}_{2,k}(\cdot; \theta)v^2$ , which belongs to  $\mathcal{P}(p)$  by Assumption D(i). The condition  $\varpi \geq (2p-1)/2(2p-r)$  clearly implies that  $\varpi \geq (p-1)/(2p-r)$ . Under Assumption LIP(i), it is easy to verify that  $h(\cdot) \in \text{LIP}(p, 0)$ . Now, we can apply Theorem 1 with  $g(\cdot)$  replaced by  $h(\cdot)$  and derive the first assertion of part (a).

As a result,  $\widehat{G}_n(\cdot) - G_n(\cdot) = o_p(1)$  uniformly on compact sets. Under Assumptions S(ii), D(i) and LIP(i), we can apply Theorem 1 to derive that  $\widehat{G}_n(\cdot) - G(\cdot) = o_p(1)$  uniformly on compact sets. From here, the second assertion of part (a) readily follows.

(b) *Step 1.* We outline the proof of part (b) in this step. Without loss of generality, we suppose Assumption SH. To simplify notation, we suppress the appearance of  $\theta^*$  by writing  $g(y, z, v)$  (resp.  $\bar{g}(\beta, z, v)$ ) in place of  $g(y, z, v; \theta^*)$  (resp.  $\bar{g}(\beta, z, v; \theta^*)$ ). We also set  $h(y, z, v) = \partial_v^2 g(y, z, v)v^2$  and  $\bar{h}(\beta, z, v) = \partial_v^2 \bar{g}(\beta, z, v)v^2$ .

The proof relies on the decomposition  $\Delta_n^{-1/2}G_n(\theta^*) = R_{1,n} + R_{2,n} + R_{3,n}$ , where

$$\begin{aligned} R_{1,n} &\equiv \Delta_n^{1/2} \sum_i \left( \bar{g}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}) - k_n^{-1} \bar{h}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \right), \\ R_{2,n} &\equiv \Delta_n^{1/2} k_n^{-1} \sum_i \left( \bar{h}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}) - h(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \right), \\ R_{3,n} &\equiv \Delta_n^{1/2} \sum_i \left( g(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}) - \bar{g}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \right). \end{aligned}$$

By Lemma A.4 with  $f(\cdot) = \bar{g}(\cdot)$ , we have  $R_{1,n} \xrightarrow{\mathcal{L}\text{-s}} \mathcal{MN}(0, \bar{S})$ ; recall (3.5) for the definition of  $\bar{S}$ . Below, we show  $R_{2,n} = o_p(1)$  in step 2. We then show (recalling Definition 2 and (3.4))  $R_{3,n} \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0, \bar{\Gamma})$  in step 4, after preparing some preliminary results in step 3. The assertion of part (b) then follows from Lemma A.3.

*Step 2.* In this step, we show that  $R_{2,n} = o_p(1)$ . By using a componentwise argument, we can assume that  $R_{2,n}$  is scalar without loss of generality. We set  $\tilde{h}_i^n \equiv \bar{h}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}) - h(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n})$  and rewrite  $R_{2,n} = \Delta_n^{1/2} k_n^{-1} \sum_i \tilde{h}_i^n$ . By Assumption S(iii) and the  $\mathcal{F}$ -measurability of  $(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n})$ , we have  $\mathbb{E}_{\mathcal{F}}[\tilde{h}_i^n] = 0$ . Furthermore, since  $\bar{g}_{2,k} \in \mathcal{P}(p-2)$  (Assumption D(i)),

$$\|\tilde{h}_i^n\|_{\mathcal{F},k} \leq K \bar{g}_{2,k}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}) \widehat{V}_{i\Delta_n}^2 \leq K(1 + |\widehat{V}_{i\Delta_n}|^p). \quad (\text{A.9})$$

By Assumption MIX, conditional on  $\mathcal{F}$ , the  $\alpha$ -mixing coefficient of the sequence  $(\tilde{h}_i^n)_{i \geq 0}$  is bounded by  $\alpha_{\text{mix}}(\cdot)$ . Observe that

$$\begin{aligned} \mathbb{E}_{\mathcal{F}}[R_{2,n}^2] &\leq \Delta_n k_n^{-2} \sum_{i,j} \left| \mathbb{E}_{\mathcal{F}}[\tilde{h}_i^n \tilde{h}_j^n] \right| \\ &\leq K \Delta_n k_n^{-2} \sum_{i,j} \alpha_{\text{mix}}(|i-j|)^{1-2/k} \|\tilde{h}_i^n\|_{\mathcal{F},k} \|\tilde{h}_j^n\|_{\mathcal{F},k}, \end{aligned} \quad (\text{A.10})$$

where the first inequality is by the triangle inequality, and the second inequality follows from the mixing inequality. We also note that the condition  $\varpi \geq (2p-1)/2(2p-r)$  implies  $\varpi \geq (2p-1)/(4p-r)$ ; hence, by the third line of (A.2),

$$\mathbb{E}|\widehat{V}_{i\Delta_n}|^{2p} \leq K. \quad (\text{A.11})$$

By (A.9)–(A.11), as well as the assumption that  $\alpha_{\text{mix}}(\cdot)$  has size  $-k/(k-2)$ , we derive  $\mathbb{E}[R_{2,n}^2] \leq K k_n^{-2} \rightarrow 0$ . Therefore,  $R_{2,n} = o_p(1)$  as wanted.

*Step 3.* It remains to show  $R_{3,n} \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0, \bar{\Gamma})$ . By the Cramer–Wold device, we can assume that  $R_{3,n}$  is one-dimensional without loss of generality. In this step, we collect some preliminary



results. For notational simplicity, we set  $\hat{z}_i^n = (\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n})$ ,  $z_i^n = (\beta_{i\Delta_n}, Z_{i\Delta_n}, V_{i\Delta_n})$  and  $\xi_i(\beta, z, v) = g(\mathcal{Y}(\beta, \chi_i), z, v) - \bar{g}(\beta, z, v)$ . We can rewrite

$$R_{3,n} = \Delta_n^{1/2} \sum_i \xi_i(\hat{z}_i^n). \quad (\text{A.12})$$

By (A.2),

$$\mathbb{E}|\widehat{V}_{i\Delta_n} - V_{i\Delta_n}|^2 \leq K\bar{a}_n, \quad \text{where} \quad \bar{a}_n \equiv \Delta_n^{(4-r)\varpi-1} + k_n^{-1} + k_n\Delta_n. \quad (\text{A.13})$$

Since  $\varpi \geq (2p-1)/2(2p-r)$ ,  $r < 1$  and  $p \geq 3$ , it is easy to see  $(4-r)\varpi > 1$ . Hence,  $\bar{a}_n \rightarrow 0$ . Under Assumption SH, we have  $\mathbb{E}\|z_{i-j}^n - z_i^n\|^2 \leq K(1 \wedge j\Delta_n)$  by a standard estimate for Itô semimartingales. By (A.13), we further have  $\mathbb{E}\|\hat{z}_{i-j}^n - \hat{z}_i^n\|^2 \leq K(\bar{a}_n + 1 \wedge j\Delta_n)$ . Recall from Assumption D the constant  $\kappa \in (0, 1]$ . Then we have, by Jensen's inequality,

$$\mathbb{E}\left[\|\hat{z}_{i-j}^n - \hat{z}_i^n\|^{2\kappa}\right] \leq K(\bar{a}_n + 1 \wedge j\Delta_n)^\kappa. \quad (\text{A.14})$$

Observe that, for  $i, j \geq 0$ ,

$$\begin{aligned} & \left| \mathbb{E}_{\mathcal{F}}[\xi_i(\hat{z}_i^n) \xi_{i-j}(\hat{z}_{i-j}^n) - \xi_i(z_i^n) \xi_{i-j}(z_i^n)] \right| \\ & \leq K\alpha_{\text{mix}}(j)^{1-2/k} \|\xi_i(\hat{z}_i^n)\|_{\mathcal{F},k} \|\xi_{i-j}(\hat{z}_{i-j}^n) - \xi_{i-j}(z_i^n)\|_{\mathcal{F},k} \\ & \quad + K\alpha_{\text{mix}}(j)^{1-2/k} \|\xi_{i-j}(z_i^n)\|_{\mathcal{F},k} \|\xi_i(\hat{z}_i^n) - \xi_i(z_i^n)\|_{\mathcal{F},k} \\ & \leq K\alpha_{\text{mix}}(j)^{1-2/k} \left( \bar{g}_{0,k}(\hat{z}_i^n) \rho_k(\hat{z}_{i-j}^n, z_i^n) + \bar{g}_{0,k}(z_i^n) \rho_k(\hat{z}_i^n, z_i^n) \right), \end{aligned} \quad (\text{A.15})$$

where the first inequality is obtained by using the triangle inequality and then the mixing inequality; the second inequality follows from  $\|\xi_i(\cdot)\|_{\mathcal{F},k} \leq K\bar{g}_{0,k}(\cdot)$  and (3.1). Note that Assumption D implies  $\bar{g}_{0,k}(\cdot) \in \mathcal{P}(p/2)$  and  $\rho_k(\hat{z}_{i-j}^n, z_i^n) \leq K(1 + \widehat{V}_{(i-j)\Delta_n}^{p/2}) \|\hat{z}_{i-j}^n - z_i^n\|^\kappa$ . Therefore, (A.15) implies

$$\begin{aligned} & \left| \mathbb{E}_{\mathcal{F}}[\xi_i(\hat{z}_i^n) \xi_{i-j}(\hat{z}_{i-j}^n) - \xi_i(z_i^n) \xi_{i-j}(z_i^n)] \right| \\ & \leq K\alpha_{\text{mix}}(j)^{1-2/k} \left( 1 + \widehat{V}_{i\Delta_n}^{p/2} \right) \left( 1 + \widehat{V}_{(i-j)\Delta_n}^{p/2} \right) \left( \|\hat{z}_{i-j}^n - z_i^n\|^\kappa + \|\hat{z}_i^n - z_i^n\|^\kappa \right). \end{aligned} \quad (\text{A.16})$$

By the Cauchy-Schwarz inequality, (A.11) and (A.14), we further deduce that

$$\mathbb{E} \left| \mathbb{E}_{\mathcal{F}}[\xi_i(\hat{z}_i^n) \xi_{i-j}(\hat{z}_{i-j}^n) - \xi_i(z_i^n) \xi_{i-j}(z_i^n)] \right| \leq K\alpha_{\text{mix}}(j)^{1-2/k} (\bar{a}_n + 1 \wedge j\Delta_n)^{\kappa/2}. \quad (\text{A.17})$$

Next, we set

$$\begin{cases} \Gamma_n \equiv \Delta_n \sum_i \mathbb{E}_{\mathcal{F}} \left[ \xi_i (\hat{z}_i^n)^2 \right] + 2\Delta_n \sum_{j=1}^{N_n} \sum_{i=j}^{N_n} \mathbb{E}_{\mathcal{F}} \left[ \xi_i (\hat{z}_i^n) \xi_{i-j} (\hat{z}_{i-j}^n) \right], \\ \tilde{\Gamma}_n \equiv \Delta_n \sum_i \mathbb{E}_{\mathcal{F}} \left[ \xi_i (z_i^n)^2 \right] + 2\Delta_n \sum_{j=1}^{N_n} \sum_{i=j}^{N_n} \mathbb{E}_{\mathcal{F}} \left[ \xi_i (z_i^n) \xi_{i-j} (z_{i-j}^n) \right]. \end{cases} \quad (\text{A.18})$$

By (A.17) and  $\sum_{l \geq 0} \alpha_{\text{mix}}(l)^{1-2/k} < \infty$ , we deduce

$$\begin{aligned} \mathbb{E} \left| \Gamma_n - \tilde{\Gamma}_n \right| &\leq K \bar{a}_n^{\kappa/2} + K \Delta_n \sum_{j=1}^{N_n} \sum_{i=j}^{N_n} \alpha_{\text{mix}}(j)^{1-2/k} (\bar{a}_n + 1 \wedge j \Delta_n)^{\kappa/2} \\ &\leq K \bar{a}_n^{\kappa/2} + K \Delta_n^{\kappa/2} \sum_{j=1}^{N_n} j^{\kappa/2} \alpha_{\text{mix}}(j)^{1-2/k}. \end{aligned}$$

As mentioned above,  $\bar{a}_n \rightarrow 0$ . Moreover, by Kronecker's lemma,  $\Delta_n^{\kappa/2} \sum_{j=1}^{N_n} j^{\kappa/2} \alpha_{\text{mix}}(j)^{1-2/k} \rightarrow 0$ . Hence,

$$\Gamma_n - \tilde{\Gamma}_n \xrightarrow{\mathbb{P}} 0. \quad (\text{A.19})$$

We now show

$$\tilde{\Gamma}_n \xrightarrow{\mathbb{P}} \bar{\Gamma}. \quad (\text{A.20})$$

To simplify notation, we denote  $\gamma_{l,s} \equiv \gamma_l(\beta_s, Z_s, V_s)$  and  $\bar{\gamma}_s \equiv \bar{\gamma}(\beta_s, Z_s, V_s)$  for  $l \geq 0$  and  $s \geq 0$ . We note that  $\bar{g}(z_i^n) \equiv 0$  because of (2.2). Hence, we can rewrite  $\tilde{\Gamma}_n$  as

$$\tilde{\Gamma}_n = \Delta_n \sum_i \gamma_{0,i\Delta_n} + 2\Delta_n \sum_{j=1}^{\infty} \sum_{i=j}^{N_n} \gamma_{j,i\Delta_n},$$

where empty sums are set to zero by convention. Therefore,

$$\tilde{\Gamma}_n - \bar{\Gamma} = \left( \Delta_n \sum_{i=0}^{N_n} \gamma_{0,i\Delta_n} - \int_0^T \gamma_{0,s} ds \right) + 2 \sum_{j=1}^{\infty} \left( \Delta_n \sum_{i=j}^{N_n} \gamma_{j,i\Delta_n} - \int_0^T \gamma_{j,s} ds \right). \quad (\text{A.21})$$

By an argument similar to (indeed simpler than) (A.16), it is easy to see that  $\gamma_j(\beta, z, v)$  is continuous in  $(\beta, z, v)$ . Under Assumption SH, the process  $(\gamma_{j,t})_{t \geq 0}$  is càdlàg and uniformly bounded. Hence, by invoking the Riemann approximation, we deduce that, for each  $j \geq 0$ ,

$\Delta_n \sum_{i=j}^{N_n} \gamma_{j,i\Delta_n} - \int_0^T \gamma_{j,s} ds \rightarrow 0$ . Moreover, observe that

$$\sum_{j=1}^{\infty} \left| \Delta_n \sum_{i=j}^{N_n} \gamma_{j,i\Delta_n} - \int_0^T \gamma_{j,s} ds \right| \leq K \sup_{t \in [0, T]} \bar{g}_{0,k}(\beta_t, Z_t, V_t)^2 \leq K,$$

where the first inequality is by the mixing inequality and  $\sum_{j \geq 1} \alpha_{\text{mix}}(j)^{1-2/k} < \infty$ , and the second inequality holds because  $(\beta_t^\top, Z_t^\top, V_t^\top)^\top$  is bounded under Assumption SH and  $\bar{g}_{0,k}(\cdot)$  is bounded on bounded sets. This dominance condition allows us to use the dominated convergence theorem to obtain the limit of the right-hand side of (A.21). From here, (A.20) readily follows.

Finally, we note that  $\mathbb{E}_{\mathcal{F}}[R_{3,n}^2] = \Gamma_n$ . Combining (A.19) and (A.20), we derive

$$\mathbb{E}_{\mathcal{F}} [R_{3,n}^2] \xrightarrow{\mathbb{P}} \bar{\Gamma}. \quad (\text{A.22})$$

*Step 4.* We now show that  $R_{3,n} \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0, \bar{\Gamma})$ . Consider a subset  $\bar{\Omega}$  of  $\Omega$  given by  $\bar{\Omega} \equiv \{\bar{\Gamma} > 0\}$  and let  $\bar{\Omega}^c$  be the complement of  $\bar{\Omega}$ . Clearly,  $\bar{\Omega}$  is  $\mathcal{F}$ -measurable. In restriction to  $\bar{\Omega}^c$ ,  $\mathbb{E}_{\mathcal{F}} [R_{3,n}^2] = o_p(1)$  and, thus, the  $\mathcal{F}$ -conditional law of  $R_{3,n}$  converges to the degenerate distribution at zero.

We now restrict attention on the event  $\bar{\Omega}$ , so we can assume  $\bar{\Gamma} > 0$ . We consider an arbitrary subsequence  $\mathbb{N}_1 \subseteq \mathbb{N}$ . By the subsequence characterization of convergence in probability, it is enough to show that there exists a further subsequence  $\mathbb{N}_2 \subseteq \mathbb{N}_1$  such that, as  $n \rightarrow \infty$  along  $\mathbb{N}_2$ , the  $\mathcal{F}$ -conditional distribution function of  $R_{3,n}$  converges uniformly to the  $\mathcal{F}$ -conditional distribution function of  $\mathcal{MN}(0, \bar{\Gamma})$  on  $\mathbb{P}$ -almost every path in  $\bar{\Omega}$ .

By (A.22), we can extract a subsequence  $\mathbb{N}_2 \subseteq \mathbb{N}_1$  such that, along  $\mathbb{N}_2$ ,  $\mathbb{E}_{\mathcal{F}} [R_{3,n}^2] \rightarrow \bar{\Gamma} > 0$  for almost every path in  $\bar{\Omega}$ . Recall from (A.12) that  $R_{3,n} = \Delta_n^{1/2} \sum_i \xi_i(\hat{z}_i^n)$ . Under Assumption MIX,  $\xi_i(\hat{z}_i^n)$  forms a sequence with zero mean and  $\alpha$ -mixing coefficients bounded by  $\alpha_{\text{mix}}(\cdot)$  under the transition probability  $\mathbb{P}^{(1)}$ . Moreover,  $\mathbb{E}|\xi_i(\hat{z}_i^n)|^k \leq K \mathbb{E}|\bar{g}_{0,k}(\hat{z}_i^n)|^k \leq K$ , where the first inequality is by repeated conditioning, Minkowski's and Jensen's inequalities; the second inequality is by  $\bar{g}_{0,k} \in \mathcal{P}(2p/k)$  and (A.11). We are now ready to apply Theorem 5.20 in White (2001) and Pólya's theorem under the transition probability  $\mathbb{P}^{(1)}$  and deduce that, along  $\mathbb{N}_2$ , the  $\mathcal{F}$ -conditional distribution function of  $R_{3,n}$  converges uniformly to the  $\mathcal{F}$ -conditional distribution function of  $\mathcal{MN}(0, \bar{\Gamma})$  for almost every path in  $\bar{\Omega}$ . As mentioned in the previous paragraph, we can use a subsequence argument to further deduce that  $R_{3,n} \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0, \bar{\Gamma})$ . As discussed in step 1, the proof of Theorem 2(b) is now complete. *Q.E.D.*

### A.3 Proof of Theorem 3

PROOF OF THEOREM 3: (a) As is typical in this type of problem, by a polarization argument, we can consider a one-dimensional setting without loss of generality. We henceforth suppose that  $g(\cdot)$  is scalar-valued. By localization, we also suppose that Assumption SH holds.

To simplify notation, we set  $\hat{g}_i^n(\theta) = g(Y_{i\Delta_n}, Z_{i\Delta_n}, \hat{V}_{i\Delta_n}; \theta)$  and  $g_i^n(\theta) = g(Y_{i\Delta_n}, Z_{i\Delta_n}, V_{i\Delta_n}; \theta)$  for  $i \geq 0$  and  $\theta \in \Theta$ . We can rewrite (3.8) as

$$\hat{\Gamma}_n(\hat{\theta}_n) = \Delta_n \sum_{i=0}^{N_n} \hat{g}_i^n(\hat{\theta}_n)^2 + 2 \sum_{j=1}^{m_n} w(j, m_n) \Delta_n \sum_{i=j}^{N_n} \hat{g}_i^n(\hat{\theta}_n) \hat{g}_{i-j}^n(\hat{\theta}_n).$$

We consider a progressive list of approximations to  $\hat{\Gamma}_n(\hat{\theta}_n)$  given by

$$\left\{ \begin{array}{l} \hat{\Gamma}_n^{(1)} \equiv \Delta_n \sum_{i=0}^{N_n} g_i^n(\theta^*)^2 + 2 \sum_{j=1}^{m_n} w(j, m_n) \Delta_n \sum_{i=j}^{N_n} g_i^n(\theta^*) g_{i-j}^n(\theta^*), \\ \hat{\Gamma}_n^{(2)} \equiv \Delta_n \sum_{i=0}^{N_n} \mathbb{E}_{\mathcal{F}} [g_i^n(\theta^*)^2] + 2 \sum_{j=1}^{m_n} w(j, m_n) \Delta_n \sum_{i=j}^{N_n} \mathbb{E}_{\mathcal{F}} [g_i^n(\theta^*) g_{i-j}^n(\theta^*)], \\ \hat{\Gamma}_n^{(3)} \equiv \Delta_n \sum_{i=0}^{N_n} \mathbb{E}_{\mathcal{F}} [g_i^n(\theta^*)^2] + 2 \Delta_n \sum_{j=1}^{m_n} \sum_{i=j}^{N_n} \mathbb{E}_{\mathcal{F}} [g_i^n(\theta^*) g_{i-j}^n(\theta^*)]. \end{array} \right.$$

We note that  $g_i^n(\theta^*)$  is identical to  $\xi_i(z_i^n)$  defined in step 3 of the proof of Theorem 2, because  $\bar{g}(\beta_{i\Delta_n}, Z_{i\Delta_n}, V_{i\Delta_n}; \theta^*) = 0$  as a result of (2.8). Therefore,  $\hat{\Gamma}_n^{(3)}$  has the same form as  $\Gamma_n$  defined in (A.18) after replacing  $\xi_i(\hat{z}_i^n)$  and  $\xi_{i-j}(\hat{z}_{i-j}^n)$  in the latter with  $\xi_i(z_i^n)$  and  $\xi_{i-j}(z_{i-j}^n)$ , respectively. From here, we can use an argument that is similar to that in step 3 of the proof of Theorem 2 to show that  $\hat{\Gamma}_n^{(3)} \xrightarrow{\mathbb{P}} \bar{\Gamma}$ ; this is actually simpler to prove because  $\hat{V}_{i\Delta_n}$  is replaced with the true value  $V_{i\Delta_n}$ . To prove  $\hat{\Gamma}_n(\hat{\theta}_n) \xrightarrow{\mathbb{P}} \bar{\Gamma}$ , it remains to verify  $\hat{\Gamma}_n(\hat{\theta}_n) - \hat{\Gamma}_n^{(1)}$ ,  $\hat{\Gamma}_n^{(1)} - \hat{\Gamma}_n^{(2)}$  and  $\hat{\Gamma}_n^{(2)} - \hat{\Gamma}_n^{(3)}$  are  $o_p(1)$ .

First consider  $\hat{\Gamma}_n(\hat{\theta}_n) - \hat{\Gamma}_n^{(1)}$ . Observe that

$$\begin{aligned} & \left| \hat{\Gamma}_n(\hat{\theta}_n) - \hat{\Gamma}_n^{(1)} \right| \\ & \leq K \sum_{j=0}^{m_n} \Delta_n \sum_{i=j}^{N_n} |\hat{g}_i^n(\hat{\theta}_n) \hat{g}_{i-j}^n(\hat{\theta}_n) - g_i^n(\theta^*) g_{i-j}^n(\theta^*)| \\ & \leq K \sum_{j=0}^{m_n} \Delta_n \sum_{i=j}^{N_n} \left( |\hat{g}_i^n(\hat{\theta}_n)| |\hat{g}_{i-j}^n(\hat{\theta}_n) - g_{i-j}^n(\theta^*)| + |\hat{g}_i^n(\hat{\theta}_n) - g_i^n(\theta^*)| |g_{i-j}^n(\theta^*)| \right) \\ & \leq Km_n \left( \Delta_n \sum_i \left( \hat{g}_i^n(\hat{\theta}_n)^2 + g_i^n(\theta^*)^2 \right) \right)^{1/2} \left( \Delta_n \sum_i |\hat{g}_i^n(\hat{\theta}_n) - g_i^n(\theta^*)|^2 \right)^{1/2}, \end{aligned} \tag{A.23}$$

where the first inequality is from the triangle inequality and the boundedness of the kernel function  $w(\cdot, \cdot)$ ; the second inequality is from the triangle inequality; the third inequality follows from the Cauchy–Schwarz inequality. We further observe that, for  $i \geq 0$ ,

$$\begin{aligned} & \Delta_n \sum_i \left| \hat{g}_i^n(\hat{\theta}_n) - g_i^n(\theta^*) \right|^2 \\ & \leq K \Delta_n \sum_i |\hat{g}_i^n(\hat{\theta}_n) - \hat{g}_i^n(\theta^*)|^2 + K \Delta_n \sum_i |\hat{g}_i^n(\theta^*) - g_i^n(\theta^*)|^2 \\ & \leq K \Delta_n \sum_i B_0(Y_{i\Delta_n}, Z_{i\Delta_n}, \hat{V}_{i\Delta_n})^2 \|\hat{\theta}_n - \theta^*\|^2 + K \Delta_n \sum_i |\hat{g}_i^n(\theta^*) - g_i^n(\theta^*)|^2, \end{aligned} \quad (\text{A.24})$$

where the first inequality is from the triangle inequality and the second inequality follows from Assumption LIP(i) (recall Definition 1). Under Assumption LIP(i), we see  $\mathbb{E}[B_0(Y_{i\Delta_n}, Z_{i\Delta_n}, \hat{V}_{i\Delta_n})^2] = \mathbb{E}[\bar{B}_0(\beta_{i\Delta_n}, Z_{i\Delta_n}, \hat{V}_{i\Delta_n})^2] \leq K$ , where the first inequality is obtained by repeated conditioning, and the second inequality is from  $\bar{B}_0 \in \mathcal{P}(p)$  and (A.11). Since  $\hat{\theta}_n - \theta^* = O_p(\Delta_n^{1/2})$  by assumption,

$$\Delta_n \sum_i B_0(Y_{i\Delta_n}, Z_{i\Delta_n}, \hat{V}_{i\Delta_n})^2 \|\hat{\theta}_n - \theta^*\|^2 = O_p(\Delta_n). \quad (\text{A.25})$$

Moreover, by Assumption D(ii), for each  $i \geq 0$ ,

$$\begin{aligned} \mathbb{E}_{\mathcal{F}} |\hat{g}_i^n(\theta^*) - g_i^n(\theta^*)|^2 & \leq \rho_k \left( (\beta_{i\Delta_n}, Z_{i\Delta_n}, \hat{V}_{i\Delta_n}), (\beta_{i\Delta_n}, Z_{i\Delta_n}, V_{i\Delta_n}) \right)^2 \\ & \leq K(1 + |\hat{V}_{i\Delta_n}|^{p-2}) \left| \hat{V}_{i\Delta_n} - V_{i\Delta_n} \right|^{2\kappa} \\ & \leq K \left( \left| \hat{V}_{i\Delta_n} - V_{i\Delta_n} \right|^{2\kappa} + \left| \hat{V}_{i\Delta_n} - V_{i\Delta_n} \right|^p \right). \end{aligned} \quad (\text{A.26})$$

By (A.2) and Jensen's inequality,

$$\begin{cases} u \geq 1 & \Rightarrow \mathbb{E} |\hat{V}_{i\Delta_n} - V_{i\Delta_n}|^u \leq K_u (\Delta_n^{(2u-r)\varpi+1-u} + k_n^{-u/2} + (k_n \Delta_n)^{(u/2) \wedge 1}), \\ 0 < u < 1 & \Rightarrow \mathbb{E} |\hat{V}_{i\Delta_n} - V_{i\Delta_n}|^u \leq K_u (\Delta_n^{(2-r)\varpi u} + k_n^{-u/2} + (k_n \Delta_n)^{u/2}). \end{cases} \quad (\text{A.27})$$

Note that  $(2p-r)\varpi + 1 - p \geq 1/2$ ,  $(2-r)\varpi \geq 1/4$  and  $\Delta_n^{1/2} \leq K k_n^{-1}$ . Then, by (A.26) and (A.27), we derive

$$\mathbb{E} |\hat{g}_i^n(\theta^*) - g_i^n(\theta^*)|^2 \leq K k_n^{-\kappa}. \quad (\text{A.28})$$

From (A.24), (A.25) and (A.28), we derive

$$\Delta_n \sum_i \left| \hat{g}_i^n(\hat{\theta}_n) - g_i^n(\theta^*) \right|^2 = O_p(k_n^{-\kappa}). \quad (\text{A.29})$$

It is easy to see that  $\mathbb{E}|g_i^n(\theta^*)|^2 \leq K$ . By (A.29), we further have

$$\Delta_n \sum_i \left( |\hat{g}_i^n(\hat{\theta}_n)|^2 + |g_i^n(\theta^*)|^2 \right) = O_p(1). \quad (\text{A.30})$$

Combining (A.23), (A.29) and (A.30), as well as Assumption HAC(ii), we have

$$\left| \hat{\Gamma}_n(\hat{\theta}_n) - \hat{\Gamma}_n^{(1)} \right| = O_p(m_n k_n^{-\kappa/2}) = o_p(1).$$

Next, we consider  $\hat{\Gamma}_n^{(1)} - \hat{\Gamma}_n^{(2)}$ . We denote  $\zeta_{j,i}^n \equiv g_i^n(\theta^*)g_{i-j}^n(\theta^*) - \mathbb{E}_{\mathcal{F}}[g_i^n(\theta^*)g_{i-j}^n(\theta^*)]$  and  $\bar{\zeta}_j^n \equiv \Delta_n \sum_{i=j}^{N_n} \zeta_{j,i}^n$ . We can then rewrite

$$\hat{\Gamma}_n^{(1)} - \hat{\Gamma}_n^{(2)} = \bar{\zeta}_0^n + 2 \sum_{j=1}^{m_n} w(j, m_n) \bar{\zeta}_j^n. \quad (\text{A.31})$$

Note that, conditional on  $\mathcal{F}$ , the sequence  $(g_i^n(\theta^*))_{i \geq 0}$  is  $\alpha$ -mixing with size  $-k/(k-2)$ . By the mixing inequality, for  $i, j \geq 0$  and  $l \geq i$ ,

$$|\mathbb{E}_{\mathcal{F}}[\zeta_{j,i}^n \zeta_{j,l}^n]| \leq K \alpha_{\text{mix}}((l-i-j)^+)^{1-2/k} \|\zeta_{j,i}^n\|_{\mathcal{F},k} \|\zeta_{j,l}^n\|_{\mathcal{F},k},$$

where  $(\cdot)^+$  denotes the positive part. By the Cauchy-Schwarz inequality,

$$\|\zeta_{j,i}^n\|_{\mathcal{F},k} \leq K \bar{g}_{0,2k}(\beta_{i\Delta_n}, Z_{i\Delta_n}, V_{i\Delta_n}; \theta^*) \bar{g}_{0,2k}(\beta_{(i-j)\Delta_n}, Z_{(i-j)\Delta_n}, V_{(i-j)\Delta_n}; \theta^*).$$

Since  $\bar{g}_{0,2k}(\cdot)$  is bounded on bounded set (Assumption HAC(iii)), we further have  $|\mathbb{E}_{\mathcal{F}}[\zeta_{j,i}^n \zeta_{j,l}^n]| \leq K \alpha_{\text{mix}}((l-i-j)^+)^{1-2/k}$ . From here, it follows that

$$\begin{aligned} \mathbb{E}[(\bar{\zeta}_j^n)^2] &\leq 2\Delta_n^2 \sum_{i=j}^{N_n} \sum_{l=i}^{N_n} \mathbb{E}|\mathbb{E}_{\mathcal{F}}[\zeta_{j,i}^n \zeta_{j,l}^n]| \\ &\leq K\Delta_n^2 \sum_{i=j}^{N_n} \sum_{l=i}^{N_n} \alpha_{\text{mix}}((l-i-j)^+)^{1-2/k} \\ &\leq K\Delta_n(j+1). \end{aligned}$$

Then, by the triangle inequality and Jensen's inequality, as well as the boundedness of the kernel function  $w(\cdot, \cdot)$ , we derive from (A.31) that

$$\mathbb{E} \left| \hat{\Gamma}_n^{(1)} - \hat{\Gamma}_n^{(2)} \right| \leq K\Delta_n^{1/2} \sum_{j=0}^{m_n} (j+1)^{1/2} = O(\Delta_n^{1/2} m_n^{3/2}).$$

Since  $m_n = o(k_n^{\kappa/2})$  by Assumption HAC(ii) and  $k_n \leq K\Delta_n^{-1/2}$  by Assumption LW, we have  $m_n = o(\Delta_n^{-1/4})$ . Hence,  $\widehat{\Gamma}_n^{(1)} - \widehat{\Gamma}_n^{(2)} = o_p(1)$ .

Finally, we show that  $\widehat{\Gamma}_n^{(2)} - \widehat{\Gamma}_n^{(3)} = o_p(1)$ . Note that

$$\begin{aligned}\widehat{\Gamma}_n^{(3)} - \widehat{\Gamma}_n^{(2)} &= 2 \sum_{j=m_n+1}^{N_n} \Delta_n \sum_{i=j}^{N_n} \mathbb{E}_{\mathcal{F}} [g_i^n(\theta^*) g_{i-j}^n(\theta^*)] \\ &\quad + 2 \sum_{j=1}^{m_n} (1 - w(j, m_n)) \Delta_n \sum_{i=j}^{N_n} \mathbb{E}_{\mathcal{F}} [g_i^n(\theta^*) g_{i-j}^n(\theta^*)].\end{aligned}$$

Observe

$$\mathbb{E} \left[ \left| 2 \sum_{j=m_n+1}^{N_n} \Delta_n \sum_{i=j}^{N_n} \mathbb{E}_{\mathcal{F}} [g_i^n(\theta^*) g_{i-j}^n(\theta^*)] \right| \right] \leq K \sum_{j=m_n+1}^{N_n} \alpha_{\text{mix}}(j)^{1-2/k} \rightarrow 0,$$

where the inequality is by the triangle inequality and the mixing inequality and the convergence is due to  $\sum_{j \geq 1} \alpha_{\text{mix}}(j)^{1-2/k} < \infty$  and  $m_n \rightarrow \infty$ . Similarly,

$$\mathbb{E} \left| 2 \sum_{j=1}^{m_n} (1 - w(j, m_n)) \Delta_n \sum_{i=j}^{N_n} \mathbb{E}_{\mathcal{F}} [g_i^n(\theta^*) g_{i-j}^n(\theta^*)] \right| \leq K \sum_{j=1}^{m_n} |1 - w(j, m_n)| \alpha_{\text{mix}}(j)^{1-2/k}.$$

Note that for each  $j$ ,  $1 - w(j, m_n) \rightarrow 0$  as  $n \rightarrow \infty$ . Since  $\sum_{j \geq 1} |1 - w(j, m_n)| \alpha_{\text{mix}}(j)^{1-2/k} \leq K \sum_{j \geq 1} \alpha_{\text{mix}}(j)^{1-2/k} < \infty$ , the majorant side of the above inequality converges to zero as  $n \rightarrow \infty$  by the dominated convergence theorem. From here, it follows that  $\widehat{\Gamma}_n^{(2)} - \widehat{\Gamma}_n^{(3)} = o_p(1)$  as claimed.

The proof of part (a) is now complete.

(b) By a polarization argument, we consider the one-dimensional setting without loss of generality. We set

$$\eta_i^n \equiv \partial_v \bar{g}(\beta_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta^*), \quad S_n \equiv 2\Delta_n \sum_i (\eta_i^n)^2 \widehat{V}_{i\Delta_n}^2.$$

Note that  $\bar{g}(\cdot) \in \mathcal{C}^{2,3}(p)$  implies that the function  $(\beta, v, z) \mapsto 2\partial_v \bar{g}(\beta, z, v; \theta^*)^2 v^2$  is in  $\mathcal{C}(2p)$ . Moreover, the condition  $\varpi \geq (2p-1)/2(2p-r)$  implies that  $\varpi \geq (2p-1)/(4p-r)$ . Hence, by applying Lemma A.2 to the function  $(\beta, v, z) \mapsto 2\partial_v \bar{g}(\beta, z, v; \theta^*)^2 v^2$ , we derive  $S_n \xrightarrow{\mathbb{P}} \bar{S}$ . It remains to show that  $\widehat{S}_{1,n}(\hat{\theta}_n) - S_n \xrightarrow{\mathbb{P}} 0$ . Below, we complete the proof by showing  $\widehat{S}_{1,n}(\hat{\theta}_n) - \widehat{S}_{1,n}(\theta^*)$  and  $\widehat{S}_{1,n}(\theta^*) - S_n$  are  $o_p(1)$ .

By the triangle inequality, we see that  $|\widehat{S}_{1,n}(\hat{\theta}_n) - \widehat{S}_{1,n}(\theta^*)| \leq K(SR_{1,n} + SR_{2,n})$ , where

$$SR_{1,n} \equiv \Delta_n \sum_i \left| \hat{\eta}_i^n(\hat{\theta}_n) - \hat{\eta}_i^n(\theta^*) \right|^2 \widehat{V}_{i\Delta_n}^2,$$

$$SR_{2,n} \equiv \Delta_n \sum_i \left| \hat{\eta}_i^n(\hat{\theta}_n) - \hat{\eta}_i^n(\theta^*) \right| |\hat{\eta}_i^n(\theta^*)| \widehat{V}_{i\Delta_n}^2.$$

Recall the notations in Definition 1. By Assumption LIP(i),  $SR_{1,n} \leq \bar{D}_{1,n} \|\hat{\theta}_n - \theta^*\|^2$ , where

$$\bar{D}_{1,n} \equiv \Delta_n \sum_i \left( \frac{1}{k'_n} \sum_{j=0}^{k'_n-1} B_1(Y_{(i+j)\Delta_n}, Z_{(i+j)\Delta_n}, \widehat{V}_{i\Delta_n}) \right)^2 \widehat{V}_{i\Delta_n}^2.$$

Note that  $\mathbb{E}_{\mathcal{F}}[\bar{D}_{1,n}] \leq K\Delta_n \sum_i (1 + \widehat{V}_{i\Delta_n}^{2(p-1)}) \widehat{V}_{i\Delta_n}^2$ . By (A.11),  $\bar{D}_{1,n} = O_p(1)$ . Since  $\hat{\theta}_n - \theta^* = o_p(1)$ ,  $SR_{1,n} = o_p(1)$ . By the Cauchy–Schwarz inequality,  $SR_{2,n} \leq SR_{1,n}^{1/2} \bar{D}_{2,n}^{1/2}$ , where  $\bar{D}_{2,n} \equiv \Delta_n \sum_i \hat{\eta}_i^n(\theta^*)^2 \widehat{V}_{i\Delta_n}^2$ . Note that Assumption AVAR1(i) implies that  $\bar{g}_{1,2}(\cdot; \theta^*) \in \mathcal{P}(p-1)$ . Hence,  $\mathbb{E}_{\mathcal{F}}[\hat{\eta}_i^n(\theta^*)^2] \leq K(1 + |\widehat{V}_{i\Delta_n}|^{2(p-1)})$ . By repeated conditioning and (A.11), we further deduce that  $\bar{D}_{2,n} = O_p(1)$ . Hence,  $SR_{2,n}$  is also  $o_p(1)$ . We have  $\widehat{S}_{1,n}(\hat{\theta}_n) - \widehat{S}_{1,n}(\theta^*) = o_p(1)$  as wanted.

Finally, we show  $\widehat{S}_{1,n}(\theta^*) - S_n \xrightarrow{\mathbb{P}} 0$ . Observe that, by the triangle inequality,

$$\mathbb{E}_{\mathcal{F}} \left| \widehat{S}_{1,n}(\theta^*) - S_n \right| \leq K\Delta_n \sum_i \left( |\eta_i^n| \mathbb{E}_{\mathcal{F}} |\hat{\eta}_i^n(\theta^*) - \eta_i^n| + \mathbb{E}_{\mathcal{F}} |\hat{\eta}_i^n(\theta^*) - \eta_i^n|^2 \right) \widehat{V}_{i\Delta_n}^2. \quad (\text{A.32})$$

We set for each  $i, j \geq 0$ ,

$$\begin{aligned} \zeta_{i,j}^n &\equiv \partial_v g \left( Y_{(i+j)\Delta_n}, Z_{(i+j)\Delta_n}, \widehat{V}_{i\Delta_n}; \theta^* \right) - \partial_v \bar{g} \left( \beta_{(i+j)\Delta_n}, Z_{(i+j)\Delta_n}, \widehat{V}_{i\Delta_n}; \theta^* \right), \\ \bar{\eta}_i^n &\equiv \frac{1}{k'_n} \sum_{j=0}^{k'_n-1} \partial_v \bar{g} \left( \beta_{(i+j)\Delta_n}, Z_{(i+j)\Delta_n}, \widehat{V}_{i\Delta_n}; \theta^* \right). \end{aligned}$$

By Assumption S(iii),  $\mathbb{E}_{\mathcal{F}}[\zeta_{i,j}^n] = 0$ . We can write  $\hat{\eta}_i^n(\theta^*) - \bar{\eta}_i^n = (1/k'_n) \sum_{j=0}^{k'_n-1} \zeta_{i,j}^n$ . Hence,

$$\begin{aligned} \mathbb{E}_{\mathcal{F}} |\hat{\eta}_i^n(\theta^*) - \bar{\eta}_i^n|^2 &\leq \frac{1}{k'_n{}^2} \sum_{j,l=0}^{k'_n-1} |\mathbb{E}_{\mathcal{F}} [\zeta_{i,j}^n \zeta_{i,l}^n]| \\ &\leq K \frac{1}{k'_n{}^2} \sum_{j,l=0}^{k'_n-1} \alpha_{\text{mix}}(|l-j|)^{1-2/k} \left( 1 + \widehat{V}_{i\Delta_n}^{p-2} \right) \\ &\leq K(1 + \widehat{V}_{i\Delta_n}^{p-2})/k'_n, \end{aligned} \quad (\text{A.33})$$

where the first inequality is by the triangle inequality; the second inequality is obtained by first using the mixing inequality and then the assumption that  $\bar{g}_{1,k}(\cdot; \theta^*) \in \mathcal{P}((p-2)/2)$ ; the third inequality is by Assumption MIX.

We set, for each  $i \geq 0$ ,  $D_i^n = (1/k'_n) \sum_{j=0}^{k'_n-1} (\|\beta_{(i+j)\Delta_n} - \beta_{i\Delta_n}\| + \|Z_{(i+j)\Delta_n} - Z_{i\Delta_n}\|)$ . Note that,



by a mean-value expansion and the assumption that  $\partial_\beta \partial_v \bar{g}(\cdot; \theta^*)$  and  $\partial_z \partial_v \bar{g}(\cdot; \theta^*)$  are in  $\mathcal{P}(p/2 - 1)$ ,

$$|\bar{\eta}_i^n - \eta_i^n| \leq K(1 + \widehat{V}_{i\Delta_n}^{p/2-1})D_i^n. \quad (\text{A.34})$$

Since  $\bar{\eta}_i^n$  and  $\eta_i^n$  are  $\mathcal{F}$ -measurable, we combine (A.33) and (A.34) to derive

$$\mathbb{E}_{\mathcal{F}} |\hat{\eta}_i^n(\theta^*) - \eta_i^n|^2 \leq K(1 + \widehat{V}_{i\Delta_n}^{p-2})((D_i^n)^2 + 1/k'_n). \quad (\text{A.35})$$

Next, note that under Assumptions S(iii) and AVAR1(i),  $\partial_v \bar{g}(\cdot; \theta^*) \in \mathcal{P}(p/2 - 1)$ . By (A.32) and (A.35),

$$\mathbb{E}_{\mathcal{F}} \left| \widehat{S}_{1,n}(\theta^*) - S_n \right| \leq K\Delta_n \sum_i \left( 1/\sqrt{k'_n} + D_i^n + (D_i^n)^2 \right) \left( 1 + \widehat{V}_{i\Delta_n}^p \right). \quad (\text{A.36})$$

Further note that  $\mathbb{E}|D_i^n|^2 + \mathbb{E}|D_i^n|^4 \leq Kk'_n\Delta_n$ . Hence, by the Cauchy–Schwarz inequality, (A.36) and (A.11), we derive  $\mathbb{E}|\widehat{S}_{1,n}(\theta^*) - S_n| \leq K(\sqrt{k'_n\Delta_n} + 1/\sqrt{k'_n}) \rightarrow 0$ . Hence,  $\widehat{S}_{1,n}(\theta^*) - S_n = o_p(1)$ . The proof of part (b) is now complete.

(c) Denote  $f(z, v; \theta) \equiv 2\tilde{\varphi}(z, v; \theta^*)\tilde{\varphi}(z, v; \theta^*)^\top v^2$ . Since  $\tilde{\varphi}(\cdot; \theta^*) \in \mathcal{C}(p-1)$ ,  $f \in \mathcal{C}(2p)$ . By Lemma A.2,  $\widehat{S}_{2,n}(\theta^*) \xrightarrow{\mathbb{P}} \bar{S}$ . Since  $\tilde{\varphi}(\cdot; \theta^*) \in \mathcal{C}(p-1)$  and  $\tilde{\varphi}(\cdot) \in \text{LIP}(p-1, 0)$ , it is easy to see that  $\sup_{\theta \in \Theta_0} \|\tilde{\varphi}(\cdot; \theta)\| \in \mathcal{P}(p-1)$  for any compact subset  $\Theta_0$  that contains  $\theta_0$ . Hence, with probability approaching one,

$$\begin{aligned} & \left| \widehat{S}_{2,n}(\hat{\theta}_n) - \widehat{S}_{2,n}(\theta^*) \right| \\ & \leq K\Delta_n \sum_i (1 + \widehat{V}_{i\Delta_n}^{p-1}) \left\| \tilde{\varphi}(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \hat{\theta}_n) - \tilde{\varphi}(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta^*) \right\| \widehat{V}_{i\Delta_n}^2 \\ & \leq K\Delta_n \sum_i \left( 1 + \widehat{V}_{i\Delta_n}^{2p} \right) \left\| \hat{\theta}_n - \theta^* \right\|. \end{aligned}$$

By (A.11) and  $\hat{\theta}_n - \theta^* = o_p(1)$ , we see  $\widehat{S}_{2,n}(\hat{\theta}_n) - \widehat{S}_{2,n}(\theta^*) = o_p(1)$ . From here, the assertion in part (c) readily follows. *Q.E.D.*

**PROOF OF COROLLARY 2:** Part (a) follows from Theorems 2, 3 and the continuous mapping theorem. To show part (b), we first show that  $cv_{n,1-\alpha}(\theta^*) \xrightarrow{\mathbb{P}} cv_{1-\alpha}$ . Fix an arbitrary subsequence  $\mathbb{N}_1 \subseteq \mathbb{N}$ . By Theorem 3,  $\widehat{\Sigma}_{g,n}(\theta^*) \xrightarrow{\mathbb{P}} \Sigma_g$  and, hence, there exists a further subsequence  $\mathbb{N}_2 \subseteq \mathbb{N}_1$ , along which  $\widehat{\Sigma}_{g,n}(\theta^*) \xrightarrow{a.s.} \Sigma_g$ . Consider a path  $\omega \in \Omega$  on which  $L(\cdot, \cdot)$  is continuous at  $\Sigma_g$  and  $\widehat{\Sigma}_{g,n}(\theta^*) \rightarrow \Sigma_g$  holds along  $\mathbb{N}_2$ ; such paths form a  $\mathbb{P}$ -full event. By the continuous mapping theorem, on path  $\omega$ , the  $\mathcal{F} \otimes \mathcal{G}$ -conditional distribution function of  $L(\widehat{\Sigma}_{g,n}(\theta^*)^{1/2}U, \widehat{\Sigma}_{g,n}(\theta^*))$  converges weakly to the  $\mathcal{F}$ -conditional distribution of  $L(\xi, \Sigma_g)$ . By assumption,  $1 - \alpha$  is a continuity point of the  $\mathcal{F}$ -conditional quantile function of  $L(\xi, \Sigma_g)$ . Hence, on path  $\omega$ , we have  $cv_{n,1-\alpha} \rightarrow cv_{1-\alpha}$  along  $\mathbb{N}_2$ . By a subsequence characterization of convergence in probability, we deduce that  $cv_{n,1-\alpha}(\theta^*) \xrightarrow{\mathbb{P}} cv_{1-\alpha}$ . This result, combined with that in part (a), implies  $\mathbb{P}(L_n(\theta^*) \leq cv_{n,1-\alpha}(\theta^*)) \rightarrow 1 - \alpha$ . *Q.E.D.*

## A.4 Proof of Propositions 1 and 2

PROOF OF PROPOSITION 1: (a) Let  $Q(\theta) \equiv G(\theta)^\top \Xi G(\theta)$ . By Theorem 2(a),  $G_n(\cdot) \xrightarrow{\mathbb{P}} G(\cdot)$  and, hence,  $Q_n(\cdot) \xrightarrow{\mathbb{P}} Q(\cdot)$  uniformly over  $\Theta$ . It is easy to see from Assumption LIP(i) that  $G(\cdot)$  is continuous and so is  $Q(\cdot)$ . Under Assumption GMIM,  $Q(\cdot)$  is uniquely minimized at  $\theta^*$ . Since  $\Theta$  is compact,  $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta^*$  follows from a standard argument (see, e.g., Theorem 2.1 in Newey and McFadden (1994)).

(b) Under Assumptions S and D, for each  $\theta \in \Theta$ , the functions  $(\beta, z, v) \mapsto \partial_\theta g(\beta, z, v; \theta)$  and  $(\beta, z, v) \mapsto \partial_\theta \partial_v^2 g(\beta, z, v; \theta) v^2$  satisfy condition (ii) of Theorem 1. Moreover, by Assumption LIP(ii), the functions  $\partial_\theta g(y, z, v; \theta)$  and  $\partial_\theta \partial_v^2 g(y, z, v; \theta) v^2$  belong to  $\text{LIP}(p, 0)$ . By Theorem 1, we have,

$$\partial_\theta G_n(\theta) \xrightarrow{\mathbb{P}} \int_0^T \partial_\theta \bar{g}(\beta_s, Z_s, V_s; \theta) ds, \quad \text{uniformly in } \theta \in \Theta. \quad (\text{A.37})$$

In particular, for any sequence  $\tilde{\theta}_n$  that satisfies  $\tilde{\theta}_n \xrightarrow{\mathbb{P}} \theta^*$ , we have  $\partial_\theta G_n(\tilde{\theta}_n) = H + o_p(1)$ . Then, under Assumption GMIM, a routine manipulation yields,

$$\Delta_n^{-1/2} \left( \hat{\theta}_n - \theta^* \right) = - (H^\top \Xi H)^{-1} H^\top \Xi \Delta_n^{-1/2} G_n(\theta^*) + o_p(1). \quad (\text{A.38})$$

The assertion then follows from Theorem 2(b).

(c) By (A.37),  $H_n \xrightarrow{\mathbb{P}} H$ . Since  $\Delta_n^{-1/2}(\hat{\theta}_n - \theta^*) = O_p(1)$  from part (b), the assertion of part (c) readily follows from Theorem 3. *Q.E.D.*

PROOF OF PROPOSITION 2: Denote  $A \equiv (I_q - H (H^\top \Xi H)^{-1} H^\top \Xi) \Sigma_g^{1/2}$ . Observe that

$$\begin{aligned} \Delta_n^{-1/2} G_n(\hat{\theta}_n) &= \Delta_n^{-1/2} G_n(\theta^*) + H \Delta_n^{-1/2} \left( \hat{\theta}_n - \theta^* \right) + o_p(1) \\ &= A \Sigma_g^{-1/2} \Delta_n^{-1/2} G_n(\theta^*) + o_p(1), \end{aligned}$$

where the first equality is by a mean-value expansion,  $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta^*$  and the uniform convergence given by (A.37); the second equality is obtained by using the asymptotic linear representation (A.38). Note that  $\Sigma_g^{-1/2} \Delta_n^{-1/2} G_n(\theta^*) \xrightarrow{\mathcal{L}-s} \mathcal{N}(0, I_q)$  by Theorem 2. It is also straightforward to show that  $A^\top \Xi A$  is idempotent with rank  $q - \dim(\theta)$ . The assertion of the proposition readily follows. *Q.E.D.*