Inference on Conditional Mean Models in Continuous Time[1]

Joon Y. Park
Department of Economics
Indiana University and SKKU

Abstract

In the paper, we develop a general methodology to estimate and test for a conditional mean model given in continuous time. Our model specifies the conditional mean of instantaneous change of a given stochastic process as a function of other covariates. The model yields a continuous time regression for the instantaneous change of an underlying process on its conditional mean change with the error process given by a general martingale. We call it a martingale regression, since the parameter in the model is identified by the residual process being a martingale. Upon an appropriate time change, the martingale regression can always be transformed into a regression with the error process given by Brownian motion. We use this property and apply a minimum distance method to estimate the parameters in the model. More specifically, the samples are collected at random time intervals so that the errors become independent normals, and the estimates are defined as the parameter values which make the empirical distribution of the residuals closest to independent and identically distributed normals. We show by simulation that our approach yields a reliable method of inference for a broad class of continuous time models.

Preliminary and incomplete. Comments are welcome.

This Version: November 2009

Key words and phrases: Martingale, time change, continuous time model, conditional mean specification, interest parity, predictive regression, diffusion models.

---

[1] I am very grateful to Yoosoon Chang and Hwagyun Kim for many helpful discussions and comments, and to Daehee Jeong, Minsoo Jeong and Minchul Shin for their excellent research assistance. This paper was prepared for an invited lecture at the 2007 Far Eastern Econometric Society Meeting, Taipei, Taiwan. I gratefully acknowledge the financial supports from the NSF under NSF Grant No. SES-0518619.

1. Introduction

This paper develops a general methodology for the statistical inference in a conditional mean model given in continuous time. Our model specifies the instantaneous rate of change in the conditional mean of a given stochastic process as a parametric function of some covariate process. As a result, it yields a continuous time regression model with a general martingale error process. The model is called the martingale regression, since it is identified by the condition that the error process is a martingale. Our model is quite general. In particular, it does not impose any restriction on the error process, allowing for a variety of conditional volatilities that are time-varying and stochastic. Our methodology is therefore applicable for a wide range of continuous time models that are used in the fields of economics and finance. General continuous-time asset pricing models given in parametric form may be well fitted into our framework. Diffusions with a parametric specification of drift function are also considered as a special case of our model, so all our results can be applied to them as well. Our approach distinguishes itself from the conventional approach in that it relies on the martingale condition for identification. The reader is referred to Bergstrom (1984) for a survey of conventional continuous time models used in econometrics.

It has been the usual practice to analyze continuous time models by applying high frequency data directly on the discretized versions of the models. However, the direct use of high frequency data on the discretized models to do inference for the underlying continuous time models is not desirable for several reasons. First, on the high frequency domain, the error process generating volatility dominates the conditional mean process of interest in many economic and financial models. The information in the sample on the conditional mean is therefore severely contaminated by the volatility component, when the sampling interval is very small. Second, the distributions of errors in many models are changing over time especially at high frequencies and very far away from being normal, due in particular to the presence of time-varying and stochastic volatilities that are often quite persistent and strongly endogenous. Consequently, the usual statistical theory relying on asymptotic normality is generally not applicable, which would invalidate the use of the standard inference in such models. Finally, we cannot generally identify conditional mean models in continuous time by the usual orthogonality condition between the errors and covariates. The orthogonality condition exploited in the conventional GMM approach holds for continuous time models, if approximated by the Euler scheme. However, it may yield substantial discretization bias.

In this paper, we propose an approach to more effectively deal with the martingale regression, i.e., the general conditional mean model given in continuous time. Our methodology is based on the use of a time change, which transforms a given martingale into Brownian motion. Indeed, it is well known that any continuous martingale becomes Brownian motion, if its sample path is read using a clock running at the speed inversely proportional to the rate of increase in the quadratic variation. This is a consequence of the celebrated theorem by Dambis, Dubins and Schwarz, which is often referred to as the DDS theorem. It has already been used in various contexts of econometric and statistical researches by several authors. Yu and Phillips (2001) exploited the DDS theorem to estimate the linear drift in diffusion models based on the Gaussian likelihood. The martingale and semimartingale tests by Park and Vasudev (2006) and Peters and de Vilder (2006) also rely on the same idea. Moreover, Andersen, Bollerslev and Dobrev (2007) used the time change given by the DDS theorem in testing the adequacy of jump-diffusion models for return distribution, and Jacewitz and Park (2009) recently employed it to allow for general stochastic volatilities in their study of the predictive regressions. Chang (2008) also used a closely related approach to invent a Gaussian panel unit root test.

We use the idea of time change to conveniently identify and estimate the martingale regression. As we mentioned earlier, the martingale regression is identified by the condition that the error process is a martingale. Unfortunately, the martingale condition for identification is very difficult to implement. If the error process is continuous, however, we may invoke the DDS theorem to identify the model after time change by the condition that the error process is Brownian motion. Needless to say, the Brownian motion condition for identification is much easier to invoke, using its Gaussianity and independent increment properties. Indeed, our estimate of the unknown parameter is defined to be the value that yields the time changed error process mostly closely follow Brownian motion. More precisely, we obtain the estimate by minimizing the Crämer-von Mises distance between the empirical distributions of the increments of error process in the time changed regression and the corresponding distributions of Brownian increments. It is shown in the paper that the estimator is consistent and asymptotically normal under suitable regularity conditions. The asymptotic variance can be estimated by the block bootstrap or sub-sampling.

In the actual implementation of our methodology, we of course have to rely on discrete samples. We assume that the observations are available at relatively high frequencies such as daily. Our methodology uses the observations at two different levels of frequencies. First, we use all available observations to estimate the time change required to identify the martingale regression by the error process being Brownian motion, instead of a general martingale process. Second, we collect the samples at a constant incremental level of the estimated quadratic variation of the error process, and use them to estimate the unknown parameter of the model by the minimum distance method. For instance, we may use the daily observations to estimate the required time change, and then obtain the samples at the average monthly increments of the quadratic variation of the error process to estimate the unknown parameters in the model. Note that we need to collect samples at random intervals in this step. Our asymptotics require that the sampling frequency is small and the sampling horizon is large. Provided in our analysis are sufficient conditions we need to ensure that the errors incurred by using discrete samples become negligible.

We perform a set of simulations to evaluate the performance of our methodology in finite samples. As is well known, the inference on conditional mean models in continuous time is mainly affected by the sampling horizon instead of the number of observations in the sample. The overall performances of our martingale estimator (MGE) and the t-ratio based on our approach are quite good, even when the sampling horizon is only modest. Across many different data generating processes, the MGE yields negligible biases and the actual probabilities of the MGE t-test are quite close to its nominal size. In contrast, the OLS methodology applied to the discretized versions of our simulation models in continuous time yield nonnegligible biases for the OLS estimators and the rejection probabilities deviating significantly from the nominal values of the standard t-tests. This is so, except for a few unrealistic cases that we consider for the purpose of comparison. Even at daily frequencies, the discretization error can be so big as the actual rejection probabilities are as high as unity. The distortion becomes worse if we introduce time-varying or stochastic volatilities in the models.

The rest of the paper is organized as follows. In Section 2, we present the model and main ideas. The conditional mean model in continuous time is introduced and the main ideas that are heavily used in the subsequent development of our methodology are explained in detail. Our martingale estimator is also defined. The inferential problems in our approach are addressed in Section 3. There we lay out how our continuous-time methodology may be implemented in practice using discrete-time observations. The feasible martingale estimator is considered and its asymptotics are developed. In particular, we show that both the infeasible and feasible martingale estimators have normal asymptotic distributions under appropriate regularity conditions. The bootstrap methods to estimate the asymptotic variance of the estimators is also discussed. Section 4 reports our simulation results. For the models specified with several different types of stochastic volatilities, we compare our approach with the OLS procedure applied to the models with the conventional discretization. The simulation results highlight the discretization bias and the distortions made by the presence of time-varying or stochastic volatilities. We conclude the paper in Section 5. Mathematical proofs are collected in Mathematical Appendix.

2. The Model and Main Ideas

2.1 Martingale Regression

Many economic and financial models can be specified in continuous time as

$$\mathbb{E}(dY_t|\mathcal{F}_t) = \mu(X_t, \theta_0)dt, \tag{1}$$

where (Y_t) and (X_t) are stochastic processes, $(\mathcal{F}_t)$ is a filtration to which both (Y_t) and (X_t) are adapted, and μ is a known function defined on $\mathbb{R} \times \Theta$ with parameter set $\Theta \subset \mathbb{R}^m$ and parameter vector $\theta_0 \in \Theta$. In the paper, we will call μ the instantaneous conditional mean function. Clearly, we may rewrite (1) as a continuous time regression

$$dY_t = \mu(X_t, \theta_0)dt + dU_t, \tag{2}$$

where (U_t) is a martingale with respect to the filtration $(\mathcal{F}_t)$, so that $\mathbb{E}(dU_t|\mathcal{F}_t) = 0$.

Over an interval $[t, t+\delta]$ for any $t > 0$ and small $\delta > 0$, we have

$$\mathbb{E}(Y_{t+\delta} - Y_t|\mathcal{F}_t) \approx \delta\mu(X_t, \theta_0)$$

if $(\mu(X_t, \theta_0))$ is continuous a.s. in time $t > 0$. Therefore, $(\mu(X_t, \theta_0))$ generally represents the rate of instantaneous change in conditional mean of (Y_t), given as a function of (X_t), which is assumed to be known up to the unknown parameter $\theta_0 \in \Theta$. For a variety of models that are commonly used in economics and finance, (Y_t) is specified as the logs of asset prices or foreign exchange rates. In this case, $Y_{t+\delta} - Y_t$ denotes the returns from holding the assets or foreign exchanges over the interval $[t, t+\delta]$. Correspondingly, (dY_t) represents their instantaneous returns at time $t > 0$.

It is important to note that the parameter $\theta_0 \in \Theta$ is identified in our model by the martingale condition. That is, if we set

$$U_t(\theta) = (Y_t - Y_0) - \int_0^t \mu(X_s, \theta) ds, \tag{3}$$

then θ_0 is the value of $\theta \in \Theta$ such that $(U_t(\theta))$ is a martingale. For this reason, we call regression (2) the *martingale regression*. To achieve identification of the martingale regression, we will assume that

Assumption 2.1 No distinctive values of $\theta \in \Theta$ yield the stochastic processes $(\mu(X_t, \theta))$, which are the same version.

Assumption 2.1 implies that the processes $(\mu(X_t, \theta))$ with different values of $\theta \in \Theta$ have all distinctive finite sample distributions, and allows us to identify $\theta_0 \in \theta$ in (2) uniquely by the martingale condition.

We let the error process (U_t) in (2) be a general martingale process, and do not need to impose any restrictive conditions. For the expositional convenience, however, we assume that

Assumption 2.2 The error process (U_t) has a.s. continuous sample path.

Assumption 2.2 is not essential, and can be relaxed. It is introduced here to convey our main ideas more effectively. In particular, it is possible to allow for the presence of jumps in the error process, as we will discuss later. For all illustrative examples that we are considering in the project, we indeed presume that there are jumps in the error process.

Under Assumption 2.2, we may write

$$dU_t = \sigma_t dW_t, \tag{4}$$

where (σ_t) is adapted to $(\mathcal{F}_t)$ and (W_t) is the standard Brownian motion with respect to $(\mathcal{F}_t)$. We leave the specification of (σ_t) in (4) totally unrestricted, and therefore, our model may have an arbitrary type of time-varying and stochastic heterogeneity. Our model for the instantaneous conditional mean changes is thus truly general.

Within the conventional framework, all parametric asset pricing models derived under no arbitrage condition in continuous time yield the continuous time regression that we specify in (2). To see this more clearly, we let (P_t) be the price of a financial asset, and let (π_t), $\pi_t = \exp(-\int_0^t r_s^f ds) D_t$, be the state-price deflator, where (r_t^f) is the risk-free rate and (D_t) is the Radon-Nykodym derivative of the equivalent martingale measure with respect to the true probability. Under no arbitrage condition, we have $\mathbb{E}(dP_t^\pi | \mathcal{F}_t) = \mathbb{E}(d(\pi_t P_t) | \mathcal{F}_t) = 0$, i.e.,

$$\mathbb{E}\left(\frac{d\pi_t}{\pi_t}\middle| \mathcal{F}_t\right) + \mathbb{E}\left(\frac{dP_t}{P_t}\middle| \mathcal{F}_t\right) + \frac{d\pi_t}{\pi_t}\frac{dP_t}{P_t} = 0,$$

and it follows from $\mathbb{E}((d\pi_t/\pi_t)|\mathcal{F}_t) = -r_t^f dt$ that $\mathbb{E}((dP_t/P_t)|\mathcal{F}_t) = r_t^f dt - (d\pi_t/\pi_t)(dP_t/P_t)$. Consequently, we have

$$\frac{dP_t}{P_t} - r_t^f = -\frac{d\pi_t}{\pi_t}\frac{dP_t}{P_t} + dU_t,$$

where (U_t) is a martingale. Both the asset price and the state-price deflator processes are conventionally specified as Ito processes, in which case $-(d\pi_t/\pi_t)(dP_t/P_t)$ reduces to the dt term in (2) with $\mu(X_t, \theta_0)$ given by the product of the volatility functions of (dP_t/P_t) and $(d\pi_t/\pi_t)$, if they are parametrized by $\theta\Theta$ with the true value θ_0.

It is easy to see that our model (2) includes as a special case with $dY_t = dX_t$ the diffusion model given by

$$dX_t = \mu_t dt + \sigma_t dW_t, \tag{5}$$

where μ_t and σ_t are referred respectively to as drift and diffusion terms. If we set $\mu_t = \mu(X_t, \theta_0)$ using some known function μ and unknown parameter $\theta_0 \in \Theta$, the diffusion model in (5) clearly reduces to our model in (2). The specification of diffusion term is totally unrestricted and left to be completely general. The most commonly used specification of drift term is a linear drift, which is given by $\mu(X_t, \theta) = \beta(\alpha - X_t)$ for $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}_{++}$. Respectively for the specifications of diffusion term as $\sigma_t = \sigma$ for $\sigma \in \mathbb{R}_{++}$ and $\sigma_t = \omega\sqrt{X_t}$ for $\omega \in \mathbb{R}_{++}$, we have Ornstein-Uhlenbeck process and Feller's square-root process. The diffusion term is often specified also as $\sigma_t = \omega|X_t|^\rho$ for $\omega \in \mathbb{R}_{++}$ and $\rho \in \mathbb{R}_+$, which is referred to as the constant elasticity of variance (CEV) diffusion. Our methodology developed subsequently in the paper allows us to estimate the drift function without specifying the functional form for the diffusion term.

2.2 Time Change

We define a time change, i.e., a non-decreasing collection of stopping times, (T_t) by

$$T_t = \inf_{s>0}\{[U]_s > t\},$$

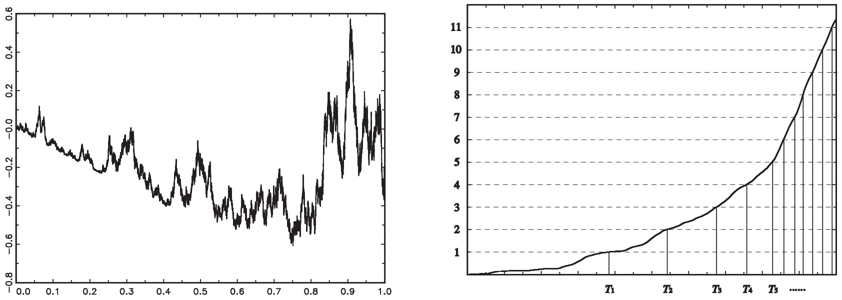
where $([U]_t)$ is the quadratic variation of (U_t). Then, as is well known, we have

$$U_{T_t} = V_t \qquad \text{or} \qquad U_t = V_{[U]_t},$$

where (V_t) is the standard Brownian motion, which is commonly called the DDS (Dambis, Dubins-Schwarz) Brownian motion of (U_t). This result, which is often referred to as the DDS theorem, plays the central role in the subsequent development of our methodology and theory. See Revuz and Yor (1994) for more details.

Roughly put, the DDS theorem implies that all continuous martingales are essentially Brownian motion with differences only in their quadratic variations, and all continuous martingales become Brownian motion if their sample paths are read using a clock running at the speed set inversely to the rate of increase in their quadratic variations. This idea was explored earlier by Park and Vasudev (2006) to develop a test for martingale in continuous time. In Figure 1 below, we will present two illustrative examples to show how we may define the time change (T_t) to convert general continuous martingales to Brownian motion.

Compensated Square Brownian Motion: $W_t^2 - t$



Exponential Brownian Motion: $\exp(W_t - t/2)$

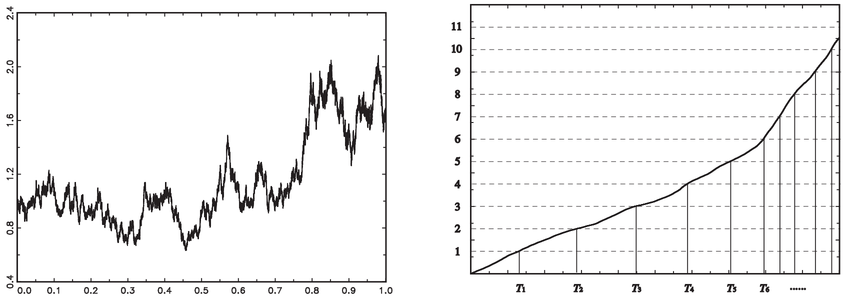


Figure 1: Sample Paths and Quadratic Variations with Time Changes

There we consider two continuous martingales, the compensated squared Brownian motion given by $U_t = W_t^2 - t$ and the exponential Brownian motion given by $U_t = \exp(W_t - t/2)$, where (W_t) is the standard Brownian motion.

The time change (T_t) is very useful for the estimation of and testing on the martingale regression (2). With the time change (T_t), we have

$$dY_{T_t} = \mu(X_{T_t}, \theta_0)dT_t + dU_{T_t} = \mu(X_{T_t}, \theta_0)dT_t + dV_t. \tag{6}$$

Note that the error process (V_t) in the model with time change (T_t) is the standard Brownian motion. This is in contrast with our original model, where the error process (U_t) is a general martingale process. The parameter $\theta_0 \in \Theta$ is identified in the time-changed martingale regression (6) by the condition

$$V_t(\theta) = (Y_{T_t} - Y_0) - \int_0^{T_t} \mu(X_s, \theta)ds \tag{7}$$

is the standard Brownian motion if and only if $\theta = \theta_0 \in \Theta$. Recall that $\theta = \theta_0 \in \Theta$ is assumed to be the only parameter value for which $(U_t(\theta))$, defined in (3), becomes a martingale.

2.3 Martingale Estimator

Now we introduce our estimator, which will be called the *martingale estimator* (MGE), for the unknown parameter $\theta \in \Theta$ in (2). We fix $\Delta > 0$ and define

$$Z_i(\theta) = \Delta^{-1/2} \left(Y_{T_{i\Delta}} - Y_{T_{(i-1)\Delta}} - \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \mu(X_t, \theta) dt \right) \tag{8}$$

for $i = 1, \ldots, N$. Note that $(Z_i(\theta))$ are the normalized increments of the error process $(V_t(\theta))$ from the time-changed martingale regression defined in (7), which are collected at time intervals of length Δ, where $\Delta > 0$ is some constant. Or equivalently, we may see that $(Z_i(\theta))$ are defined to be the normalized increments of the error process $(U_t(\theta))$ of our model in (3), collected at random time intervals given by $(T_{i\Delta})$. The choice of Δ will be discussed below.

To introduce our estimator, we let i_d denote the d-dimensional vector index running from i, and define $Z_{i_d}(\theta) = (Z_i(\theta), \ldots, Z_{i-d+1}(\theta))'$. Moreover, assuming that $(Z_i(\theta))$ is strictly stationary, we signify for each $\theta \in \Theta$ by $\Pi(\cdot, \theta)$ and $\Pi_N(\cdot, \theta)$ respectively the joint distribution function and empirical distribution function of (Z_{i_d}). Consequently, we have

$$\Pi_N(z, \theta) = \frac{1}{N} \sum_{i=1}^{N} 1\{Z_{i_d}(\theta) \leq z\} = \frac{1}{N} \sum_{i=1}^{N} 1\{Z_i(\theta) \leq z_1\} \cdots 1\{Z_{i-d+1}(\theta) \leq z_d\} \tag{9}$$

for $z = (z_j) \in \mathbb{R}^d$. Here and elsewhere indicator functions with vector arguments are defined as the product of indicators with their scalar components. Note that $\Pi(\cdot, \theta_0)$, which we will also write as $\Pi_0(\cdot)$, reduces to the d-dimensional multivariate standard normal distribution function. Therefore, we have

$$\Pi(z, \theta_0) \equiv \Pi_0(z) = \Phi(z_1) \cdots \Phi(z_d),$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Our estimator $\hat{\theta}_N$, the d-dimensional MGE, of the parameter $\theta \in \Theta$ is defined as

$$\hat{\theta}_N = \operatorname*{argmin}_{\theta \in \Theta} Q_N(\theta),$$

where

$$Q_N(\theta) = \int [\Pi_N(z, \theta) - \Pi(z, \theta_0)]^2 \varpi(dz) \tag{10}$$

with any bounded measure $\varpi(\cdot)$ on $\mathbb{R}^d$. This type of minimum distance estimator was defined earlier by Manski (1983). The objective function $Q_N(\theta)$ in (10) becomes

$$Q(\theta) = \int [\Pi(z, \theta) - \Pi(z, \theta_0)]^2 \varpi(dz)$$

in the limit as $N \to \infty$.

The natural choice of ϖ is the measure given by the distribution function $\Pi(\cdot, \theta_0)$. In this case, we do not need any numerical integration to obtain the MGE $\hat{\theta}_N$. Indeed, the objective function $Q_N(\theta)$ can be readily evaluated using simple algebraic computational procedures for each $\theta \in \Theta$. To show this more explicitly, fix $\theta \in \Theta$ and let $(z_{(i)})$ be the observed values of $(Z_i(\theta))$ arranged in the ascending order, i.e., $z_{(1)} < \cdots < z_{(N)}$, and define $w_i = \Phi(z_{(i)})$, where Φ is the standard normal distribution function as defined earlier. Then the value of the objective function Q_N is given by

$$Q_N = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{2i-1}{2N} - w_i\right)^2 + \frac{1}{12N^2}$$

for the 1-dimensional MGE, and

$$Q_N = \frac{1}{N^2}\sum_{i,j=2}^{N}(1 - w_i \vee w_j)(1 - w_{i-1} \vee w_{j-1}) - \frac{1}{2N}\sum_{k=2}^{N}(1 - w_k^2)(1 - w_{k-1}^2) + \frac{1}{9}$$

for the 2-dimensional MGE, where $p \vee q = \max(p, q)$.

We may check the adequacy of specification for our model using

$$\tau_N = NQ_N(\hat{\theta}_N)$$

as a test statistic. Under the correct specification, it follows that $\hat{\theta}_N \approx \theta_0$ and $\Pi_N(\cdot, \hat{\theta}_N) \approx \Pi(\cdot, \theta_0)$, and therefore, $Q_N(\hat{\theta}_N) \approx 0$ for large N. Indeed, as we show later in the paper, we have $Q(\hat{\theta}_N) = O_p(N^{-1})$ and the statistic τ_N has a proper limit distribution as $N \to \infty$ under the correct specification. Of course, this would not be so, if the underlying model is misspecified. If there is no value of $\theta \in \Theta$ for which the error process $(U_t(\theta))$ in (3) is a martingale, then the time changed error process $(V_t(\theta))$ in (7) does not become a Brownian motion for any value of $\theta \in \Theta$. It therefore follows that the limit of $Q_N(\theta)$ does not vanish for any value of $\theta \in \Theta$, so we would have in particular that $Q_N(\hat{\theta}_N) \not\to_p 0$. Consequently, we would have $\tau_N \to_p \infty$ under misspecification, and the test becomes consistent if we reject the null of correct specification when τ_N takes large values.

Both the finite sample performance and the limit distribution of the MGE depend on the choice of Δ. Note that both $\Pi_N(\cdot, \theta)$ and $\Pi(\cdot, \theta)$ depend on Δ though we suppress it for the sake of notational brevity. Moreover, for a given T, the choice of Δ determines the size N of the normalized increments of the error process $(Z_i(\theta))$, which affects the behavior of the MGE directly. In general, we may expect that the distribution $\Pi(\cdot, \theta)$ of $(Z_i(\theta))$ departs more sharply from standard normal as θ takes values away from θ_0 for larger values of Δ. This is because the conditional mean component $\mu(X_t, \theta_0)dt$ becomes more important than the error component dU_t in our model (2) as Δ increases. Therefore, all other things being equal, the MGE would have a smaller variance for a larger value of Δ. On the other hand, the marginal effect of Δ via the size N of the normalized increments of the error process is the opposite. For a fixed T, N decreases as Δ increases, which would make the variance of the MGE larger. Indeed, we may find an optimal choice of Δ for some simple cases, as will be discussed in more detail later. However, we assume at the moment that Δ is just a constant fixed a priori.

3. Statistical Procedure and Asymptotic Theory

3.1 Estimation of Time Change

To implement our methodology, we need to estimate the time change (T_t). Suppose that we have n-observations on (Y_t) and (X_t) with sampling interval $\delta > 0$, which are denoted by

$$(X_1, Y_1), \ldots, (X_{i\delta}, Y_{i\delta}), \ldots, (X_{n\delta}, Y_{n\delta}). \tag{11}$$

with the initial value (X_0, Y_0). Throughout the paper, we let $T = n\delta$ denote the sampling horizon. Note that the conditional mean component $\mu(X_t, \theta_0)dt$ in our model (2) is of bounded variation, whose quadratic variation vanishes at all $t \geq 0$. Therefore, we have $[U]_t = [Y]_t$ for all $t \geq 0$, which can be estimated by

$$[Y]_t^\delta = \sum_{i\delta \leq t} (Y_{i\delta} - Y_{(i-1)\delta})^2.$$

For the hypothesis testing, we may also impose the null value θ_0 of $\theta \in \Theta$, so that $[U]_t^\delta = \sum_{i\delta \leq t} (U_{i\delta} - U_{(i-1)\delta})^2$ can be obtained directly under the null hypothesis.

In the subsequent development of our theory, we require $\delta \to 0$ fast enough so that $([Y]_t^\delta)$ becomes a consistent estimate for $([U]_t)$ over the entire sampling horizon. Below we provide some simple sufficient conditions for the uniform consistency of both $([U]_t^\delta)$ and $([Y]_t^\delta)$. For the expositional brevity, we assume throughout that the observations are made over equi-spaced sampling interval. The extension to allow for irregular and random sampling is possible, as long as the modulus of the sampling interval is small and decreases down to zero.

Assumption 3.1 For all $0 \leq s \leq t \leq T$,

$$a_T(t-s) \leq [U]_t - [U]_s \leq b_T(t-s),$$

where $a_T, b_T > 0$ are some constants depending only upon T.

Assumption 3.1 is satisfied for a large class of continuous martingales. For the Brownian motion, we may easily deduce that the condition holds with $a_T = b_T = 1$. More generally, the condition holds for any martingale (M_t), defined as $dM_t = \sigma_t dW_t$ for some volatility process (σ_t) and Brownian motion (W_t), if we have

$$a_T \leq \inf_{0 \leq t \leq T} \sigma_t^2 \qquad \text{and} \qquad \sup_{0 \leq t \leq T} \sigma_t^2 \leq b_T.$$

Many diffusion processes satisfy this condition.

Lemma 3.1 *Under* Assumptions 2.2 *and* 3.1, *we have*

$$\mathbb{E}\left(\sup_{0 \leq t \leq T} \left| [U]_t^\delta - [U]_t \right| \right) = O\left(b_T (\delta T)^{1/2} \right)$$

for b_T *introduced in* Assumption 3.1.

Therefore, the estimated quadratic variation ($[U]_t^\delta$) obtained from the discrete samples is uniformly consistent for the true quadratic variation ($[U]_t$), as long as $b_T^2(\delta T) \to 0$. Note that the longer horizon the data spans (i.e., as T becomes larger), we need to observe them more frequently (i.e., δ should be smaller) to ensure the uniform consistency of the estimated quadratic variation over an expanded time interval $[0, T]$. This is more so, if the quadratic variation increases more sharply (i.e., b_T increases faster).

Assumption 3.2 We assume that

$$\mathbb{E}\left(\sup_{0\le t\le T}\sup_{\theta\in\Theta}|\mu(X_t,\theta)|\right)^4 = O(c_T^2)$$

for all large T.

Assumption 3.2 specifies the maximal growth rate of $(\mu(X_t, \theta))$ over $t \in [0, T]$ and $\theta \in \Theta$. If the instantaneous conditional mean function μ is bounded, then Assumption 3.2 is trivially satisfied with $c_T = 1$. More generally, we let (a) $\sup_{\theta\in\Theta}|\mu(x,\theta)| \le c|x|^p$ for some constants $c > 0$ and $p \ge 0$, and let (b) $\mathbb{E}\bigl(\sup_{0\le t\le T}|X_t|\bigr)^r = O(T^q)$ for some $r \ge 4p$ and $q \ge 0$. Then we have

$$\begin{aligned}\mathbb{E}\left(\sup_{0\le t\le T}\sup_{\theta\in\Theta}|\mu(X_t,\theta)|\right)^4 &\le c^4\mathbb{E}\left(\sup_{0\le t\le T}|X_t|\right)^{4p}\\ &\le c^4\left[\mathbb{E}\left(\sup_{0\le t\le T}|X_t|\right)^r\right]^{4p/r} = O(T^{4pq/r}),\end{aligned}$$

and therefore, Assumption 3.2 holds with $c_T = T^{2pq/r}$.

Condition (a) is not stringent and met for virtually all instantaneous conditional mean functions used in empirical applications. Condition (b) is also not restrictive and satisfied by all stochastic processes commonly used in practice. Many processes that are used in practical applications, such as interest rates, certain growth rates and various financial ratios, have natural boundaries, and consequently, the condition holds with $q = 0$ for any value of $r \ge 0$. In this case, we would therefore have $c_T = 1$ in Assumption 3.2 for all values of p. Stationary Gaussian processes satisfy the condition with any $q > 0$ for any choice of $r \ge 4p$ if some mild additional requirements are met. The reader is referred to Berman (1992) for more details. For instance, it is well known that the running maximum $\sup_{0\le t\le T}|X_t|$ of Ornstein-Uhlenbeck process has any integral moments and grows at the rate of $(\log T)^{1/2}$. Therefore, Assumption 3.2 is satisfied with $c_T = T^\varepsilon$ for any $\varepsilon > 0$, regardless of the value of p. Finally, the condition holds for Brownian motion (W_t) with $q = r/2$, since

$$T^{-1/2}\sup_{0\le t\le T}|W_t| = \sup_{0\le t\le 1}T^{-1/2}|W_{Tt}| =_d \sup_{0\le t\le 1}|W_t|$$

due to the scaling property of Brownian motion. Note that $\sup_{0\leq t\leq 1}|W_t|$ is distributed as the modulus of standard normal, and therefore, has the infinite number of moments. See, e.g., Revuz and Yor (1991, p.19). As a result, Assumption 3.2 holds with $c_T = T^p$. It can be readily seen that the condition is also satisfied for Brownian motion with drift if we set $q = r$, which yields Assumption 3.2 with $c_T = T^{2p}$.

Lemma 3.2 *Under* Assumptions 2.2 *and* 3.1 - 3.2*, we have*

$$\mathbb{E}\left(\sup_{0\leq t\leq T}\left|[Y]_t^\delta - [U]_t^\delta\right|\right)^2 = O\left(c_T^2(\delta T)^2\right) + O\left((b_T c_T)\delta T^2\right)$$

with b_T and c_T introduced in Assumptions 3.1 *and* 3.2.

As in Lemma 3.1, we require in Lemma 3.2 that $\delta \to 0$ as $T \to \infty$. Lemma 3.2, together with Lemma 3.1, establishes the uniform consistency of $([Y]_t^\delta)$ for $([U]_t)$. In particular, it allows us to use $([Y]_t^\delta)$ to estimate the time change (T_t). Naturally, the required condition in Lemma 3.2 is more stringent than the one in Lemma 3.1, since (Y_t) has the conditional mean component and we need to ensure that its contribution to the estimation of quadratic variation is asymptotically negligible. In particular, Lemma 3.2 makes it necessary that δ decreases at least faster than T^2 increases.[2]

Now we define

$$R_{\delta,T} = \sup_{0\leq t\leq T}\left|[Y]_t^\delta - [U]_t\right|. \tag{12}$$

Then it follows that

$$\mathbb{E}R_{\delta,T}^2 = O\left(b_T(\delta T)^{1/2}\right) + O\left(c_T^2(\delta T)^2\right) + O\left((b_T c_T)\delta T^2\right) \tag{13}$$

from Lemmas 3.1 and 3.2. If we let $S = [U]_T$, we may readily deduce that

Corollary 3.3 *Under* Assumptions 2.2 *and* 3.1 - 3.2*, we have*

$$\mathbb{E}\left(\sup_{0\leq t\leq S}\left|T_t^\delta - T_t\right|\right)^2 = O\left(\frac{b_T}{a_T^2}(\delta T)^{1/2}\right) + O\left(\frac{c_T^2}{a_T^2}(\delta T)^2\right) + O\left(\frac{b_T c_T}{a_T^2}(\delta T^2)\right),$$

with a_T, b_T and c_T introduced in Assumptions 3.1 *and* 3.2.

The time change based on realized variance of (Y_t) may therefore be used instead of the required theoretical time change, if $\delta \to 0$ sufficiently faster than $T \to \infty$.

---

[2] We may use the fitted residuals $\hat{U}_{i\delta} - \hat{U}_{(i-1)\delta} = (Y_{i\delta} - Y_{(i-1)\delta}) - \mu(X_{(i-1)\delta}, \hat{\theta})\delta$, $i = 1, \ldots, n$, obtained using any consistent estimator $\hat{\theta}$ of θ_0 to estimate the quadratic variation of (U_t). Our subsequent results would then hold under less stringent conditions.

3.2 Feasible Martingale Estimator

Let

$$Z_i^\delta(\theta) = \Delta^{-1/2}\left(Y_{T_{i\Delta}^\delta} - Y_{T_{(i-1)\Delta}^\delta} - \int_{T_{(i-1)\Delta}^\delta}^{T_{i\Delta}^\delta} \mu(X_t,\theta)dt\right) \tag{14}$$

for $i = 1,\ldots,N$. The feasible MGE is based on $(Z_i(\theta)^\delta)$ in (14), in place of $(Z_i(\theta))$ introduced earlier in (8). Note that we construct the samples $(Z_i(\theta)^\delta)$ of size N, from the observations $(X_{i\delta}, Y_{i\delta})$ of size n on the underlying stochastic processes (X,Y) given in (11), to estimate the parameter θ in the model. Throughout the paper, we call $(X_{i\delta}, Y_{i\delta})$ and $(Z_i(\theta)^\delta)$, respectively, the *original* and *estimation* samples to avoid confusion. Of course, we have $n > N$.

Given our results in the previous subsection, we may well expect that the estimation samples $(Z_i(\theta)^\delta)$ get close to $(Z_i(\theta))$ as $\delta \to 0$ for each $i = 1,\ldots,N$. Furthermore, they are expected to be close to each other uniformly in $i = 1,\ldots,N$, if $\delta \to 0$ sufficiently fast relative to $T \to \infty$. Below we present the exact condition that we need to require to make the error in approximating $(Z_i(\theta))$ by $(Z_i(\theta)^\delta)$ uniformly negligible for $i = 1,\ldots,N$ for all large N. In what follows, we use the notation $p \wedge q = \min(p,q)$.

Assumption 3.3 We assume $N = N_1 \wedge N_2 \to \infty$, where

$$N_1 = o\left(\frac{a_T^2}{b_T c_T}(\delta T)^{-1/2} \wedge \frac{a_T^2}{c_T^3}(\delta T)^{-2} \wedge \frac{a_T^2}{b_T c_T^2}(\delta T^2)^{-1}\right)$$

and that, for some $\varepsilon > 0$,

$$N_2 = O\left(\frac{\delta^{-1/4+\varepsilon}T^{-1/4}}{b_T^{1/2}} \wedge \frac{\delta^{-1+\varepsilon}T^{-1}}{c_T} \wedge \frac{\delta^{-1/2+\varepsilon}T^{-1}}{(b_T c_T)^{1/2}}\right)$$

for all small δ and large T.

Lemma 3.4 *Under* Assumptions 2.2 *and* 3.1 - 3.3*, we have*

$$\mathbb{E} \sup_{1\le i\le N} \sup_{\theta\in\theta} \left| Z_i^\delta(\theta) - Z_i(\theta) \right| = o(N^{-1/2})$$

for all large N.

Lemma 3.4 requires that we have at least $\delta T^2 \to 0$, as in Lemma 3.2. We may set N larger as $\delta \to 0$ faster, compared to the rate $T \to \infty$. The resulting approximation error in $(Z_i(\theta)^\delta)$ would then become smaller.

Using the estimation samples $(Z_i^\delta(\theta))$ in (14), we now define

$$\Pi_N^\delta(z,\theta) = \frac{1}{N}\sum_{i=1}^N 1\{Z_{i_d}^\delta(\theta) \le z\} = \frac{1}{N}\sum_{i=1}^N 1\{Z_i^\delta(\theta) \le z_1\}\cdots 1\{Z_{i-d+1}^\delta(\theta) \le z_d\}. \tag{15}$$

Clearly Π_N^δ in (15) corresponds to Π_N introduced in (9).

Assumption 3.4 We assume that the conditional distribution of $Z_i(\theta)$ on $(Z_i^\delta(\theta)-Z_i(\theta))$ is absolutely continuous respect to Lebesgue measure having density bounded uniformly in $1 \leq i \leq N$ and $\theta \in \Theta$.

Lemma 3.5 *Under* Assumptions 2.2 *and* 3.1 - 3.4 *we have*

$$\mathbb{E} \sup_{z \in \mathbb{R}^d} \sup_{\theta \in \Theta} \left| \Pi_N^\delta(z,\theta) - \Pi_N(z,\theta) \right| = o(N^{-1/2})$$

for all large N.

The result in Lemma 3.5 follows immediately from Lemma 3.4 under Assumption 3.4. Though it is difficult to check, the condition in Assumption 3.4 does not seem to be overly stringent.

The feasible MGE is given by

$$\hat{\theta}_N^\delta = \operatorname*{argmin}_{\theta \in \Theta} Q_N^\delta(\theta),$$

where

$$Q_N^\delta(\theta) = \int [\Pi_N^\delta(z,\theta) - \Pi(z,\theta_0)]^2 \varpi(dz), \tag{16}$$

which is defined correspondingly with Q_N in (10). It is well expected from Lemma 3.5 that the feasible MGE would behave similarly as $\hat{\theta}_N$ in asymptotics under suitable regularity conditions. We will show below that this is indeed true.

3.3 Asymptotic Theory for Martingale Estimator

The following are a set of sufficient conditions we impose to obtain the limit distribution of our martingale estimator $\hat{\theta}_N$.

Assumption 3.5 We assume that
(a) For all $\theta \in \Theta$, $(Z_i(\theta))$ is strictly stationary and α-mixing with the mixing coefficient $\alpha(k) = O(k^{-c})$ for some $c > (2d+1)(4d-1)$.
(b) For all $\theta \in \Theta$ near θ_0, we have $|\mu(x,\theta) - \mu(x,\theta_0)| \leq \nu(x)\|\theta - \theta_0\|$ for a measurable real-valued function ν on $\mathbb{R}$. Moreover, we let

$$Z_i = U_{T_{i\Delta}} - U_{T_{(i-1)\Delta}} \qquad \text{and} \qquad W_i = \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \nu(X_t)dt,$$

and assume that the conditional distribution of Z_i on W_i is absolutely continuous with respect to Lebesgue measure having density bounded uniformly in $i \geq 1$, and that $\sup_{i \geq 1} \mathbb{E} W_i^2 < \infty$.

We have

Lemma 3.6 *Under Assumption* 3.5, *we have*

$$\sqrt{N}[\Pi_N(z,\theta) - \Pi(z,\theta)]$$

is stochastically equicontinuous at $\theta_0 \in \Theta$ *with respect to the Euclidean metric on* $\Theta \subset \mathbb{R}^m$ *for all* $z \in \mathbb{R}^d$.

Assumption 3.6 We assume that
(a) The parameter space Θ is compact, and θ_0 is an interior point of Θ.
(b) The function $\Pi(\cdot,\theta)$ of θ is differentiable at θ_0 in $\mathcal{L}^2(\varpi)$, i.e., there exists $\dot{\Pi} \in \mathcal{L}^2(\varpi)$ such that

$$\int \left(\frac{\Pi(z,\theta) - \Pi(z,\theta_0) - (\theta-\theta_0)'\dot{\Pi}(z)}{\|\theta-\theta_0\|} \right)^2 \varpi(dz) \to 0$$

as $\|\theta - \theta_0\| \to 0$, where $\mathcal{L}^2(\varpi)$ denotes the Hilbert space of functions that are square integrable with respect to measure ϖ.
(c) The function Q has a positive second derivative matrix $\ddot{Q}$ at θ_0.

Under Assumptions 2.1 - 2.2 and 3.5 - 3.6, we will establish that

$$\sqrt{N}(\hat{\theta}_N - \theta_0) = -2\ddot{Q}(\theta_0)^{-1}\sqrt{N}\int \dot{\Pi}(z)[\Pi_N(z,\theta_0) - \Pi(z,\theta_0)]\varpi(dz) + o_p(1) \tag{17}$$

for all large N. Moreover, by the functional central limit theory given by, e.g., Deo (1978), we have

$$\sqrt{N}[\Pi_N(\cdot,\theta_0) - \Pi(\cdot,\theta_0)] \to_d \Lambda(\cdot) \tag{18}$$

as $N \to \infty$, where Λ is the Gaussian process with covariance kernel $\Sigma(x,y) = \mathbb{E}\Lambda(x)\Lambda(y)$. Note that $\Pi_N(\cdot,\theta_0)$ is the empirical process defined from d-dimensional multivariate normal samples that are $(d-1)$-dependent.

To define $\Sigma(x,y)$ more explicitly, we let $x = (x_j) \in \mathbb{R}^d$ and $y = (y_j) \in \mathbb{R}^d$, and define for $|k| \le d-1$

$$\Gamma_k(x,y) = \mathbb{E}[1\{Z_{i_d} \le x\} - \Pi_0(x)][1\{Z_{(i-k)_d} \le y\} - \Pi_0(y)]$$

with $Z_{i_d} = Z_{i_d}(\theta_0)$. The covariance kernel of the Gaussian process Λ is then given by

$$\Sigma(x,y) = \sum_{|k|\le d-1} \Gamma_k(x,y)$$

for $x, y \in \mathbb{R}^d$. We may easily see that

$$\Gamma_0(x,y) = \Phi(x_1 \wedge y_1)\cdots\Phi(x_d \wedge y_d) - \Phi(x_1)\cdots\Phi(x_d)\Phi(y_1)\cdots\Phi(y_d)$$

and, for $1 \le k \le d-1$,

$$\begin{aligned}\Gamma_k(x,y) =& \Phi(x_1)\cdots\Phi(x_k)\Big[\Phi(x_{k+1}\wedge y_1)\cdots\Phi(x_d \wedge y_{d-k}) \\ & - \Phi(x_{k+1})\cdots\Phi(x_d)\Phi(y_1)\Phi(y_{d-k})\Big]\Phi(y_{d-k+1})\cdots\Phi(y_d).\end{aligned}$$

Furthermore, it follows that $\Gamma_{-k}(x,y) = \Gamma_k(y,x)$.

It is easy to deduce from (17) and (18) that

Theorem 3.7 *Under* Assumptions 2.1 - 2.2 *and* 3.1 - 3.6, *we have*

$$\hat{\theta}_N^\delta = \hat{\theta}_N + o_p(N^{-1/2})$$

for all large N, *and*

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \to_d \mathbb{N}(0, 4\Omega),$$

where $\Omega = \ddot{Q}(\theta_0)^{-1} P \ddot{Q}(\theta_0)^{-1}$ *with*

$$P = \int \int \dot{\Pi}(x)\Sigma(x,y)\dot{\Pi}(y)'\varpi(dx)\varpi(dy)$$

as $N \to \infty$.

Theorem 3.7 establishes the asymptotic normality of the MGE. The proof of the theorem relies on the results in Andrews and Pollard (1994), Wegkamp (1999) and Brown and Wegkamp (2002). If $\Pi(z,\cdot)$ is twice differentiable, then we may expect to have

$$\ddot{Q}(\theta_0) = 2\int \dot{\Pi}(z)\dot{\Pi}(z)'\varpi(dz)$$

under suitable conditions required to interchange the order of differentiation and integration. In general, the asymptotic variance Ω can be estimated by the usual resampling methods such as sub-sampling and bootstrapping.

The usual subsampling method can be directly applicable for our model to estimate the asymptotic variance Ω. For the bootstrap method, the most natural way to implement it in our framework is to use a block bootstrap and resample from the pairs $(X_{i\delta}, Y_{i\delta})$, say, $(X_{i\delta}^*, Y_{i\delta}^*)$ for $i = 1, \ldots, n$. It is important to resample the pairs to preserve the dependency between X and Y. Of course, we need to introduce some additional conditions on (X, Y), such as stationarity and strong geometric mixing conditions, to make the block bootstrap valid. See, e.g., Horowitz (2001). Those conditions are not very stringent, since all stationary diffusion processes are strongly geometrically mixing. Under the required extra conditions for the validity of the block bootstrap, we may expect the block bootstrap to be consistent. If we denote by $\hat{\theta}_N^{\delta *}$ the MGE obtained from the bootstrap samples, then we may indeed follow Brown and Wegkamp (2002) to show that the conditional distribution of $\sqrt{N}(\hat{\theta}_N^{\delta *} - \hat{\theta}_N^\delta)$ consistently estimates the distribution of $\sqrt{N}(\hat{\theta}_N^\delta - \theta_0)$ in probability.[3] We may therefore use the bootstrap sample variance of $\sqrt{N}(\hat{\theta}_N^{\delta *} - \hat{\theta}_N^\delta)$ as a consistent estimate for the asymptotic variance Ω.

Subsampling or bootstrapping entire samples can be computationally burdensome. It is unnecessary, if the conditional mean model is linear in parameter and given by $\mu(x,\theta) = \theta_0'\nu(x)$. In this case, we may directly resample

$$\left(Y_{T_{i\Delta}^\delta} - Y_{T_{(i-1)\Delta}^\delta}, \int_{T_{(i-1)\Delta}^\delta}^{T_{i\Delta}^\delta} \nu(X_t)dt \right)$$

---

[3] They consider only i.i.d. case under simpler conditions. However, their main arguments for the bootstrap consistency readily extends to our framework given the validity of block bootstrap and all our previous results in the paper.

using the subsampling or block bootstrap method. The dimension of the resampling is now significantly reduced from n to N. Moreover, the steps to obtain the MGE using the subsamples or bootstrap samples become greatly simplified. In particular, it is not necessary to re-estimate the time change and collect the samples at the random intervals given by the time change. The computational burden of bootstrap is therefore minimal in this case. We use this approach in our simulation reported in the next section.

3.4 Other Issues in Implementation

Obviously, our results here are not applicable if the error process has jumps. The DDS theorem holds only for continuous martingales, and therefore, our methodology relying on the theorem breaks down. Note that we may allow jumps in (X, Y) as long as they are synchronized and do not disturb the relationship between X and Y in any discrete fashion, since we only require the continuity of sample paths for the error process. There are several ways to avoid the discontinuity in the error process. First, we may use samples collected at relatively lower frequency. It is well observed in many economic and financial time series that jumps are relatively rare for samples at the frequencies of daily or lower, though they are frequently observed for many intra-day samples. Second, as a preliminary step, we may detect the presence of jumps using a test by, e.g., Lee and Mykland (2007). Once we identify the locations of jumps, we may simply discard the observations corresponding to (dY_t) and $(\mu(X_t, \theta)dt)$ before we implement the martingale methodology. The subsequent analysis, of course, will be made conditional on the preliminary jump test, which is not entirely satisfactory. However, as long as the number of jumps is not substantial, the procedure should not have any major impact on the performance of our MGE.

Our result in Theorem 3.7 allows us to find an asymptotic optimal choice of Δ, if the distribution of (X, Y) is known and (X, Y) is continuously observed. Indeed, for the simplest case of Ornstein-Uhlenbeck diffusion $dX_t = \kappa(\mu - X_t)dt + \sigma dW_t$, the value of Δ which minimizes the variance of $\hat{\theta}_N$ is given by $\Delta^* = 2.15\sigma^2/\kappa$, and the corresponding size of the estimation sample becomes $N^* = \kappa T/2.15$, if the sample path of the process is continuously observable.[4] With the choice of optimal Δ^* or N^*, the MGE has the asymptotic variance $1.54\sqrt{2\kappa}$. As is well known, the MLE of κ has the asymptotic variance $\sqrt{2\kappa}$. The MGE has the asymptotic variance that is 1.54 times bigger than the MLE. This, of course, does not imply that the finite sample behavior of the MLE is better than that of the MGE. Indeed, as we show in our simulations, the MGE performs well and often better than the MLE in finite samples. However, we expect that the MLE outperforms the MGE if T becomes large because it uses more information.

It seems very difficult to find the optimal value of Δ or N in more general models. Furthermore, if we consider $\hat{\theta}_N^\delta$ based on the discrete observations on (X, Y), it would be impossible to obtain the analytical solution for an optimal choice of Δ. In this case, we may also take into consideration the errors in approximating the required time change $(T_{i\Delta})$ by its estimate $(T_{i\Delta}^\delta)$. Clearly, the relative magnitude of the error incurred in this approximation becomes smaller as Δ gets large, since we have a greater number of observations in the

---

[4] The optimal value of Δ for the Ornstein-Uhlenbeck process reported here is made available by Minchul Shin, to whom I am very grateful.

original sample for each interval $[T^\delta_{(i-1)\Delta}, T^\delta_{i\Delta}]$. Of course, we may try various values of Δ and find an optimal choice numerically. It will be explained in detail how we may do this in a later section. In case we need to compare our methodology with other competing approaches based on the fixed sampling schemes, we may also set Δ comparably to them so that they have the same number of observations. This can be done by dividing $[U]^\delta_T$ by a fixed value of N to find the corresponding level of Δ. For instance, setting N to be the number of months in the sampling horizon, we may obtain $\Delta = (1/N)[U]^\delta_T$ corresponding to the monthly observations.

4. Illustrations and Simulations

In this section, we provide some empirical illustrations on how to implement our methodology in practical applications, and present some simulation results to evaluate the finite sample performance of the MGE. For our purpose, we consider two continuous time models. The first model, Model I, is the Feller square-root process given by

$$dX_t = (\alpha + \beta X_t)dt + \gamma\sqrt{X_t}dW_t, \tag{19}$$

which has been widely used to fit interest rate models since the influential work by Cox, Ingersol and Ross (1985). As discussed, the model can be regarded as a special case of our model with $dY_t = dX_t$ and $dU_t = \gamma\sqrt{X_t}dW_t$. The second model, Model II, is specified as

$$dY_t = (\alpha + \beta X_t)dt + \sqrt{X_t}dW_t \tag{20}$$

with

$$dX_t = (\mu + \nu X_t)dt + \omega\sqrt{X_t}dZ_t,$$

where (W_t) and (Z_t) are Brownian motions with $dW_t dZ_t = \rho dt$. It is commonly referred to as the Heston model, since it was used by Heston (1993) to model stochastic volatility.

4.1 Empirical Illustrations

To estimate Model I, we use the annualized three-months T-bill rates in the secondary market for (X_t). The data are collected at the daily frequency from January 4, 1954 to September 30, 2009, with the sample size $n = 13,927$. The observed rates were zero on December 10, 18 and 24, 2008, which were deleted since they are not comparable to the model. To detect for jumps, we use the test by Lee and Mykland (2007). The test results are somewhat dependent upon the size of window, so we tried various lengths of window from 5 to 100 days. For the 5% level test, the numbers of jumps that the test detect are 332, 112, 75, 74 and 69 corresponding to the window sizes 8, 16, 32, 64 and 96. The number of detected jumps is pretty stable after the window size 45, which yields 70 jumps. We delete them before we implement the martingale method, so the actual number of daily observations is 13,857.

For Model II, we use the S&P 500 Index (SPX) and the squares of the Chicago Board Options Exchange (CBOE) Volatility Index (VIX) respectively for (Y_t) and (X_t), as in Aït-Sahalia and Kimmel (2007). The daily observations of size $n = 4,977$ from January 4, 1990 to September 30, 2009 are used for our estimation. The VIX is not available on March 1, 1991, January 31, 1997 and November 26, 1997, so the corresponding observations of the SPX are also deleted. Once again, we use the test by Lee and Mykland (2007) with size 5% to detect the jumps. For the SPX, the test detects 41, 22, 14, 13 and 14 jumps respectively for the selection of window size 8, 16, 32, 64 and 96. We choose the window size 25, which yields 15 jumps. For the squared VIX, we get 74, 51, 44, 44,and 51 jumps for the window size 8, 16, 32, 64 and 96. For the window size 35 that we choose, we have 44 jumps. Some of the jumps in the SPX and VIX are overlapped. Once we delete all the detected jumps, there are 4922 observations that we use to estimate the model.

The actual sequence $(T^{\delta}_{i\Delta})$ of required time change is obtained by

$$T^{\delta}_{i\Delta} = \delta \underset{k > T_{(i-1)\Delta}}{\operatorname{argmin}} \left| \sum_{j=T_{(i-1)\Delta}+1}^{k} (Y_{j\delta} - Y_{(j-1)\delta})^2 - \Delta \right|$$

sequentially for $i = 1, \ldots, N$ for any fixed $\Delta > 0$. The value of Δ is chosen so that the standard error of the MGE of β is minimized. To estimate the standard error, we use the block bootstrap with the block size $N^{1/3}$, as suggested by Horowitz (2001), and 1,000 bootstrap iterations are made. The search for the optimal Δ is made for the range of $N \geq 20$ and $\Delta \geq (20/n)[Y]^{\delta}_{T}$. This is to avoid having too small values of N and Δ. If N is too small, the size of the estimation sample becomes too small and we expect that the estimate is unstable and the bootstrap procedure to compute the standard error of the estimate is unlikely to perform well. On the other hand, if Δ is too small, there are not enough number of original samples in each of the interval from which we collect the estimation sample. Then the time change will not be very effective. In fact, we will end up using all the observations nullifying the effect of time change if we set Δ sufficiently small.

To search for the optimal value of Δ, we use the 1-dimensional MGE. For Model I, we find that $\Delta = 0.0000887$ minimizes the standard error of the MGE for β, which gives us $N = 141$. For Model II, the optimal value is $\Delta = 0.02829$, and it yields $N = 23$. As expected, the estimates are dependent upon the choice of Δ. However, they are not overly sensitive to Δ. The MGE yields similar estimates for β for all the values of Δ we investigate, though the standard errors of the estimates vary somewhat irregularly as we change the value of Δ. In particular, we do not have the U-shaped standard errors with the unique minimum at some optimal value of Δ.

In Table 1, we report the MGE of β for Model I and Model II. For comparison, we also report the exact MLE for Model I and the approximated MLE proposed by Aït-Sahalia and Kimmel (2007) for Model II.

4.2 Monte Carlo Simulations

This section is to be written.

5. Conclusion

This section is to be written.

Mathematical Appendix

Proof of Lemma 3.1 For any $s \leq t$, we have

$$(U_t - U_s)^2 - [U]_{s,t} = 2\int_s^t (U_r - U_s)dU_r \tag{21}$$

due to Ito's formula, and therefore,

$$\mathbb{E}\left((U_t - U_s)^2 - [U]_{s,t}\right)^2 = 4\mathbb{E}\int_s^t (U_r - U_s)^2 d[U]_r. \tag{22}$$

Moreover, we have

$$\begin{aligned}
\mathbb{E}\int_s^t (U_r - U_s)^2 d[U]_r &\leq b_T \mathbb{E}\int_s^t (U_r - U_s)^2 dr \\
&= b_T \int_s^t \left(\mathbb{E}U_r^2 - \mathbb{E}U_s^2\right) dr \\
&= b_T \int_s^t \left(\mathbb{E}[U]_r - \mathbb{E}[U]_s\right) dr \\
&\leq b_T^2 \int_s^t (r - s) dr \\
&\leq \frac{1}{2} b_T^2 (t - s)^2, \qquad (23)
\end{aligned}$$

due to Assumption 3.1, and the fact that $\mathbb{E}U_t^2 = \mathbb{E}[U]_t$ for all $t \geq 0$. Consequently, we may easily deduce from (22) and (23) that

$$\mathbb{E}\left((U_t - U_s)^2 - [U]_{s,t}\right)^2 \leq 2b_T^2 (t - s)^2 \tag{24}$$

for all $0 \leq s \leq t \leq T$.

We define a stochastic process $[U^\delta]$ on $[0, T]$ by

$$[U^\delta]_t = \sum_{j=1}^{i-1} (U_{j\delta} - U_{(j-1)\delta})^2 + (U_t - U_{(j-1)\delta})^2$$

for $(i-1)\delta \leq t < i\delta$ with $i = 1, \ldots, n$, and

$$[U^\delta]_T = \sum_{i=1}^{n} (U_{i\delta} - U_{(i-1)\delta})^2.$$

It follows from (21) that

$$\left(U_t - U_{(i-1)\delta}\right)^2 - [U]_{(i-1)\delta,t} = 2\int_{(i-1)\delta}^t (U_s - U_{(i-1)\delta})dU_s,$$

and we may easily deduce that $[U^\delta]-[U]$ is a martingale.

We use Doob's L^p-inequality [see, e.g., Revuz and Yor (1994, p.52)] to have

$$\mathbb{E}\sup_{0\le t\le T}\Big([U^\delta]_t-[U]_t\Big)^2\le 4\mathbb{E}\Big([U^\delta]_T-[U]_T\Big)^2. \tag{25}$$

Furthermore, we have

$$\begin{aligned}[U^\delta]_T-[U]_T&=\sum_{i=1}^n\big(U_{i\delta}-U_{(i-1)\delta}\big)^2-[U]_T\\&=\sum_{i=1}^n\Big\{\big(U_{i\delta}-U_{(i-1)\delta}\big)^2-[U]_{(i-1)\delta,i\delta}\Big\},\end{aligned}$$

and it follows from (24) and the orthogonality of martingale differences that

$$\begin{aligned}\mathbb{E}\Big([U^\delta]_T-[U]_T\Big)^2&=\sum_{i=1}^n\mathbb{E}\Big\{\big(U_{i\delta}-U_{(i-1)\delta}\big)^2-[U]_{(i-1)\delta,i\delta}\Big\}^2\\&\le 2nb_T^2\delta^2=2b_T^2(\delta T).\end{aligned} \tag{26}$$

Consequently, we have

$$\mathbb{E}\sup_{0\le t\le T}\Big([U^\delta]_t-[U]_t|\Big)^2=O(b_T^2(\delta T)) \tag{27}$$

due to (25) and (26).

We consider

$$\begin{aligned}\sup_{0\le t\le T}\Big([U]_t^\delta-[U^\delta]_t\Big)^2&=\max_{1\le i\le n}\sup_{(i-1)\delta\le t<i\delta}\big(U_t-U_{(i-1)\delta}\big)^2\\&\le\max_{1\le i\le n}\sup_{(i-1)\delta\le t<i\delta}\Big\{\big(U_t-U_{(i-1)\delta}\big)^2-[U]_{(i-1)\delta,t}\Big\}\\&\quad+\max_{1\le i\le n}\sup_{(i-1)\delta\le t<i\delta}[U]_{(i-1)\delta,t}.\end{aligned} \tag{28}$$

We have

$$\sup_{(i-1)\delta\le t<i\delta}[U]_{(i-1)\delta,t}\le[U]_{(i-1)\delta,i\delta}\le\delta b_T \tag{29}$$

for all $i=1,\ldots,n$. Moreover, it follows from Doob's L^p-inequality and (24) that

$$\begin{aligned}&\mathbb{E}\left[\max_{1\le i\le n}\sup_{(i-1)\delta\le t<i\delta}\Big\{\big(U_t-U_{(i-1)\delta}\big)^2-[U]_{(i-1)\delta,t}\Big\}\right]^2\\&\le\sum_{i=1}^n\mathbb{E}\left[\sup_{(i-1)\delta\le t<i\delta}\Big\{\big(U_t-U_{(i-1)\delta}\big)^2-[U]_{(i-1)\delta,t}\Big\}\right]^2\\&\le 4\sum_{i=1}^n\mathbb{E}\Big\{(U_{i\delta}-U_{(i-1)\delta})^2-[U]_{(i-1)\delta,i\delta}\Big\}^2\\&\le 8nb_T^2\delta^2=8b_T^2(\delta T).\end{aligned} \tag{30}$$

Therefore, we have

$$\mathbb{E} \sup_{0\leq t\leq T} \left([U]_t^\delta - [U^\delta]_t\right)^2 = O(b_T(\delta T)^{1/2}) + O(b_T\delta) = O(b_T(\delta T)^{1/2}) \tag{31}$$

from (28), (29) and (30). Now it follows from (27) and (31) that

$$\begin{aligned}\mathbb{E} \sup_{0\leq t\leq T} \left([U]_t^\delta - [U]_t\right)^2 &\leq \mathbb{E} \sup_{0\leq t\leq T} \left([U]_t^\delta - [U^\delta]_t\right)^2 + \mathbb{E} \sup_{0\leq t\leq T} \left([U^\delta]_t - [U]_t\right)^2 \\ &= O(b_T(\delta T)^{1/2}) + O(b_T^2(\delta T)) = O(b_T(\delta T)^{1/2}),\end{aligned}$$

as was to be shown. □

Proof of Lemma 3.2 Write

$$[Y]_t^\delta = [U]_t^\delta + A_t + 2B_t, \tag{32}$$

where

$$\begin{aligned}A_t &= \sum_{i\delta\leq t} \left(\int_{(i-1)\delta}^{i\delta} \mu(X_t,\theta)dt\right)^2 \\ B_t &= \sum_{i\delta\leq t} \left(\int_{(i-1)\delta}^{i\delta} \mu(X_t,\theta)dt\right)(U_{i\delta} - U_{(i-1)\delta}).\end{aligned}$$

We have

$$\sup_{0\leq t\leq T} A_t^2 \leq \left[\sum_{i=1}^n \left(\int_{(i-1)\delta}^{i\delta} \mu(X_t,\theta)dt\right)^2\right]^2 \leq \left(\sup_{0\leq t\leq T}\sup_{\theta\in\Theta}|\mu(X_t,\theta)|\right)^4 (n\delta^2)^2,$$

and therefore,

$$\mathbb{E}\left(\sup_{0\leq t\leq T} A_t\right)^2 = O\left(c_T^2(\delta T)^2\right), \tag{33}$$

due to Assumption 3.2.

On the other hand, it follows from Cauchy-Schwarz inequality that

$$\begin{aligned}|B_t|^2 &= \left|\sum_{i\delta\leq t} \left(\int_{(i-1)\delta}^{i\delta} \mu(X_t,\theta)dt\right)(U_{i\delta} - U_{(i-1)\delta})\right|^2 \\ &\leq \left[\sum_{i\delta\leq t} \left(\int_{(i-1)\delta}^{i\delta} \mu(X_t,\theta)dt\right)^2\right]\left[\sum_{i\delta\leq t}(U_{i\delta} - U_{(i-1)\delta})^2\right].\end{aligned}$$

Therefore, we have

$$\begin{aligned}\sup_{0\leq t\leq T}|B_t|^2 &\leq \left[\sum_{i=1}^{n}\left(\int_{(i-1)\delta}^{i\delta}\mu(X_t,\theta)dt\right)^2\right]\left[\sum_{i=1}^{n}(U_{i\delta}-U_{(i-1)\delta})^2\right]\\ &\leq \left(\sup_{0\leq t\leq T}\sup_{\theta\in\Theta}|\mu(X_t,\theta)|\right)^2(n\delta^2)(b_T T),\end{aligned}$$

from which it follows immediately that

$$\mathbb{E}\left(\sup_{0\leq t\leq T}|B_t|\right)^2 = O\left((b_T c_T)\delta T^2\right). \tag{34}$$

Note that $[U]_T^\delta \leq b_T T$, due to Assumption 3.1.

Now we have from (32), (33) and (34) that

$$\begin{aligned}\mathbb{E}\left(\sup_{0\leq t\leq T}\left|[Y]_t^\delta-[U]_t^\delta\right|\right)^2 &\leq 2\mathbb{E}\left(\sup_{0\leq t\leq T}A_t\right)^2+8\mathbb{E}\left(\sup_{0\leq t\leq T}|B_t|\right)^2\\ &= O\left(c_T^2(\delta T)^2\right)+O\left((b_T c_T)\delta T^2\right),\end{aligned}$$

and the proof is complete. □

Proof of Corollary 3.3 Let $R_{\delta,T}$ be defined as in (12). Then we have

$$T_{t-R_{\delta,T}} \leq T_t^\delta \leq T_{t-R_{\delta,T}},$$

and therefore,

$$\left|T_t^\delta - T_t\right| \leq \left|T_{t+R_{\delta,T}} - T_{t-R_{\delta,T}}\right| \tag{35}$$

for all $0\leq t\leq S$, since (T_t) is monotonic increasing in $t>0$.

However, it follows from Assumption 3.1 that

$$a_T(T_t - T_s) \leq [U]_{T_t} - [U]_{T_s} = t-s$$

for all $0\leq s\leq t\leq S$, and that

$$|T_t - T_s| \leq a_T^{-1}(t-s) \tag{36}$$

for all $0\leq s\leq t\leq S$. Consequently, we may easily deduce from (35) and (36) that

$$\sup_{0\leq t\leq S}\left|T_t^\delta - T_t\right| \leq 2a_T^{-1}R_{\delta,T},$$

and that

$$\mathbb{E}\left(\sup_{0\leq t\leq S}\left|T_t^\delta - T_t\right|\right)^2 \leq 4a_T^{-2}\mathbb{E}R_{\delta,T}^2,$$

from which the stated result follows immediately. □

Proof of Lemma 3.4 We will show that

$$\mathbb{E} \sup_{0\le t\le S} \sup_{\theta\in\Theta} \left| \int_0^{T_t^\delta} \mu(X_s,\theta)ds - \int_0^{T_t} \mu(X_s,\theta)ds \right| = o(N^{-1/2}) \tag{37}$$

and that

$$\mathbb{E} \sup_{0\le t\le S} \left| Y_{T_t^\delta} - Y_{T_t} \right| = o(N^{-1/2}), \tag{38}$$

from which the stated result follows readily. To establish (37), we simply note that

$$\left| \int_0^{T_t^\delta} \mu(X_s,\theta)ds - \int_0^{T_t} \mu(X_s,\theta)ds \right| \le \left(\sup_{0\le t\le T} \sup_{\theta\in\Theta} |\mu(X_t,\theta)| \right) \left(\sup_{0\le t\le S} \left| T_t^\delta - T_t \right| \right),$$

which holds for all $0 \le t \le S$ and $\theta \in \Theta$. Therefore, it follows that

$$\begin{aligned}
&\mathbb{E} \sup_{0\le t\le S} \left| \int_0^{T_t^\delta} \mu(X_s,\theta)ds - \int_0^{T_t} \mu(X_s,\theta)ds \right| \\
&\quad \le \left[\mathbb{E} \left(\sup_{0\le t\le T} \sup_{\theta\in\Theta} |\mu(X_t,\theta)| \right)^2 \right]^{1/2} \left[\mathbb{E} \left(\sup_{0\le t\le S} \left| T_t^\delta - T_t \right| \right)^2 \right]^{1/2} \\
&\quad = O(c_T^{1/2}) \left[O\left(\frac{b_T^{1/2}}{a_T} (\delta T)^{1/4} \right) + O\left(\frac{c_T}{a_T} (\delta T) \right) + O\left(\frac{(b_T c_T)^{1/2}}{a_T} (\delta^{1/2} T) \right) \right] \\
&\quad = O\left(\frac{(b_T c_T)^{1/2}}{a_T} (\delta T)^{1/4} \right) + O\left(\frac{c_T^{3/2}}{a_T} (\delta T) \right) + O\left(\frac{(b_T^{1/2} c_T)}{a_T} (\delta^{1/2} T) \right),
\end{aligned} \tag{39}$$

from Cauchy-Schwarz inequality, Assumption 3.2 and Corollary 3.3. Now we may easily deduce (37) from (39), upon noticing that

$$O\left(\frac{(b_T c_T)^{1/2}}{a_T} (\delta T)^{1/4} \right) + O\left(\frac{c_T^{3/2}}{a_T} (\delta T) \right) + O\left(\frac{(b_T^{1/2} c_T)}{a_T} (\delta^{1/2} T) \right) = o(N_1^{-1/2}),$$

as we assume in Assumption 3.3.

To deduce (38), we note that

$$\left| Y_{T_t^\delta} - Y_{T_t} \right| \le \left| \int_0^{T_t^\delta} \mu(X_s,\theta_0)ds - \int_0^{T_t} \mu(X_s,\theta_0)ds \right| + \left| U_{T_t^\delta} - U_{T_t} \right|$$

for all $0 \le t \le S$. Since it follows immediately from (37) that

$$\mathbb{E} \sup_{0\le t\le S} \left| \int_0^{T_t^\delta} \mu(X_s,\theta_0)ds - \int_0^{T_t} \mu(X_s,\theta_0)ds \right| = o(N_1^{-1/2}),$$

we only need to prove

$$\mathbb{E} \sup_{0\le t\le S} \left| U_{T_t^\delta} - U_{T_t} \right| = o(N_2^{-1/2}) \tag{40}$$

to establish (38).

Let $R_{\delta,T}$ be defined as in (12). Then we have

$$\inf_{T_{t-R_{\delta,T}}\leq t\leq T_{t+R_{\delta,T}}} U_t \leq U_{T_t^\delta} \leq \sup_{T_{t-R_{\delta,T}}\leq t\leq T_{t+R_{\delta,T}}} U_t,$$

i.e.,

$$\inf_{|s-t|\leq R_{\delta,T}} V_s \leq U_{T_t^\delta} \leq \sup_{|s-t|\leq R_{\delta,T}} V_s$$

for all $0 \leq t \leq S$. Consequently, it follows that

$$\left|U_{T_t^\delta} - U_{T_t}\right| \leq \sup_{|s-t|\leq R_{\delta,T}} |V_t - V_s|$$

for all $0 \leq t \leq S$, and therefore,

$$\sup_{0\leq t\leq S}\left|U_{T_t^\delta} - U_{T_t}\right| \leq \sup_{0\leq t\leq S}\sup_{|s-t|\leq R_{\delta,T}} |V_t - V_s|. \tag{41}$$

However, we have

$$\begin{aligned}\sup_{0\leq t\leq S}\sup_{|s-t|\leq R_{\delta,T}} |V_t - V_s| &= S^{1/2}\sup_{0\leq u\leq 1}\sup_{|u-v|\leq R_{\delta,T}/S} S^{-1/2}\left|V_{Su} - V_{Sv}\right| \\ &= S^{1/2}\left(2\frac{R_{\delta,T}}{S}\log\frac{S}{R_{\delta,T}}\right)^{1/2} \\ &= \left(2R_{\delta,T}\log\frac{S}{R_{\delta,T}}\right)^{1/2}, \end{aligned} \tag{42}$$

due to the Lévy's modulus of continuity of Brownian motion [see, for instance, Karatzas and Shreve (1988, p.114)].

Note that $S = [U]_T \leq b_T T$. Therefore, it follows from (41) and (42) that

$$\sup_{0\leq t\leq S}\left|U_{T_t^\delta} - U_{T_t}\right| \leq \sqrt{2}(b_T T)^\varepsilon R_{\delta,T}^{(1-\varepsilon)/2}$$

for any $\varepsilon > 0$, and we have

$$\begin{aligned}&\mathbb{E}\sup_{0\leq t\leq S}\left|U_{T_t^\delta} - U_{T_t}\right| \leq \sqrt{2}(b_T T)^\varepsilon \left(\mathbb{E}R_{\delta,T}^2\right)^{(1-\varepsilon)/4} \\ &\quad = O\left((b_T T)^\varepsilon\right)\left[O\left(b_T(\delta T)^{1/2}\right) + O\left(c_T^2(\delta T)^2\right) + O\left((b_T c_T)\delta T^2\right)\right]^{(1-\varepsilon)/4} \\ &\quad = O\left(b_T^{1/4}(\delta^{1/8-\varepsilon}T^{1/8})\right) + O\left(c_T^{1/2}(\delta^{1/2-\varepsilon}T^{1/2})\right) + O\left((b_T c_T)^{1/4}(\delta^{1/4-\varepsilon}T^{1/2})\right). \end{aligned} \tag{43}$$

However, it follows from Assumption 3.3 that

$$O\left(b_T^{1/4}(\delta^{1/8-\varepsilon}T^{1/8})\right) + O\left(c_T^{1/2}(\delta^{1/2-\varepsilon}T^{1/2})\right) + O\left((b_T c_T)^{1/4}(\delta^{1/4-\varepsilon}T^{1/2})\right) = o(N_2^{-1/2}),$$

and (40) follows directly from (43). The proof is therefore complete. □

Proof of Lemma 3.5 We have

$$\left|\Pi_N^\delta(z,\theta)-\Pi_N(z,\theta)\right|\leq\frac{1}{N}\sum_{i=1}^N\left|1\{Z_{i_d}^\delta(\theta)\leq z\}-1\{Z_{i_d}(\theta)\leq z\}\right|$$

and

$$\begin{aligned}&\left|1\{Z_{i_d}^\delta(\theta)\leq z\}-1\{Z_{i_d}(\theta)\leq z\}\right|\\&\quad\leq\left|1\{Z_i^\delta(\theta)\leq z_1\}\cdots 1\{Z_{i-d+1}^\delta(\theta)\leq z_d\}-1\{Z_i(\theta)\leq z_1\}\cdots 1\{Z_{i-d+1}(\theta)\leq z_d\}\right|\\&\quad\leq\sum_{j=1}^d\left|1\{Z_{i-j+1}^\delta(\theta)\leq z_j\}-1\{Z_{i-j+1}(\theta)\leq z_j\}\right|.\end{aligned}$$

Moreover, it follows that

$$\left|1\{Z_{i-j+1}^\delta(\theta)\leq z_j\}-1\{Z_{i-j+1}(\theta)\leq z_j\}\right|\leq 1\{|Z_{i-j+1}(\theta)-z_j|\leq|Z_{i-j+1}^\delta(\theta)-Z_{i-j+1}(\theta)|\}.$$

However, it follows from Assumption 3.4 that

$$\mathbb{P}\{|Z_{i-j+1}(\theta)-z_j|\leq|Z_{i-j+1}^\delta(\theta)-Z_{i-j+1}(\theta)|\}\leq K\mathbb{E}|Z_{i-j+1}^\delta(\theta)-Z_{i-j+1}(\theta)|,$$

where K is a constant independent of $i,j\geq 1$ and $\theta\in\Theta$. Finally, note that

$$\mathbb{E}|Z_{i-j+1}^\delta(\theta)-Z_{i-j+1}(\theta)|=o(N^{-1/2}).$$

and the proof is complete. □

Proof of Lemma 3.6 We define a pseudometric r on $\mathbb{R}^d$ as

$$r(z_1,z_2)=\max_{1\leq j\leq d}[\Phi(z_{1j}\vee z_{2j})-\Phi(z_{1j}\wedge z_{2j})],$$

where $z_1=(z_{1j})$ and $z_2=(z_{2j})$, and subsequently, introduce a pseudometric ρ on $\mathbb{R}^d\times\Theta$ given by

$$\rho((z_1,\theta_1),(z_2,\theta_2))=r(z_1,z_2)\vee\|\theta_1-\theta_2\|. \tag{44}$$

Moreover, we let $\mathcal{F}$ be a class of random functions

$$f(z,\theta)=1\big\{Z_{i_d}(\theta)\leq z\big\} \tag{45}$$

indexed by $(z,\theta)\in\mathbb{R}^d\times\Theta$. Our proof heavily relies on Andrews and Pollard (1994) applied with the pseudometric ρ and the class of functions defined in (44) and (45).

Let $\varepsilon>0$ be given, and consider a rectangle given by $\mathcal{R}=[\underline{z},\overline{z}]=\prod_{j=1}^d[\underline{z}_j,\overline{z}_j]\subset\mathbb{R}^d$ with $r(\underline{z},\overline{z})\leq\varepsilon^2$ and a neighborhood $\mathcal{N}=\{\theta\in\Theta|\|\theta-\theta_0\|\leq\varepsilon\}$ of $\theta_0\in\Theta$. Then we may deduce that

$$\left[\mathbb{E}\sup_{x,y\in\mathcal{R},\theta\in\mathcal{N}}\Big(1\{Z_{i_d}(\theta)\leq x\}-1\{Z_{i_d}(\theta_0)\leq y\}\Big)^2\right]^{1/2}\leq K\varepsilon \tag{46}$$

as shown below, where and elsewhere in the proof K denotes the generic constant whose actual value may vary from line to line. To show (46), we note that

$$\begin{aligned}
&\Big|1\{Z_{i_d}(\theta)\le x\}-1\{Z_{i_d}(\theta_0)\le y\}\Big|\\
&\quad\le\Big|1\{Z_{i_d}(\theta)\le x\}-1\{Z_{i_d}(\theta_0)\le x\}\Big|+\Big|1\{Z_{i_d}(\theta_0)\le x\}-1\{Z_{i_d}(\theta_0)\le y\}\Big|\\
&\quad\le\sum_{j=1}^{d}\Big|1\{Z_{i-j+1}(\theta)\le x_j\}-1\{Z_{i-j+1}(\theta_0)\le x_j\}\Big|\\
&\qquad+\sum_{j=1}^{d}\Big|1\{Z_{i+j-1}(\theta_0)\le x_j\}-1\{Z_{i-j+1}(\theta_0)\le y_j\}\Big|, \qquad (47)
\end{aligned}$$

and that we have

$$\begin{aligned}
&\Big|1\{Z_{i-j+1}(\theta)\le x_j\}-1\{Z_{i-j+1}(\theta_0)\le x_j\}\Big|\\
&\quad\le 1\Big\{|Z_{i-j+1}(\theta_0)-x_j|\le|Z_{i-j+1}(\theta)-Z_{i-j+1}(\theta_0)|\Big\}\\
&\quad\le 1\left\{|Z_{i-j+1}(\theta_0)-x_j|\le\|\theta-\theta_0\|\int_{T_{(i-j)\Delta}}^{T_{(i-j+1)\Delta}}\nu(X_t)dt\right\}\\
&\quad\le K\varepsilon\int_{T_{(i-j)\Delta}}^{T_{(i-j+1)\Delta}}\nu(X_t)dt, \qquad (48)
\end{aligned}$$

and

$$\Big|1\{Z_{i-j+1}(\theta_0)\le x_j\}-1\{Z_{i-j+1}(\theta_0)\le y_j\}\Big|\le 1\Big\{\underline{z}_j\le Z_{i-j+1}(\theta_0)\le\overline{z}_j\Big\} \qquad (49)$$

for $j=1,\ldots,d$. Consequently, due to (48) and (49), it follows immediately from (47) that

$$\mathbb{E}\sup_{x,y\in\mathcal{R},\theta\in\mathcal{N}}\Big(1\{Z_{i_d}(\theta)\le x\}-1\{Z_{i_d}(\theta_0)\le y\}\Big)^2\le K\varepsilon^2\left[\mathbb{E}\left(\int_{T_{(i-j)\Delta}}^{T_{(i-j+1)\Delta}}\nu(X_t)dt\right)^2+1\right]$$

from which we may easily deduce (46), by redefining constant K appropriately.

Now we define $N(x,\mathcal{F})$ to be the bracketing number for the set $\mathbb{R}^d\times\mathcal{N}$ using the class of functions introduced in (44). It is obvious that we may cover the entire $\mathbb{R}^d$ by a set of $O(\varepsilon^{-2d})$ many rectangles of r-length ε^2. Therefore, we have

$$N(x,\mathcal{F})=x^{-2d}.$$

To employ the result by Andrews and Pollard (1994, Theorem 2.2), we need to show that for $\alpha(k)=k^{-c}$ and $N(x,\mathcal{F})=x^{-2d}$

$$\sum_{k=1}^{\infty}k^{a-2}\alpha(k)^{b/(a+b)}<\infty\quad\text{and}\quad\int_0^1 x^{-b/(2+b)}N(x,\mathcal{F})^{1/a}dx<\infty \qquad (50)$$

hold with some even integers $a \geq 2$ and $b > 0$. Note that the conditions in (50) are satisfied if and only if

$$(a-2) - \frac{cb}{a+b} < -1, \quad -\frac{d}{a} - \frac{b}{2+b} > -1,$$

which hold if and only if

$$\frac{a(a-1)}{c-(a-1)} < b < \frac{a}{d} - 2.$$

Therefore, the required a and b exist if and only if

$$\frac{a}{2d} > 1$$

and

$$c > \frac{(a-1)\left[\left(\frac{1}{d}+1\right)a - 2\right]}{2\left(\frac{a}{2d}-1\right)}.$$

In particular, if we set

$$a = 4d$$

and

$$c > (2d+1)(4d-1),$$

the required conditions are all met. The proof is therefore complete. □

Proof of Theorem 3.7 We first derive the asymptotic distribution of $\hat{\theta}_N$. Given Lemma 3.6, our proof of Theorem 3.7 is largely identical to that of Theorem 5 in Brown and Wegkamp (2002), which in turn heavily relies on Theorem 3.2 in Wegkamp (1998). Since our setup and notation are slightly different from theirs, we include a brief sketch of the proof here. The inclusion of the proof would also make straightforward the proof for the asymptotic equivalence of $\hat{\theta}_N^\delta$ and $\hat{\theta}_N$, which will be given later.

Note that

$$\begin{aligned} Q_N(\theta) - Q(\theta) = &\int [\Pi_N(z,\theta) - \Pi(z,\theta)]^2 \varpi(dz) \\ &+ 2\int [\Pi(z,\theta) - \Pi(z,\theta_0)][\Pi_N(z,\theta) - \Pi(z,\theta)]\varpi(dz) \end{aligned} \tag{51}$$

However, due to Lemma 3.6, $\sqrt{N}[\Pi_N(z,\theta) - \Pi(z,\theta)]$ is stochastically equicontinuous at θ_0 with respect to the Euclidean metric on Θ for all $z \in \mathbb{R}^d$, i.e.,

$$\sqrt{N}[\Pi_N(z,\theta) - \Pi(z,\theta)] - \sqrt{N}[\Pi_N(z,\theta_0) - \Pi(z,\theta_0)] \to_p 0 \tag{52}$$

uniformly in $z \in \mathbb{R}^d$, as $\theta \to \theta_0$. Moreover, it follows from Assumption 3.6(b) that

$$\Pi(z,\theta) - \Pi(z,\theta_0) = (\theta - \theta_0)'\dot{\Pi}(z,\theta_0) + R(z,\theta), \tag{53}$$

where $\int R(z,\theta)^2 \varpi(dz) = o(\|\theta - \theta_0\|^2)$ near θ_0.

It follows from (52) and (53) that

$$\int[\Pi_N(z,\theta)-\Pi(z,\theta)]^2\varpi(dz)=\int[\Pi_N(z,\theta_0)-\Pi(z,\theta_0)]^2\varpi(dz)+o_p(N^{-1})$$

and

$$\begin{aligned}&\int[\Pi(z,\theta)-\Pi(z,\theta_0)][\Pi_N(z,\theta)-\Pi(z,\theta)]\varpi(dz)\\&\quad=(\theta-\theta_0)'\int\dot\Pi(z,\theta_0)[\Pi_N(z,\theta_0)-\Pi(z,\theta_0)]\varpi(dz)+o_p\left(N^{-1/2}\|\theta-\theta_0\|\right)\end{aligned}$$

near θ_0. Therefore, we may easily deduce from (51) that

$$\begin{aligned}Q_N(\theta)-Q(\theta)&=\int[\Pi_N(z,\theta_0)-\Pi(z,\theta_0)]^2\varpi(dz)\\&\quad+2(\theta-\theta_0)'\int\dot\Pi(z,\theta_0)[\Pi_N(z,\theta_0)-\Pi(z,\theta_0)]\varpi(dz)+o_p(N^{-1})+o_p\left(N^{-1/2}\|\theta-\theta_0\|\right)\end{aligned}\tag{54}$$

near θ_0.

However, due to the second-order differentiability of Q, it follows that

$$\begin{aligned}Q(\theta)&=Q(\theta_0)+\dot Q(\theta_0)(\theta-\theta_0)+\frac{1}{2}(\theta-\theta_0)'\ddot Q(\theta_0)(\theta-\theta_0)+o(\|\theta-\theta_0\|^2)\\&=\frac{1}{2}(\theta-\theta_0)'\ddot Q(\theta_0)(\theta-\theta_0)+o(\|\theta-\theta_0\|^2)\end{aligned}\tag{55}$$

near θ_0, where $\dot Q$ and $\ddot Q$ are respectively the first-order and second-order derivatives of Q. Note that $Q(\theta_0)=\dot Q(\theta_0)=0$. Consequently, we have from (54) and (55) that

$$\begin{aligned}Q_N(\theta)=&\int[\Pi_N(z,\theta_0)-\Pi(z,\theta_0)]^2\varpi(dz)\\&+2(\theta-\theta_0)'\int\dot\Pi(z,\theta_0)[\Pi_N(z,\theta_0)-\Pi(z,\theta_0)]\varpi(dz)\\&+\frac{1}{2}(\theta-\theta_0)'\ddot Q(\theta_0)(\theta-\theta_0)\\&+o_p(N^{-1})+o_p\left(N^{-1/2}\|\theta-\theta_0\|\right)+o(\|\theta-\theta_0\|^2)\end{aligned}\tag{56}$$

near θ_0. Given Assumption 3.6(c) it can therefore be deduced that

$$\sqrt{N}(\hat\theta_N-\theta_0)=-2\ddot Q(\theta_0)^{-1}\sqrt{N}\int\dot\Pi(z,\theta_0)[\Pi_N(z,\theta_0)-\Pi(z,\theta_0)]\varpi(dz)+o_p(1)\tag{57}$$

for large n, as in the proof of Theorem 3.2 in Wegkamp (1998). The asymptotic distribution of $\hat\theta_N$ may be easily derived from (57) as explained in the discussion prior to Theorem 3.7.

Now we show the asymptotic equivalence between $\hat{\theta}_N$ and $\hat{\theta}_N^\delta$. To do so, we write

$$\begin{aligned} Q_N^\delta(\theta) - Q_N(\theta) &= \int [\Pi_N^\delta(z,\theta) - \Pi_N(z,\theta)]^2 \varpi(dz) \\ &\quad + 2\int [\Pi_N^\delta(z,\theta) - \Pi_N(z,\theta)][\Pi_N(z,\theta) - \Pi(z,\theta_0)]\varpi(dz). \end{aligned} \tag{58}$$

It follows from Lemma 3.5 that

$$\int [\Pi_N^\delta(z,\theta) - \Pi_N(z,\theta)]^2 \varpi(dz) = o(N^{-1}) \tag{59}$$

uniformly in $\theta \in \Theta$. Moreover, we have

$$\begin{aligned} &\left| \int [\Pi_N^\delta(z,\theta) - \Pi_N(z,\theta)][\Pi_N(z,\theta) - \Pi(z,\theta_0)]\varpi(dz) \right| \\ &\leq \left(\int [\Pi_N^\delta(z,\theta) - \Pi_N(z,\theta)]^2 \varpi(dz) \right)^{1/2} \left(\int [\Pi_N(z,\theta) - \Pi(z,\theta_0)]^2 \varpi(dz) \right)^{1/2} \\ &= o_p(N^{-1}) \end{aligned} \tag{60}$$

near θ_0. Therefore, we may deduce from (58), (59) and (60) that

$$\left| Q_N^\delta(\theta) - Q_N(\theta) \right| = o_p(N^{-1})$$

near θ_0. The asymptotic equivalence between $\hat{\theta}_N$ and $\hat{\theta}_N^\delta$ can therefore be seen easily from (56), and the proof is complete. □

References

Aït-Sahalia, Y. and R. Kimmel (2007), "Maximum likelihood estimation of stochastic volatility models," *Journal of Financial Economics*, 83, 413-452.

Andersen, T.G., T. Bollerslev and D. Dobrev (2007), "No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications," *Journal of Econometrics*, 138, 125-180.

Andrews, D.W.K. and D. Pollard (1994), "An introduction to functional central limit theorems for dependent stochastic processes," *International Statistical Review*, 62, 119-132.

Bergstrom, A.R. (1984), "Continuous time stochastic models and issues of aggregation," *Handbook of Econometrics*, vol. 2, 1146-1212, ed. by Z. Griliches and M.D. Intriligator, Amsterdam: Noth-Holland.

Berman, S.M. (1992). *Sojourns and Extremes of Stochastic Processes*, California: Wadsworth & Brooks/Cole.

Brown, D. and M.H. Wegkamp (2002), “Weighted minimum mean-square distance from independent estimation,” *Econometrica*, 70, 2035-2051.

Chang, Y. (2008), “Taking a new contour: A novel approach to panel unit root tests,” Working paper, Department of Economics, Indiana University.

Cox, J., J. Ingersol and S. Ross (1985), “A theory of the term structure of interest rates,” *Econometrica*, 53, 385-407.

Deo, C.M. (1973), “A note on empirical processes of strong-mixing sequences,” *Annals of Probability*, 1, 870-875.

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.

Horowitz, J. (2001), “The bootstrap,” *Handbook of Econometrics*, vol. 5, 3159-3228, ed. by J.J. Heckman and E.E. Leamer, Amsterdam: North-Holland.

Karatzas, I. and S.E. Shreve (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed., New York: Springer-Verlag.

Lee, S.S. and P.A. Mykland (2007), “Jumps in financial markets: A new nonparametric test and jump dynamics,” *Review of Financial Studies*, 21, 2535-2563.

Manski, C.F. (1983), “Closest empirical distribution estimation,” *Econometrica*, 51, 305-320.

Park, J.Y. and R. Vasudev (2006), “Testing for martingales in continuous time,” Working paper, Department of Economics, Rice University.

Peters, R.T. and R.G. de Vilder (2006), “Testing the continuous semimartingale hypothesis for the S&P 500,” *Journal of Business and Economic Statistics*, 24, 444-454.

Revuz, D. and M. Yor (1994). *Continuous Martingale and Brownian Motion*, 2nd ed., New York: Springer-Verlag.

Wegkamp, M. (1998). *Entropy Methods in Statistical Estimation*, Amsterdam: CWI-tract 125.

Yu, J. and P.C.B. Phillips (2001), “A Gaussian approach for continuous time models of the short-term interest rate,” *Econometrics Journal*, 4, 210-224.