Martingale Regressions for Conditional Mean Models in Continuous Time[1]

Joon Y. Park
Department of Economics
Indiana University

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 the martingale regression, since the parameter in the model is identified by the residual process being a martingale. Upon an appropriate time change, the continuous part of the error process in the martingale regression can always be transformed into a Brownian motion. We use this property and apply a minimum distance method to estimate the parameters in the model. To implement our methodology, we may simply collect the samples at the required random time intervals, and define our estimates to be the parameter values which make the empirical distribution of the residuals closest to independent and identically distributed normals. It is shown by simulation that our approach yields a very reliable method of inference applicable for the general continuous time conditional mean model.

This Version: November 2010
JEL Classification Number: C22

Key words and phrases: conditional mean models in continuous time, martingale estimation, time change, diffusion, continuous time asset pricing models

[1] I am very grateful to Yoosoon Chang, Yongok Choi 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, and various versions of the paper have been circulated under the titles "Martingale Regression and Time Change" and "Inference on Conditional Mean Models in Continuous Time". 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. This is in contrast with the conventional approach based on the traditional framework of classical regressions, for which the reader is referred to Bergstrom (1984). 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.

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, it necessarily yields some discretization bias, which may be substantial unless the discretization is appropriately done and carefully taken care of. Second, 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 too small. Third, 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.

In this paper, we propose an approach to more effectively deal with the martingale regression. Our methodology uses some fundamental properties of martingales and does not rely on any orthogonality condition. It is based on the use of a time change, which transforms a general martingale into Brownian motion, given by the celebrated theorem of Dambis, Dubins and Schwarz. The DDS theorem, for short, implies 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. The DDS theorem has already been used by several authors in various contexts. Yu and Phillips (2001) exploited it 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 our approach, we only estimate and test for the continuous part of the martingale regression. In most of the existing economic and financial models, jumps are generated exogenously and do not include any information on model parameters. Therefore, for the applications of our methodology, we simply regard jumps as pure noise and employ a preliminary test to identify and discard the observations contaminated by noisy jumps. To test for jumps at the preliminary step, we may use the test by Lee and Mykland (2008), which serves our purpose well and performs quite satisfactorily. Of course, our method is subject to the potential problems of statistical procedures based on preliminary tests, such as increased variances for estimators and size distortions in tests. To minimize the negative effect of preliminary test, we recommend to use samples collected at relatively lower frequencies such as daily rather than intra-day observations at ultra-high frequencies.[2] At daily frequency, for instance, moderately large jumps occur only intermittently and the impact of preliminary test appears to be minimal and insignificant in practical applications of our methodology.

For the actual implementation of our methodology, we of course have to rely on discrete samples. We assume that the observations are available at high frequencies such as daily.[3] 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 obser-

[2]For the type of models considered in the paper, it is well known that the efficiency of estimator is determined by the sampling horizon, not by the sample size. Using lower frequency observations, therefore, do not deteriorate the efficiency of estimator, as long as they are sampled frequently enough to study the underlying continuous time model.

[3]Our methodology and its asymptotic theory are applicable as long as the sampling interval is sufficiently small relative to the sampling horizon. Therefore, we may allow for observations at monthly or quarterly frequency as long as the sampling horizon is large enough.

vations 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 martingale estimator (MGE). For comparison, we also consider the maximum likelihood estimator (MLE), which uses full information on the entire structure of our simulation models. The overall performance of our martingale estimator (MGE) is quite good, and at least comparable to that of the MLE for all models considered in our simulations. We notice some tendency that the MLE does relatively better than the MGE as the sampling horizon increases. However, for the normal range of sampling horizons that we encounter in most practical applications, we do not observe any evidence of asymptotic relative efficiency of the MLE. In fact, in terms of bias, the MGE does substantially better than the MLE. This is especially so, when the sampling horizon is only modest.[4] In general, the bias of the MLE is of an unacceptably magnitude except for the case that the sampling horizon is unrealistically large. In sharp contrast, the bias of the MGE is generally very small and becomes truly negligible in many cases, even when the sampling horizon is quite small. The standard deviation of the MGE is largely the same as that of the MLE.

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 the martingale estimator has normal asymptotic distribution under appropriate regularity conditions. The resampling method to estimate the asymptotic variance of the estimators is also discussed. Section 4 presents some empirical illustrations and reports our simulation results. Various issues arising in implementing our methodology are discussed in detail, and the finite sample performance of our MGE is compared with that of the MLE by Monte Carlo simulations. The effect of jumps on the performance of the MGE is also studied. 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

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

[4]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.

where (Y_t) and (X_t) are stochastic processes,[5] $(\mathcal{F}_t)$ 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 currencies 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 essential for the development of our methodology and its asymptotic theory in the paper. However, the presence of jumps in the error process can be readily

[5]Throught the paper, (X_t) is defined as a vector process though it is often specialized as a scalar process.

accommodated in practical applications by employing a preliminary test for jumps. This will be explained in detail in later sections.

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.,

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

and therefore, it follows from $\mathbb{E}((d\pi_t/\pi_t)|\mathcal{F}_t) = -r_t^f dt$ that

$$\frac{dP_t}{P_t} - r_t^f = -\mathbb{E}\left(\frac{d\pi_t}{\pi_t}\frac{dP_t}{P_t}\middle|\mathcal{F}_t\right) + dU_t,$$

where (U_t) is a martingale.[6] Both the asset price and the state-price deflator processes are conventionally specified as Ito processes, in which case $-\mathbb{E}((d\pi_t/\pi_t)(dP_t/P_t)|\mathcal{F}_t)$ yields 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 $\theta \in \Theta$ with the true value θ_0.[7]

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

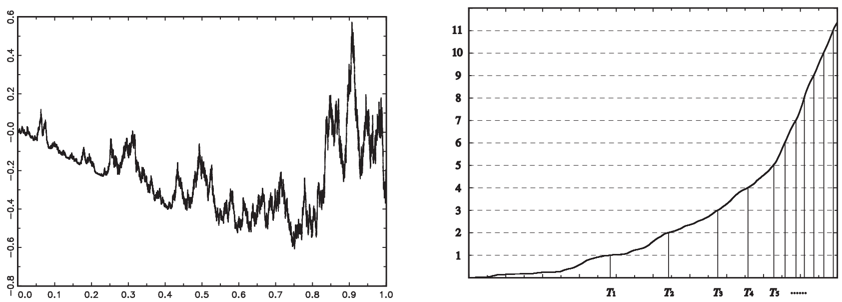
$$dX_t = \mu_t dt + \sigma_t dW_t, \tag{5}$$

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

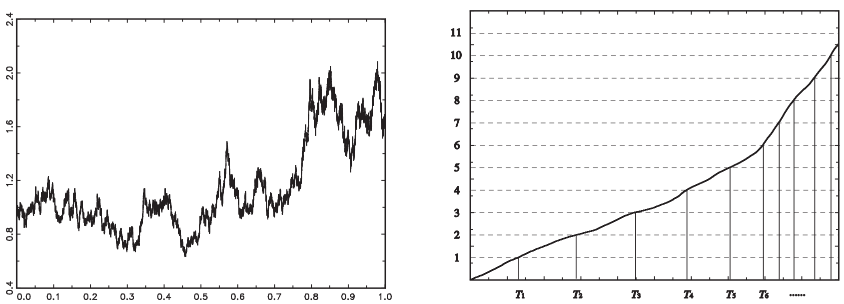
[6] Following the convention used in finance and financial economics literature, we denote by $(d\pi_t dP_t)$ the differential of quadratic covariation between (π_t) and (P_t).

[7] In general, the volatility functions of (dP_t/P_t) and $(d\pi_t/\pi_t)$ are not directly observable and we have to use their proxies or estimates. Later we will explain in detail how we accommodate the use of proxies and estimates for covariates in our methodology.

Figure 1: Sample Paths and Quadratic Variations with Time Changes



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



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

$\rho \in \mathbb{R}_+$, 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.[8]

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

[8] There is a clear advantage of our approach here. As shown in Jeong and Park (2010), the MLE's for the drift term and diffusion term parameters are asymptotically independent for a wide class of diffusion models, both stationary and nonstationary, and there will be no efficiency gain in joint estimation.

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, 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. 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

$$\tau_N = NQ_N(\hat{\theta}_N) \tag{11}$$

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 $Q_N(\hat{\theta}_N) \nrightarrow_p 0$. 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 M-observations on (Y_t) and (X_t) with sampling interval $\delta > 0$, which are denoted by

$$(X_\delta, Y_\delta), \ldots, (X_{j\delta}, Y_{j\delta}), \ldots, (X_{M\delta}, Y_{M\delta}) \tag{12}$$

with the initial value (X_0, Y_0). Throughout the paper, we let $T = M\delta$ 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

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

i.e., the realized variance of (Y_t) over time interval $[0,t]$. For the hypothesis testing, we may also impose the null value θ_0 of $\theta \in \Theta$, so that $[U]_t^\delta = \sum_{j\delta \le t}(U_{j\delta} - U_{(j-1)\delta})^2$ 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 $0 \le s \le t \le T$,

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

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 (U_t), defined as $dU_t = \sigma_t dW_t$ for some volatility process (σ_t) and Brownian motion (W_t), if we have

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

Many diffusion processes satisfy this condition.

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

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

with 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) $\sup_{\theta\in\Theta}|\mu(x,\theta)|\leq c\|x\|^p$ for some constants $c>0$ and $p\geq 0$, and let (b) $\mathbb{E}\big(\sup_{0\leq t\leq T}\|X_t\|\big)^r=O(T^q)$ for some $r\geq 4p$ and $q\geq 0$. Then we have

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

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 $r\geq 0$. 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 $r\geq 4p$ 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 $\sup_{0\leq t\leq T}|X_t|$ 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

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

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*

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

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). In general, the required condition in Lemma 3.2 is more stringent than the one in Lemma 3.1. This is because (Y_t) has the conditional mean component and we need to ensure that its contribution to the estimation of quadratic variation is asymptotically negligible.[9]

Now we define
$$R_{\delta,T} = \sup_{0\le t\le T} \left| [Y]_t^\delta - [U]_t \right| . \tag{13}$$

Then it follows that

$$\mathbb{E}R_{\delta,T}^2 = O\left(\delta T b_T^2\right) + O\left((\delta T)^2 c_T^2\right) + O\left(\delta T^2 (b_T c_T)\right) \tag{14}$$

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*

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

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 $T \to \infty$.

3.2 Feasible Martingale Estimator

In place of $(Z_i(\theta))$ introduced earlier in (8), the feasible MGE is based on $(Z_i(\theta)^\delta)$, which is defined for $i = 1, \ldots, N$ by

$$Z_i^\delta(\theta) = \Delta^{-1/2} \left(Y_{T_{i\Delta}^\delta} - Y_{T_{(i-1)\Delta}^\delta} - \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta}, \theta) \right) \tag{15}$$

with $\delta M_i = T_{i\Delta}^\delta$. We may easily see that

$$\sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta}, \theta) \approx \int_{T_{(i-1)\Delta}^\delta}^{T_{i\Delta}^\delta} \mu(X_t, \theta)dt \tag{16}$$

for small $\delta > 0$, if $(\mu(X_\cdot, \theta))$ is Riemann integrable on the interval $[T_{(i-1)\Delta}^\delta, T_{i\Delta}^\delta]$.

Note that we construct the samples $(Z_i(\theta)^\delta)$ of size N in (15), from the observations $(X_{j\delta}, Y_{j\delta})$ of size M on the underlying stochastic processes (X, Y) given in (12), to estimate

[9] We may use the fitted residuals $\hat{U}_{j\delta} - \hat{U}_{(j-1)\delta} = (Y_{j\delta} - Y_{(j-1)\delta}) - \mu(X_{(j-1)\delta}, \hat{\theta})\delta$, $j = 1, \ldots, M$, 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. From the empirical perspective, however, such a two step procedure does not seem to be very meaningful. For the typical models that we may apply our methodology, the conditional mean components are of order significantly smaller than their martingale components.

the parameter θ in the model. Throughout the paper, we call $(X_{j\delta}, Y_{j\delta})$ and $(Z_i(\theta)^\delta)$, respectively, the *original* and *estimation* samples to avoid confusion. Of course, we have $M > N$, and we use a fewer number of estimation samples collected from the original samples. In general, this does not incur any loss in efficiency. First, for the implementation of our martingale procedure we use all observations to compute the quadratic variation of the error process and to obtain the integral value of the conditional mean function. Second, even if the observations are available continuously in time, we would not use them all in our methodology for the estimation of conditional mean function. As discussed earlier, it is expected that the optimal value of Δ exists and is strictly positive for a wide class of models.

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.

Assumption 3.3 We assume that

$$\mathbb{E}\left(\sup_{0\le s\le t\le T}\sup_{\theta\in\Theta}\Big|\mu(X_t,\theta)-\mu(X_s,\theta)\Big|\right)^2 \le d_T(t-s)$$

for some constant d_T depending only upon T.

The condition in Assumption 3.3 is sufficient to make the approximation in (16) valid. The required uniform continuity in expectation is expected to hold for a wide class of diffusion type processes (X_t) if $\mu(\cdot,\theta)$ is Lipshcitz continuous uniformly in $\theta \in \Theta$. If (X_t) is generated as a diffusion given by $dX_t = \nu_t dt + \omega_t dW_t$ with standard Brownian motion (W_t), then we have

$$\mathbb{E}\left(\sup_{0\le s\le t\le T}\big|X_t - X_s\big|\right)^2 \le 2\left[(t-s)^2\left(\mathbb{E}\sup_{0\le t\le T}\nu_t^2\right) + (t-s)\left(\mathbb{E}\sup_{0\le t\le T}\omega_t^2\right)\right].$$

Consequently, if we have $\sup_{\theta\in\Theta}|\mu(x,\theta)-\mu(y,\theta)| \le c|x-y|$ for some constant $c > 0$, then the condition is satisfied for all $|t-s| \le 1$ with $d_T = c^2\left[\left(\mathbb{E}\sup_{0\le t\le T}\nu_t^2\right)\vee\left(\mathbb{E}\sup_{0\le t\le T}\omega_t^2\right)\right]$, where we use the notation $p \vee q = \max(p,q)$.

Assumption 3.4 We assume that $b_T/c_T = O(T)$, $c_T/(a_T^2 b_T) = O(T)$, $d_T/(a_T^2 b_T^2 c_T) = O(T^3)$, and

$$\delta = O\left(\frac{1}{T^{4+\varepsilon}b_T^3 c_T}\right)$$

for some $\varepsilon > 0$.

Assumption 3.4 introduces some conditions on the relative magnitudes of a_T, b_T, c_T and d_T and specifies how fast δ should decrease as T gets large. In usual applications, it appears that all of a_T, b_T, c_T and d_T either remain constant or grow very slowly. Consequently, the conditions we impose on their relative magnitudes are likely to hold widely. In fact, they are not essential and mostly expositional. They are introduced merely to simplify the required condition for δ. For the validity of our methodology and subsequent theory, it is necessary that we have at least $\delta = o(T^{-4})$. The condition may look stringent. However, for many economic and financial time series, observations are made and available at all levels of frequencies and we may even choose their frequencies to meet our needs. It is therefore not very restrictive. As discussed earlier, the required rate would be reduced if we may assume to know the value of θ_0 as in the case of hypothesis testing or use any iterated procedure based on a preliminary estimate for θ_0.

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

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

for all large N.

Lemma 3.4 allows us to use the approximated sample $(Z_i^\delta(\theta))$ instead of the unobservable estimation sample $(Z_i(\theta))$. Indeed, we will show subsequently that the resulting approximation error would not affect the asymptotic theory of our methodology.

Using the estimation samples $(Z_i^\delta(\theta))$ in (15), we now define

$$\Pi_N^\delta(z, \theta) = \frac{1}{N} \sum_{i=1}^N 1\{Z_{i_d}^\delta(\theta) \le z\} = \frac{1}{N} \sum_{i=1}^N 1\{Z_i^\delta(\theta) \le z_1\} \cdots 1\{Z_{i-d+1}^\delta(\theta) \le z_d\}. \tag{17}$$

Clearly Π_N^δ in (17) corresponds to Π_N introduced in (9).

Assumption 3.5 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 $1 \le i \le N$ and $\theta \in \Theta$.

Lemma 3.5 *Under* Assumptions 2.2 *and* 3.1 - 3.5 *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 3.5. Though it is difficult to check, the condition in Assumption 3.5 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

$$Q_N^\delta(\theta) = \int [\Pi_N^\delta(z,\theta) - \Pi(z,\theta_0)]^2 \varpi(dz), \tag{18}$$

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 3.6 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 $|\mu(x,\theta) - \mu(x,\theta_0)| \le \nu(x)\|\theta - \theta_0\|$ for a measurable real-valued function ν. 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 $i \ge 1$, and that $\sup_{i\ge 1} \mathbb{E}W_i^2 < \infty$.

We have

Lemma 3.6 *Under Assumption* 3.6, *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 3.7 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 3.6 - 3.7, we will establish that

$$\sqrt{N}(\hat{\theta}_N - \theta_0) = -2\ddot{Q}(\theta_0)^{-1}\sqrt{N}\int \dot{\Pi}(z)[\Pi_N(z,\theta_0) - \Pi(z,\theta_0)]\varpi(dz) + o_p(1) \tag{19}$$

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

$$\sqrt{N}[\Pi_N(\cdot,\theta_0) - \Pi(\cdot,\theta_0)] \to_d \Lambda(\cdot) \tag{20}$$

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

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

where we use the notation $p \wedge q = \min(p,q)$. Furthermore, it follows that $\Gamma_{-k}(x,y) = \Gamma_k(y,x)$.

It is easy to deduce from (19) and (20) that

Theorem 3.7 *Under* Assumptions 2.1 - 2.2 *and* 3.1 - 3.7, *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 $N \to \infty$.

The MGE is consistent with the convergence rate $\sqrt{N}$, which we would normally expect to be identical to $\sqrt{T}$.[10] As is well known, this is the convergence rate of the MLE for the drift term parameter in a fully specified stationary diffusion model.[11] Therefore, the MGE has the same rate of convergence as the MLE in this case. Moreover, the MGE has limit normal distribution. The proof of the asymptotic normality 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 procedure with various existing methods to select the subsample size can be applied 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 $(X_{j\delta}, Y_{j\delta})$, say, $(X_{j\delta}^*, Y_{j\delta}^*)$ for $j = 1, \ldots, M$. 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.[12] 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

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

using the subsampling or block bootstrap method. The dimension of the resampling is now significantly reduced from M 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

[10] More precisely, this would be the case if we have $a_T^{-1} = O(1)$ in Assumption 3.1 since $[U]_T = N\Delta \geq Ta_T$.

[11] In particular, the convergence rate is determined by the sampling span T, and not by the sample size, which is given by M in our notation.

[12] 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.

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.

As is well expected, the statistic

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

has the same limit distribution as its continuous version τ_N introduced in (11) under the null hypothesis of correct specification. Moreover, their limit distribution can be represented as a functional of the Gaussian process Λ defined in (20). This is shown in the following corollary.

Corollary 3.8 *Under* Assumptions 2.1 - 2.2 *and* 3.1 - 3.7, *we have*

$$\tau_N^\delta = \tau_N + o_p(1)$$

for all large N, *and*

$$\tau_N \to_d \int \Lambda^2(z)\varpi(dz) - 2\int\int \Lambda(x)\Lambda(y)\left[\dot{\Pi}(x)'\ddot{Q}(\theta_0)^{-1}\dot{\Pi}(y)\right]\varpi(dx)\varpi(dy)$$

as $N \to \infty$.

Therefore, we may use the statistic τ_N^δ to test for the correct specification of the martingale regression. The critical values of the test τ_N^δ can be obtained by the subsampling or bootstrap method discussed above.

3.4 Other Issues in Implementation

3.4.1 Jumps

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. To deal with the problem of discontinuity in the error process, we suggest to detect the presence of jumps using the test by Lee and Mykland (2008) as a preliminary step. 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. If jumps are generated exogenously, as is the case for virtually all economic and financial models used in practical applications, they do not include any information on the unknown parameters in the model and can be regarded as pure noise. In our approach, we do not attempt to extract useful information from the observations contaminated with noisy jumps, since it would necessarily require a rather complete and tight specification of the jump process in the model.

If we test for jumps and delete them in a preliminary step, the subsequent analysis of course will be made conditional on the preliminary jump test. This may result in increased variances for estimators and size distortions in tests, as in the case of other statistical procedures based on preliminary tests. To minimize the problem resulting from the reliance on a preliminary jump test, we recommend to use samples collected at relatively lower frequency such as daily rather than intra-day observations at ultra-high frequencies. It is well observed in many economic and financial time series that jumps are rare for samples at the frequencies of daily or lower, though they are frequently observed for many intra-day samples. Moreover, jumps tend to be more indistinguishable from continuous realizations as the frequency increases. The negative impact of relying on a preliminary test for jumps would certainly be larger if jumps are more frequent and have sizes not distinguishable from the realizations of the continuous part of the underlying model. If we use daily observations, the impact of a preliminary jump test appears to be insubstantial for models used in practical applications, where we expect that relatively large jumps occur intermittently.

3.4.2 Choice of Δ

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 $\hat{\theta}_N$ can be found analytically and 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.[13] With the choice of optimal Δ^* or N^*, we may readily show that the MGE has the asymptotic standard deviation $1.54\sqrt{2\kappa}$. As is well known, the MLE of κ has the asymptotic standard deviation $\sqrt{2\kappa}$, and therefore, the MGE has the asymptotic standard deviation that is 1.54 times bigger than the MLE. This, of course, does not imply that the finite sample behavior of the MLE is necessarily better than that of the MGE. Indeed, as we clearly show in our simulations, the MGE often performs significantly better than the MLE in finite samples. Therefore, there is a strong incentive to use the MGE even in the case that we have a fully specified likelihood function and the MLE is available.

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 original sample for each interval $[T_{(i-1)\Delta}^\delta, T_{i\Delta}^\delta]$. 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 the realized variance $[Y]_T^\delta$ of (Y_t) over the entire sampling period by a fixed value of N to find the corresponding

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

level of Δ. For instance, setting N to be the number of months in the sampling horizon, we may obtain $\Delta = (1/N)[Y]_T^\delta$ corresponding to the monthly observations.

3.4.3 Unobserved Covariates

In many continuous time conditional mean models used in economics and finance, some of the explanatory variables are not directly observable at any discrete time intervals and we necessarily have to use either their proxies or their estimates. In fact, it is quite common that the volatility processes representing the market and macroeconomic risks appear as covariates in continuous time asset pricing models derived from no arbitrage condition, and they are not directly observable unless their generating processes are observed in continuous time. Obviously, however, we may use the proxies or estimates $(\bar{X}_t)$ for the covariate (X_t) in our approach, as long as they are close enough each other. In what follows, we let $(\bar{Z}_i^\delta(\theta))$ be defined from $(\bar{X}_t)$ correspondingly as $(Z_i^\delta(\theta))$ in (15).

Assumption 3.8 We assume that
(a) $|\mu(x,\theta) - \mu(y,\theta)| \leq \nu(x)\|x-y\|$ and $\sup_{0\leq t\leq T} \nu(X_t) = O(e_T)$ with some constant e_T depending only upon T.
(b) $\mathbb{E}\left(\sup_{0\leq t\leq T}|X_t - \bar{X}_t|\right)^2 = O(\delta^p T^q)$ with $\delta = o\left((a_T^2/b_T e_T^2)^{1/2p} T^{-(2q+1)/2p}\right)$ for some constants $p > 0$ and $q \geq 0$.

Lemma 3.9 *Under* Assumptions 2.2, 3.1 - 3.2 *and* 3.8, *we have*

$$\mathbb{E} \max_{1\leq i\leq N} \sup_{\theta\in\Theta} \left|\bar{Z}_i^\delta(\theta) - Z_i^\delta(\theta)\right| = o(N^{-1/2}),$$

for all large N.

Lemma 3.9 implies that our methodology and its asymptotic theory remain to be valid if in particular Assumption 3.8 is satisfied.

The conditions in Assumption 3.8 are not stringent. Clearly, condition (a) is expected to be widely met, as long as the instantaneous conditional mean function $\mu(x,\theta)$ is differentiable with respect to x. In particular, it is trivially satisfied with $e_T = 1$, if $\mu(x,\theta)$ is linear and given by $\mu(x,\theta) = \theta' x$. We may also readily see that condition (b) holds for a broad class of continuous time models in economics and finance. In many cases, a more precise estimate or a better proxy for (X_t) becomes available as $\delta \to 0$, and therefore, it is reasonable to assume that the discrepancy between (X_t) and $(\bar{X}_t)$ decreases at the rate of δ^p with some $p > 0$. For example, as shown by Barndorff-Nielsen and Shephard (2002, 2004), Aït-Sahalia, Mykland and Zhang (2005) and Bandi and Russel (2008), we may estimate the unobserved integrated volatilities of general continuous time processes at the rate of $O_p(\delta^{1/2})$ over any fixed interval, if their generating processes are observed at time intervals of length δ. In general, the discrepancy between (X_t) and $(\bar{X}_t)$ over the time interval $[0,T]$ would increase as $T \to \infty$, the rate of which we assume to be given by T^q for some $q \geq 0$.

Once again, we only require that $\delta > 0$ be small relative to T here. Moreover, though we do not make it explicit in the paper for the simplicity and clarity of our exposition, it is quite obvious that our theory can be easily extended to accommodate the possibility that Y and X are observed at different sampling frequencies. In particular, we may readily allow for the proxies and estimates of (X_t) to be observed at lower frequencies than (Y_t). The only essential requirement is that the sampling interval should be sufficiently small in comparison with the time span of sample. For instance, we may even use macroeconomic time series observed at monthly or quarterly intervals to obtain measures of macroeconomic fluctuations and study their effects on asset prices, which are observed much more frequently, if the sampling horizon for macroeconomic time series is long enough.

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 our estimator. For our purpose, we consider two continuous time models. The first model, Model I, is the Feller's square-root process given by

$$dX_t = (\alpha + \beta X_t)dt + \gamma\sqrt{X_t}dW_t, \tag{21}$$

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

$$dY_t = (\alpha + \beta X_t)dt + \sqrt{X_t}dW_t \tag{22}$$

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 earlier by Heston (1993) to model stochastic volatility. The MGE is employed to estimate the parameters α and β of Models I and II in (21) and (22).

For comparison, we also consider the estimators based on the ML approach for Models I and II. For Model I, we employ the exact MLE method to obtain the estimates for the parameters α and β. In contrast, for Model II, we use the approximation MLE method proposed recently by Aït-Sahalia and Kimmel (2007). In what follows, however, we do not distinguish the two methods and simply call both of the resulting estimators the MLE's. Needless to say, the MLE's are expected to perform better for our simulation models, since they use more information. In both Models I and II, the MLE uses information on the precise structure of the process generating stochastic volatility, whereas the MGE ignores any structural information on stochastic volatility and only treats it nonparametrically. Of course, the MLE is subject to the potential problem of misspecification of volatility process.[14] This problem, however, will not be investigated in the paper, since our objective

[14]Through simulations we find that the effect of misspecification can be substantial, especially when the underlying model is nearly nonstationary.

is to evaluate the performance of the MGE, compared with the MLE applied to the correctly specified model.

To study the effect of preliminary jump test on the performance of the MGE, we consider the model defined as

$$dY_t = dX_t + dJ_t, \tag{23}$$

where (X_t) is the Feller's square-root process introduced above in (21) and (J_t) is a jump process independent of (X_t). We specify the jump process (J_t) as a compound Poisson process which is given by

$$J_t = \sum_{k=1}^{N_t} D_k,$$

where (N_t) is a Poisson process with intensity $\lambda > 0$ and (D_k) is a sequence of independent and identically distributed uniform random variables that is independent of (N_t). In what follows, (23) will be referred to as Model III. For Model III, we apply the MGE after we identify and discard observations contaminated with jumps using the test developed by Lee and Mykland (2008). In our simulation, the performance of the MGE for Model III implemented with the preliminary test for jumps is evaluated against that of the MGE for Model I with no jumps.

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 December 31, 2009, with the sample size $M = 13,989$. The observed rates were zero on December 10, 18 and 24, 2008, which were discarded since they are not compatible with the model. Furthermore, we delete the observations that are believed to be contaminated by jumps and cannot be generated by our model. To detect for jumps, we use the test by Lee and Mykland (2008). The test results are somewhat dependent upon the size of window, so we tried various lengths of window ranging from 8 to 96 days.[15] For the 5% level test, the numbers of jumps that the test detect are 335, 113, 77, 73, 76, 74 and 69 corresponding to the window sizes 8, 16, 32, 48, 64, 80 and 96. In our empirical analysis, we set the window size to be 48, and after deleting the observations corresponding to the detected jumps the actual number of daily observations we use to estimate the model is 13,913.

For Model II, we use the log of 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), following Aït-Sahalia and Kimmel (2007) which uses the VIX as a proxy for the latent volatility process.[16] The daily observations from January 2, 1990 to December 31, 2009 are initially downloaded for our estimation. However, the VIX is not available on March 1,

[15]Lee and Mykland (2008) suggest to use the window size 16 for daily observations. We try various window lengths here to check the sensitivity of our empirical results with respect to the window size. In fact, the choice of window size has virtually no effect on our estimates for both the MLE and MGE.

[16]The deleterious effect on the parameter estimation of using the VIX as a proxy for the latent factor representing the volatility state is beyond the scope of our paper, and will not be investigated in our empirical analysis.

1991, January 31, 1997 and November 26, 1997, and the corresponding observations of the SPX are also deleted leaving the daily observations of size $M = 5,040$. Once again, we use the test by Lee and Mykland (2008) with size 5% to detect the jumps. The test detects 38, 20, 12, 9, 11, 12 and 12 jumps for the SPX, and 76, 51, 45, 44, 45, 50 and 52 for the squared VIX, respectively for the selection of window size 8, 16, 32, 48, 64, 80 and 96. We choose the window size 32, for which we detect 49 jumps. Some of the jumps in the SPX and VIX are overlapped. Once we delete all observations associated with the detected jumps, there are 4,991 observations that we use to estimate the model.

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

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

with $M_{i-1} = \delta^{-1} T^{\delta}_{(i-1)\Delta}$, sequentially for $i = 1, \ldots, N$ with any fixed $\Delta > 0$. Furthermore, we define the estimation sample by

$$Z^{\delta}_i(\theta) = \Delta_i^{-1/2} \left(Y_{T^{\delta}_{i\Delta}} - Y_{T^{\delta}_{(i-1)\Delta}} - \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta}, \theta) \right),$$

where $\Delta_i = \sum_{j=M_{i-1}+1}^{M_i} (Y_{j\delta} - Y_{(j-1)\delta})^2$. Note that the estimation sample defined here is normalized by the actual realized variance, rather than the target level of increment in quadratic variation. Obviously, our actual definitions of time change and estimation sample here would not have any effect on the development of our asymptotic theory. It seems, however, that the performance of the MGE improves slightly in finite samples.

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 $\Delta \geq (20/M)[Y]^{\delta}_{T}$. 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. For Model I, we find that $\Delta = 8.76 \times 10^{-5}$ minimizes the standard error of the MGE for β, which gives us $N = 143$. This is true for both the 1-dimensional and 2-dimensional MGE's. For Model II, the optimal values of Δ are given by $\Delta = 3.16 \times 10^{-2}$ and $\Delta = 2.21 \times 10^{-2}$, which yields $N = 20$ and $N = 29$, respectively for the 1-dimensional and 2-dimensional MGE's.[17]

[17]For both models, the standard errors of the estimates vary somewhat irregularly as we change the values of Δ. In particular, we do not have the U-shaped standard errors with the unique minimum at some optimal value of Δ.

Table 1: Estimation Results for Models I and II

Model I

	MLE	MGE 1D	MGE 2D
α	0.0021	0.0048	0.0040
	N/A	(0.0036)	(0.0028)
β	–0.0339	–0.0720	–0.0612
	N/A	(0.0861)	(0.0905)

Model II

	MLE	MGE 1D	MGE 2D
α	–0.4758	0.0809	0.1221
	(0.0608)	(0.0762)	(0.0673)
β	11.1217	–0.6297	–1.5127
	(1.6311)	(1.0334)	(1.0491)

In parenthesis are the estimated standard errors for the estimates. For the MGE, 1D and 2D refer to the 1-dimensional and 2-dimensional MGE's respectively.

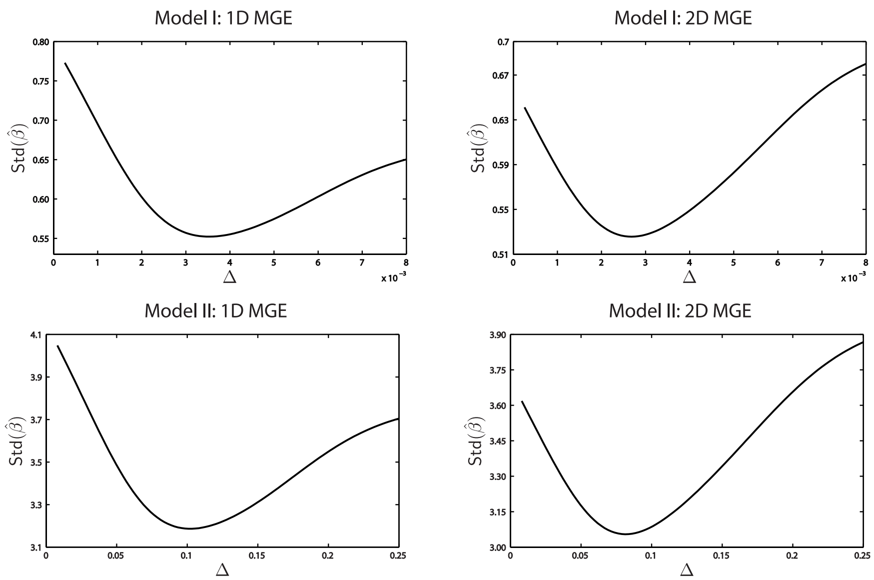
In Table 1, we report the estimates of the parameters α and β in Models I and II with their estimated standard errors. For comparison, we also present the MLE's as well as the MGE's.[18] For Model I, the MLE and MGE yield roughly comparable estimates for both parameters α and β. In particular, we do not observe any significant differences between their estimates. Moreover, the discrepancies in the estimates based on the 1-dimensional and 2-dimensional MGE's are small and seem to be statistically insignificant. We cannot report the standard errors of the MLE's, because their values are negative and nonsensical. For Model II, the two estimates given by the MLE and MGE are quite distinctive. The MLE's seem insensible, and we expect α to be in opposite sign and β to be smaller. In contrast, the MGE's appears to be in reasonable ranges for both parameters α and β. The reader is referred to Aït-Sahalia and Kimmel (2007) for more discussions on the model and what these parameters represent. The MLE and MGE yield the estimated standard errors that are largely of the same magnitude.

4.2 Monte Carlo Simulations

For the benchmark model we use in our simulations of Model I, we set the parameter values $(\alpha, \beta, \gamma^2) = (\alpha_0, \beta_0, \gamma_0^2)$ with $\alpha_0 = 0.01579, \beta_0 = -0.219$ and $\gamma_0^2 = 0.06665^2$, which were obtained earlier by Aït-Sahalia (1999) using the exact MLE for the three-months T-bill rates. Other parameter values are also considered to see how sensitive our results are to the choice of parameter values. In particular, we report the results from the variant models with parameter values $(\alpha, \beta, \gamma^2) = (\alpha_0/10, \beta_0/10, \gamma_0^2/10)$ and $(\alpha, \beta, \gamma^2) = (10\alpha_0, 10\beta_0, 10\gamma_0^2)$, as well as our benchmark model. As is well known, the Feller's square-root process introduced in (21) has time-invariant marginal distribution that is given by Gamma with parameters $2\alpha/\gamma^2$ and $-2\beta/\gamma^2$, if $2\alpha/\gamma^2 \geq 1$ and $\beta < 0$. Therefore, our variant models have the same time-invariant marginal distributions as the benchmark model. However, they generate

[18] We employ the simplex method provided by the `fminsearch` function of Matlab to find the MLE's and MGE's reported in the table, using as the initial values the exact MLE obtained by Aït-Sahalia (1999) for Model I, and the parameter values used in the simulation model of Aït-Sahalia and Kimmel (2007) for Model II.

Figure 2: Optimal Selection of Δ



The figure shows the plot of standard deviation of the MGE for the parameter β as a function of Δ. The 1-dimensional and 2-dimensional MGE's are considered for the second variant version of Model I and the benchmark version of Model II.

processes with differing levels of persistency, as they have distinctive values for the mean-reversion parameter β. The first variant model generates a more persistent process, while the second variant model generates a less persistent process, than the benchmark model.

To simulate Model II, we use for our benchmark model the same parameter values as those used by Aït-Sahalia and Kimmel (2007) in their simulations of the model, which are given by $(\alpha, \beta, \rho) = (0.025, 0.94, -0.8)$ and $(\mu, \nu, \omega^2) = (\mu_0, \nu_0, \omega_0^2)$ with $\mu_0 = 0.3, \nu_0 = -3$ and $\omega_0^2 = 0.25^2$. Note that our model satisfies the required conditions $2\mu/\omega^2 \geq 1$ and $\nu < 0$ for the stationarity of volatility process. As in our simulations of Model I, we also consider the variant models defined with the same values of (α, β, ρ), and $(\mu, \nu, \omega^2) = (\mu_0/10, \nu_0/10, \omega_0^2/10)$ and $(\mu, \nu, \omega^2) = (10\mu_0, 10\nu_0, 10\omega_0^2)$. The volatility processes in our variant models for Model II defined in (22) therefore have the same time-invariant marginal distributions as the benchmark model. However, they are more and less persistent, respectively for the first and second variant models, than for the benchmark model. Though we also look at the models with different values of parameter β in our simulations, we do not report the details of their results since they are qualitatively the same. The performances of both the MLE and MGE do not depend much on the value of parameter β in Model II.

As in our empirical illustrations, we numerically search the optimal Δ for each of the benchmark and variant models. The actual search is made over the 100 equi-spaced grid points in the interval, the left and right end points of which are given respectively by the expected values of the 20 days of quadratic variation and one-twentieth of the total quadratic variation of the error process. The values of the end points are obtained analytically using the parameter values of our simulation models. Consequently, on average, each sampling interval for the estimation samples has at least 20 observations, and the size of the estimation samples is no less than 20. In some cases, the optimal Δ is nicely and uniquely defined within the interval. For the purpose of illustration, we present some of these cases in Figure 2, which plots the standard deviation of the MGE of the parameter β as a function of Δ for the second variant model of Model I and the benchmark model of Model II. In other cases, it is either less conspicuous or given at one of the end points of the interval. Roughly speaking, the optimal Δ tends to be more clearly defined as the sampling horizon gets larger and as the underlying process becomes more stationary.

The simulation results are provided in Tables 2 and 3, respectively, for Models I and II. In the tables, we report the bias, standard deviation (Std) and root mean squared error (RMSE) of the MLE and MGE of the parameter β. The reported results are based on 5,000 iterations.[19] For Model I, the simulation samples are generated as $1,260$, $5,040$ and $12,600$ daily observations, corresponding respectively to 5, 20 and 50 years of sampling horizon, using the exact transition starting from the initial stationary distribution. The exact transition of the process generated by Model II is unknown. Therefore, we follow Aït-Sahalia and Kimmel (2007) and use the Euler discretization to generate 30 intra-day samples per day starting from the initial values of 0.1 and $\ln 100$, and collect the required daily observations of sizes $1,260$, $5,040$, and $12,600$ again for each of the sampling horizons, 5, 20 and 50 years, after discarding the initial 500 daily observations.

The overall performance of the MGE is at least comparable to that of the MLE for all models considered in our simulations. In particular, the ML approach which fully uses the structure of the model does not provide any noticeable improvement in finite samples. We may notice some tendency that the MLE does relatively better than the MGE as the sampling horizon increases. However, for the normal range of sampling horizons that we encounter in most practical applications, it is not likely that we see any evidence of the relative efficiency of the MLE in asymptotic theory. As is well known, the MLE is subject to some severe bias. The bias of the MLE is unacceptably huge especially when the sampling horizon is small or only moderately large. The problem mitigates as the sampling horizon increases, yet the magnitude of bias is far away from being negligible even when the sampling horizon is as large as 50 years, longer than the most of empirical studies in the literature. In sharp contrast, the bias of the MGE is generally quite small and becomes truly negligible in many cases. This is so, even when the sampling horizon is as small as 5 years.

At least for the models we consider in our simulations, we cannot find any evidence that the 2-dimensional MGE does better than the 1-dimensional MGE. They are largely comparable across all models and all sampling horizons used in our simulations. We may see some evidence that the performance of the 2-dimensional MGE tends to be better than that of the 1-dimensional MGE in terms of RMSE, as the sampling horizon increases. This is well expected, since the asymptotic variance of the 2-dimensional MGE is smaller than that of the 1-dimensional MGE. However, the relative advantage of the 2-dimensional MGE

[19] As for the estimation results in Table 1, we employ the simplex method provided by the `fminsearch` of Matlab for both the MLE and MGE using the true parameter values as the initial values.

over the 1-dimensional MGE does not seem to be significant in any realistic setup. Finally, for both the MLE and MGE, the performances of estimators improve more fastly as the sampling horizon increases as the underlying models become more nonstationary. For the first variant models of both Models I and II, we consistently have faster improvements of the finite sample performances as the sampling horizon increases, compared with the corresponding second variant models. Recall that the formers generate processes that are less mean-reverting in mean and volatility than the latters.

We also examine the finite sample performance of the MGE in the presence of jumps. For our simulation, we set the jump intensity and the distribution of jump size to be largely comparable to our results from the jump test for the daily T-bill rates that we use to fit Model I. For the jump intensity, we let $\lambda = 1$ and 1.5, which correspond respectively to the estimated intensities from the 10% and 1% jump tests.[20] Likewise, we let the distribution of jump size be uniform and symmetrical around the origin taking values between the 25th and 100th percentiles in modulus of the daily differentials of T-bill rates tested positive for jumps, net of their fitted drifts.[21] As expected, the performance of the MGE deteriorates as we include jumps in our model. However, as we may see clearly in Table 4, the impact of jumps on the performance of the MGE is insignificant at least for the jumps we normally encounter in daily observations. The performance of the MGE tends to depend upon that of the jump test used in the preliminary step, which is in turn determined by the size and intensity of jumps. In general, the 1-dimensional MGE is more robust to the presence of jumps, compared to the 2-dimensional MGE.

5. Conclusion

In the paper, we consider the general conditional mean model in continuous time and develop a methodology for the statistical inference to analyze the model. Our framework is quite flexible and accommodates a wide class of continuous time conditional mean models, including general continuous time asset pricing models derived from no arbitrage condition and diffusion models as special cases. In particular, we do not impose any restriction on the error process, allowing for arbitrage forms of stochastic volatilities that are endogenous and persistent. For the identification and estimation of our model, we only use the condition that the error process is a martingale. No other conditions, such as the orthogonality conditions used commonly in conventional approach, are not required. The estimator derived from our approach, the MGE, is consistent and asymptotically normal under general regularity conditions. Overall, it performs quite well in finite samples, at least comparably with the MLE which uses full information on the model, and it greatly improves upon the MLE in terms of finite sample bias. Unlike the MLE, the bias of the MGE is quite acceptable even when the sampling horizon is small.

As discussed, we do not consider and utilize any orthogonality conditions implied by our

[20] Note that λ denotes the expected number of jumps in one year.

[21] In specifying the distribution of jump size, we exclude the daily differentials of the jump contaminated T-bill rates net of their fitted drifts up to their 25th percentile in modulus, since they appear to be indistinguishable from the realizations of the continuous diffusion term in our model.

model. It would certainly be possible to combine our martingale condition with any available orthogonality conditions, whenever we have proper instruments. There are several ways to jointly implement the orthogonality conditions together with the martingale condition. One obvious way is to jointly consider the two conditions with appropriate relative weights. It is also possible that we may derive some moment conditions implied by our martingale condition, and implement them jointly with other moment conditions given by available orthogonality conditions. For instance, the estimation samples obtained after time change in the same way as in the paper have zero mean and unit standard deviation, which can be added as additional moment conditions to other existing moment conditions. This approach, which relies exclusively on moment conditions, does not fully utilize all the aspects of the martingale condition. However, it has some obvious advantages, since it does not require the error process to be continuous martingale and permits the presence of jumps. Some of the researches along this line are underway.

Mathematical Appendix

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

$$(U_t - U_s)^2 - [U]_{s,t} = 2\int_s^t (U_r - U_s)dU_r \tag{24}$$

due to Ito's formula, and therefore,

$$\mathbb{E}\left((U_t - U_s)^2 - [U]_{s,t}\right)^2 = 4\mathbb{E}\int_s^t (U_r - U_s)^2 d[U]_r. \tag{25}$$

Moreover, we have

$$\begin{aligned}
\mathbb{E}\int_s^t (U_r - U_s)^2 d[U]_r &\leq b_T \mathbb{E}\int_s^t (U_r - U_s)^2 dr \\
&= b_T \int_s^t \left(\mathbb{E}U_r^2 - \mathbb{E}U_s^2\right) dr \\
&= b_T \int_s^t \left(\mathbb{E}[U]_r - \mathbb{E}[U]_s\right) dr \\
&\leq b_T^2 \int_s^t (r - s) dr \\
&\leq \frac{1}{2} b_T^2 (t - s)^2,
\end{aligned} \tag{26}$$

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 (25) and (26) that

$$\mathbb{E}\left((U_t - U_s)^2 - [U]_{s,t}\right)^2 \leq 2b_T^2 (t - s)^2 \tag{27}$$

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

We consider

$$[U]_t - [U]_t^\delta = [U]_t - \sum_{k=1}^{j-1}(U_{k\delta} - U_{(k-1)\delta})^2$$
$$= \Big([U]_t - [U]_{(j-1)\delta}\Big) - \left(\sum_{k=1}^{j-1}\Big((U_{k\delta} - U_{(k-1)\delta})^2 - [U]_{(k-1)\delta,k\delta}\Big)\right) \qquad (28)$$

for $(j-1)\delta \le t < j\delta$, $j = 1, \dots, M$. We have

$$\max_{1\le j\le M} \sup_{(j-1)\delta\le t<j\delta} \Big|[U]_t - [U]_{(j-1)\delta}\Big| \le \delta b_T \qquad (29)$$

due to Assumption 3.1. Moreover,

$$(U_{j\delta} - U_{(j-1)\delta})^2 - [U]_{(j-1)\delta,j\delta} = 2\int_{(j-1)\delta}^{j\delta}(U_t - U_{(j-1)\delta})dU_t,$$

and

$$\sum_{k=1}^{j}\Big((U_{k\delta} - U_{(k-1)\delta})^2 - [U]_{(k-1)\delta,k\delta}\Big)$$

is a discrete time martingale, as a sequence in j, with respect to the filtration $(\mathcal{F}_{j\delta})$. Therefore, we may use Doob's L^p-inequality [see, e.g., Revuz and Yor (1994, p.52)] and (27) to have

$$\mathbb{E}\left(\max_{1\le j\le M}\sum_{k=1}^{j-1}\Big((U_{k\delta} - U_{(k-1)\delta})^2 - [U]_{(k-1)\delta,k\delta}\Big)\right)^2$$
$$\le 4\mathbb{E}\left(\sum_{j=1}^{M}\Big((U_{j\delta} - U_{(j-1)\delta})^2 - [U]_{(j-1)\delta,j\delta}\Big)\right)^2$$
$$= 4\sum_{j=1}^{M}\mathbb{E}\Big((U_{j\delta} - U_{(j-1)\delta})^2 - [U]_{(j-1)\delta,j\delta}\Big)^2$$
$$\le 4M\left(2\delta^2 b_T^2\right) = O\left(\delta T b_T^2\right). \qquad (30)$$

Consequently, it follows from (28), (29) and (30) that

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

as was to be shown. □

Proof of Lemma 3.2 Write

$$[Y]_t^\delta = [U]_t^\delta + A_t + 2B_t, \tag{31}$$

where

$$A_t = \sum_{j\delta \le t} \left(\int_{(j-1)\delta}^{j\delta} \mu(X_t, \theta) dt \right)^2$$
$$B_t = \sum_{j\delta \le t} \left(\int_{(j-1)\delta}^{j\delta} \mu(X_t, \theta) dt \right) (U_{j\delta} - U_{(j-1)\delta}).$$

We have

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

and therefore,

$$\mathbb{E} \left(\sup_{0\le t\le T} A_t \right)^2 = O\left((\delta T)^2 c_T^2 \right), \tag{32}$$

due to Assumption 3.2.

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

$$|B_t|^2 = \left| \sum_{j\delta \le t} \left(\int_{(j-1)\delta}^{j\delta} \mu(X_t, \theta) dt \right) (U_{j\delta} - U_{(j-1)\delta}) \right|^2$$
$$\le \left[\sum_{j\delta \le t} \left(\int_{(j-1)\delta}^{j\delta} \mu(X_t, \theta) dt \right)^2 \right] \left[\sum_{j\delta \le t} (U_{j\delta} - U_{(j-1)\delta})^2 \right].$$

Therefore, we have

$$\sup_{0\le t\le T} |B_t|^2 \le \left[\sum_{j=1}^{M} \left(\int_{(j-1)\delta}^{j\delta} \mu(X_t, \theta) dt \right)^2 \right] \left[\sum_{j=1}^{M} (U_{j\delta} - U_{(j-1)\delta})^2 \right]$$
$$\le \left(\sup_{0\le t\le T} \sup_{\theta\in\Theta} |\mu(X_t, \theta)| \right)^2 (M\delta^2)(Tb_T),$$

from which it follows immediately that

$$\mathbb{E} \left(\sup_{0\le t\le T} |B_t| \right)^2 = O\left(\delta T^2 (b_T c_T) \right). \tag{33}$$

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

Now we have from (31), (32) and (33) that

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

and the proof is complete. □

Proof of Corollary 3.3 Let $R_{\delta,T}$ be defined as in (13). Then we have

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

and therefore,

$$\left|T_t^\delta - T_t\right| \le \left|T_{t+R_{\delta,T}} - T_{t-R_{\delta,T}}\right| \tag{34}$$

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

However, it follows from Assumption 3.1 that

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

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

$$|T_t - T_s| \le a_T^{-1}(t-s) \tag{35}$$

for all $0 \le s \le t \le S$. Consequently, we may easily deduce from (34) and (35) that

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

and that

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

from which the stated result follows immediately. □

Proof of Lemma 3.4 Let

$$K_{\delta,T} = \delta T^2(b_T c_T).$$

We have $b_T/c_T = O(T)$ and $\delta = o(T^{-4}c_T^{-1})$ under Assumption 3.4, from which it follows that

$$\delta T b_T^2 = O\Big(\delta T^2(b_T c_T)\Big) = O(K_{\delta,T}), \quad (\delta T)^2 c_T^2 = o\Big(\delta T^2(b_T c_T)\Big) = o(K_{\delta,T}).$$

For $R_{\delta,T}^2$ defined in (13), we therefore have

$$\mathbb{E}R_{\delta,T}^2 = O\Big(\delta T^2(b_T c_T)\Big) = O(K_{\delta,T}). \tag{36}$$

It also follows from the conditions $c_T/(a_T^2 b_T) = O(T)$ and $\delta = o(T^{-4} b_T^{-3} c_T^{-1})$ in Assumption 3.4 that

$$\frac{c_T^{1/2}}{a_T}\Big(\mathbb{E}R_{\delta,T}^2\Big)^{1/2} = O\left(\frac{c_T^{1/2}}{a_T}\Big(\delta T^2(b_T c_T)\Big)^{1/2}\right) = o\left(\Big(\delta T^2(b_T c_T)\Big)^{1/4}\right) = o(K_{\delta,T}^{1/4}). \qquad (37)$$

Moreover, since $d_T/(a_T^2 b_T^2 c_T) = O(T^3)$ and $\delta = o(T^{-4} b_T^{-3} c_T^{-1})$ as given in Assumption 3.4, we may easily deduce that

$$\frac{\delta^{1/2} d_T^{1/2}}{a_T} = o\left(\Big(\delta T^2(b_T c_T)\Big)^{1/4}\right) = o(K_{\delta,T}^{1/4}). \qquad (38)$$

Finally, if $\delta = O(T^{-4-\varepsilon} b_T^{-3} c_T^{-1})$ for some $\varepsilon > 0$ as assumed in Assumption 3.4, then we have

$$K_{\delta,T}^{1/4-\varepsilon} = o\Big((Tb_T)^{-1/2}\Big) = o(N^{-1/2}) \qquad (39)$$

for some small enough $\varepsilon > 0$. Note that $N\Delta \leq Tb_T$ and Δ is fixed.

We have

$$\begin{aligned}\sqrt{\Delta}\left|Z_i^\delta(\theta) - Z_i(\theta)\right| \leq& \left|\Big(Y_{T_{i\Delta}^\delta} - Y_{T_{(i-1)\Delta}^\delta}\Big) - \Big(Y_{T_{i\Delta}} - Y_{T_{(i-1)\Delta}}\Big)\right| \\ &+ \left|\int_{T_{(i-1)\Delta}^\delta}^{T_{i\Delta}^\delta} \mu(X_t,\theta)dt - \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \mu(X_t,\theta)dt\right| \\ &+ \left|\int_{T_{(i-1)\Delta}^\delta}^{T_{i\Delta}^\delta} \mu(X_t,\theta)dt - \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta},\theta)\right|. \qquad (40)\end{aligned}$$

Subsequently, we consider each of the three terms in (40) separately. First, we establish

$$\mathbb{E}\max_{1\leq i\leq N}\sup_{\theta\in\Theta}\left|\int_{T_{(i-1)\Delta}^\delta}^{T_{i\Delta}^\delta} \mu(X_t,\theta)dt - \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \mu(X_t,\theta)dt\right| = o(N^{-1/2}). \qquad (41)$$

To prove (41), it suffices to show that

$$\mathbb{E}\sup_{0\leq t\leq S}\sup_{\theta\in\Theta}\left|\int_0^{T_t^\delta} \mu(X_s,\theta)ds - \int_0^{T_t} \mu(X_s,\theta)ds\right| = o(N^{-1/2}), \qquad (42)$$

since

$$\begin{aligned}&\left|\int_{T_{(i-1)\Delta}^\delta}^{T_{i\Delta}^\delta} \mu(X_t,\theta)dt - \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \mu(X_t,\theta)dt\right| \\ &\qquad \leq 2\max_{1\leq i\leq N}\left|\int_0^{T_{i\Delta}^\delta} \mu(X_t,\theta)dt - \int_0^{T_{i\Delta}} \mu(X_t,\theta)dt\right|\end{aligned}$$

for all $1 \leq i \leq N$ and $\theta \in \Theta$. Note that

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

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

$$\begin{aligned} \mathbb{E} \sup_{0 \leq t \leq S} \sup_{\theta \in \Theta} & \left| \int_0^{T_t^\delta} \mu(X_s, \theta) ds - \int_0^{T_t} \mu(X_s, \theta) ds \right| \\ & \leq \left[\mathbb{E} \left(\sup_{0 \leq t \leq T} \sup_{\theta \in \Theta} |\mu(X_t, \theta)| \right)^2 \right]^{1/2} \left[\mathbb{E} \left(\sup_{0 \leq t \leq S} \left| T_t^\delta - T_t \right| \right)^2 \right]^{1/2} \\ & \leq O(c_T^{1/2}) \frac{\left(\mathbb{E} R_{\delta,T}^2 \right)^{1/2}}{a_T} \\ & = o(K_{\delta,T}^{1/4}), \end{aligned} \tag{43}$$

from Cauchy-Schwarz inequality, Assumption 3.2, Corollary 3.3 and (37). Now we may easily deduce (42) from (43) and (39).

Second, we show that

$$\mathbb{E} \max_{1 \leq i \leq N} \left| \left(Y_{T_{i\Delta}^\delta} - Y_{T_{(i-1)\Delta}^\delta} \right) - \left(Y_{T_{i\Delta}} - Y_{T_{(i-1)\Delta}} \right) \right| = o(N^{-1/2}), \tag{44}$$

for which it suffices to prove that

$$\mathbb{E} \sup_{0 \leq t \leq S} \left| Y_{T_t^\delta} - Y_{T_t} \right| = o(N^{-1/2}), \tag{45}$$

since

$$\left| \left(Y_{T_{i\Delta}^\delta} - Y_{T_{(i-1)\Delta}^\delta} \right) - \left(Y_{T_{i\Delta}} - Y_{T_{(i-1)\Delta}} \right) \right| \leq 2 \max_{1 \leq i \leq N} \left| Y_{T_{i\Delta}^\delta} - Y_{T_{i\Delta}} \right|$$

for all $1 \leq i \leq N$. To establish (45), we note that

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

for all $0 \leq t \leq S$. Since it follows immediately from (42) that

$$\mathbb{E} \sup_{0 \leq t \leq S} \left| \int_0^{T_t^\delta} \mu(X_s, \theta_0) ds - \int_0^{T_t} \mu(X_s, \theta_0) ds \right| = o(N^{-1/2}), \tag{46}$$

we only need to prove

$$\mathbb{E} \sup_{0 \leq t \leq S} \left| U_{T_t^\delta} - U_{T_t} \right| = o(N^{-1/2}) \tag{47}$$

to establish (45).

Note that

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

i.e.,

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

for all $0 \le t \le S$. It follows that

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

for all $0 \le t \le S$. Therefore, we may deduce that

$$\sup_{0\le t\le S} \left| U_{T_t^\delta} - U_{T_t} \right| \le \sup_{0\le t\le S} \sup_{|s-t|\le R_{\delta,T}} |V_t - V_s| \le \Big(2R_{\delta,T}\Big)^{1/2-\varepsilon}$$

for any $\varepsilon > 0$, due to the Lévy's modulus of continuity of Brownian motion [see, for instance, Karatzas and Shreve (1988, p.114) and Kanaya (2008)]. Consequently, it follows from (36) that

$$\mathbb{E} \sup_{0\le t\le S} \left| U_{T_t^\delta} - U_{T_t} \right| \le \mathbb{E}\Big(2R_{\delta,T}\Big)^{1/2-\varepsilon} = O\left(\left(\mathbb{E} R_{\delta,T}^2 \right)^{(1-\varepsilon)/4} \right) = O(K_{\delta,T}^{1/4-\varepsilon})$$

for any $\varepsilon > 0$, from which and (39) we may readily establish (47). As explained above, (45) follows immediately from (46) and (47).

Finally, we show that

$$\mathbb{E} \max_{1\le i\le N} \sup_{\theta\in\Theta} \left| \int_{T_{(i-1)\Delta}^\delta}^{T_{i\Delta}^\delta} \mu(X_t,\theta)dt - \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta},\theta) \right| = o(N^{-1/2}). \tag{48}$$

Note that we have

$$\begin{aligned}
&\max_{1\le i\le N} \sup_{\theta\in\Theta} \left| \int_{T_{(i-1)\Delta}^\delta}^{T_{i\Delta}^\delta} \mu(X_t,\theta)dt - \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta},\theta) \right| \\
&\le \max_{1\le i\le N} \sum_{j=M_{i-1}+1}^{M_i} \int_{(j-1)\delta}^{j\delta} \sup_{\theta\in\Theta} \Big| \mu(X_t,\theta) - \mu(X_{j\delta},\theta) \Big| dt \\
&\le \delta \left(\max_{1\le i\le N} |M_i - M_{i-1}| \right) \left(\sup_{\substack{|t-s|\le\delta \\ 0\le s,t\le T}} \sup_{\theta\in\Theta} \Big| \mu(X_t,\theta) - \mu(X_s,\theta) \Big| \right) \\
&= \left(\max_{1\le i\le N} \Big| T_{i\Delta}^\delta - T_{(i-1)\Delta}^\delta \Big| \right) \left(\sup_{\substack{|t-s|\le\delta \\ 0\le s,t\le T}} \sup_{\theta\in\Theta} \Big| \mu(X_t,\theta) - \mu(X_s,\theta) \Big| \right),
\end{aligned}$$

and therefore,

$$\begin{aligned}
&\mathbb{E} \max_{1\le i\le N} \sup_{\theta\in\Theta} \left| \int_{T^\delta_{(i-1)\Delta}}^{T^\delta_{i\Delta}} \mu(X_t,\theta)dt - \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta},\theta) \right| \\
&\le \left[\mathbb{E}\left(\max_{1\le i\le N} \left| T^\delta_{i\Delta} - T^\delta_{(i-1)\Delta} \right| \right)^2 \right]^{1/2} \left[\mathbb{E}\left(\sup_{\substack{|t-s|\le\delta \\ 0\le s,t\le T}} \sup_{\theta\in\Theta} \left| \mu(X_t,\theta) - \mu(X_s,\theta) \right| \right)^2 \right]^{1/2} \\
&\le \delta^{1/2} d_T^{1/2} \left[\mathbb{E}\left(\max_{1\le i\le N} \left| T^\delta_{i\Delta} - T^\delta_{(i-1)\Delta} \right| \right)^2 \right]^{1/2}, \qquad (49)
\end{aligned}$$

due in particular to Assumption 3.3. However, we have

$$\max_{1\le i\le N} \left| T^\delta_{i\Delta} - T^\delta_{(i-1)\Delta} \right| \le \max_{1\le i\le N} \left| T_{i\Delta} - T_{(i-1)\Delta} \right| + 2 \max_{1\le i\le N} \left| T^\delta_{i\Delta} - T_{i\Delta} \right|,$$

and therefore,

$$\mathbb{E}\left(\max_{1\le i\le N} \left| T^\delta_{i\Delta} - T^\delta_{(i-1)\Delta} \right| \right)^2 = O\left(\frac{1}{a_T^2}\right) + O\left(\frac{\mathbb{E}R^2_{\delta,T}}{a_T^2}\right) = O\left(\frac{1}{a_T^2}\right). \qquad (50)$$

Consequently, it follows from (49) and (50) that

$$\mathbb{E} \max_{1\le i\le N} \sup_{\theta\in\Theta} \left| \int_{T^\delta_{(i-1)\Delta}}^{T^\delta_{i\Delta}} \mu(X_t,\theta)dt - \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta},\theta) \right| = O\left(\delta^{1/2} \frac{d_T^{1/2}}{a_T} \right),$$

and we have (48) from (38) and (39). The stated result now follows immediately from (41), (44) and (48), and the proof is complete. □

Proof of Lemma 3.5 We have

$$\left| \Pi^\delta_N(z,\theta) - \Pi_N(z,\theta) \right| \le \frac{1}{N} \sum_{i=1}^N \left| 1\{Z^\delta_{i_d}(\theta) \le z\} - 1\{Z_{i_d}(\theta) \le z\} \right|$$

and

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

Moreover, it follows that

$$\left| 1\{Z^\delta_{i-j+1}(\theta) \le z_j\} - 1\{Z_{i-j+1}(\theta) \le z_j\} \right| \le 1\{|Z_{i-j+1}(\theta) - z_j| \le |Z^\delta_{i-j+1}(\theta) - Z_{i-j+1}(\theta)|\}.$$

However, it follows from Assumption 3.5 that

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

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

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

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

$$\rho((z_1, \theta_1), (z_2, \theta_2)) = r(z_1, z_2) \vee \|\theta_1 - \theta_2\|. \tag{51}$$

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

$$f(z, \theta) = 1\big\{Z_{i_d}(\theta) \leq z\big\} \tag{52}$$

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 (51) and (52).

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

$$\left[\mathbb{E} \sup_{x,y\in\mathcal{R},\theta\in\mathcal{N}} \Big(1\{Z_{i_d}(\theta) \leq x\} - 1\{Z_{i_d}(\theta_0) \leq y\}\Big)^2\right]^{1/2} \leq K\varepsilon \tag{53}$$

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 (53), we note that

$$\begin{aligned}
&\Big|1\{Z_{i_d}(\theta) \leq x\} - 1\{Z_{i_d}(\theta_0) \leq y\}\Big| \\
&\quad \leq \Big|1\{Z_{i_d}(\theta) \leq x\} - 1\{Z_{i_d}(\theta_0) \leq x\}\Big| + \Big|1\{Z_{i_d}(\theta_0) \leq x\} - 1\{Z_{i_d}(\theta_0) \leq y\}\Big| \\
&\quad \leq \sum_{j=1}^d \Big|1\{Z_{i-j+1}(\theta) \leq x_j\} - 1\{Z_{i-j+1}(\theta_0) \leq x_j\}\Big| \\
&\qquad + \sum_{j=1}^d \Big|1\{Z_{i+j-1}(\theta_0) \leq x_j\} - 1\{Z_{i-j+1}(\theta_0) \leq y_j\}\Big|,
\end{aligned} \tag{54}$$

and that we have

$$\begin{aligned}
&\Big|1\{Z_{i-j+1}(\theta) \le x_j\} - 1\{Z_{i-j+1}(\theta_0) \le x_j\}\Big| \\
&\quad \le 1\Big\{|Z_{i-j+1}(\theta_0) - x_j| \le |Z_{i-j+1}(\theta) - Z_{i-j+1}(\theta_0)|\Big\} \\
&\quad \le 1\left\{|Z_{i-j+1}(\theta_0) - x_j| \le \|\theta - \theta_0\| \int_{T_{(i-j)\Delta}}^{T_{(i-j+1)\Delta}} \nu(X_t)dt\right\} \\
&\quad \le K\varepsilon \int_{T_{(i-j)\Delta}}^{T_{(i-j+1)\Delta}} \nu(X_t)dt, \qquad (55)
\end{aligned}$$

and

$$\Big|1\{Z_{i-j+1}(\theta_0) \le x_j\} - 1\{Z_{i-j+1}(\theta_0) \le y_j\}\Big| \le 1\Big\{\underline{z}_j \le Z_{i-j+1}(\theta_0) \le \overline{z}_j\Big\} \qquad (56)$$

for $j = 1, \ldots, d$. Consequently, due to (55) and (56), it follows immediately from (54) 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 (53), 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 (51). 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 $N(x, \mathcal{F}) = x^{-2d}$

$$\sum_{k=1}^{\infty} k^{a-2}\alpha(k)^{b/(a+b)} < \infty \quad \text{and} \quad \int_0^1 x^{-b/(2+b)} N(x, \mathcal{F})^{1/a} dx < \infty \qquad (57)$$

hold with some even integers $a \ge 2$ and $b > 0$. Note that the conditions in (57) 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

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

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.,

$$\sqrt{N}[\Pi_N(z,\theta) - \Pi(z,\theta)] - \sqrt{N}[\Pi_N(z,\theta_0) - \Pi(z,\theta_0)] \to_p 0 \tag{59}$$

uniformly in $z \in \mathbb{R}^d$, as $\theta \to \theta_0$. Moreover, it follows from Assumption 3.7(b) that

$$\Pi(z,\theta) - \Pi(z,\theta_0) = (\theta - \theta_0)'\dot{\Pi}(z,\theta_0) + R(z,\theta), \tag{60}$$

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

It follows from (59) and (60) 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 (58) that

$$\begin{aligned} Q_N(\theta) - Q(\theta) = & \int [\Pi_N(z,\theta_0) - \Pi(z,\theta_0)]^2 \varpi(dz) \\ & + 2(\theta - \theta_0)' \int \dot{\Pi}(z,\theta_0)[\Pi_N(z,\theta_0) - \Pi(z,\theta_0)]\varpi(dz) \\ & + o_p(N^{-1}) + o_p\left(N^{-1/2}\|\theta - \theta_0\|\right) \end{aligned} \tag{61}$$

near θ_0.

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

$$Q(\theta) = Q(\theta_0) + \dot{Q}(\theta_0)(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)'\ddot{Q}(\theta_0)(\theta - \theta_0) + o(\|\theta - \theta_0\|^2)$$
$$= \frac{1}{2}(\theta - \theta_0)'\ddot{Q}(\theta_0)(\theta - \theta_0) + o(\|\theta - \theta_0\|^2) \quad (62)$$

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 (61) and (62) that

$$\begin{aligned} Q_N(\theta) = &\int [\Pi_N(z, \theta_0) - \Pi(z, \theta_0)]^2 \varpi(dz) \\ &+ 2(\theta - \theta_0)' \int \dot{\Pi}(z, \theta_0)[\Pi_N(z, \theta_0) - \Pi(z, \theta_0)]\varpi(dz) \\ &+ \frac{1}{2}(\theta - \theta_0)'\ddot{Q}(\theta_0)(\theta - \theta_0) \\ &+ o_p(N^{-1}) + o_p\left(N^{-1/2}\|\theta - \theta_0\|\right) + o(\|\theta - \theta_0\|^2) \end{aligned} \quad (63)$$

near θ_0. Given Assumption 3.7(c) it can therefore be deduced that

$$\sqrt{N}(\hat{\theta}_N - \theta_0) = -2\ddot{Q}(\theta_0)^{-1}\sqrt{N} \int \dot{\Pi}(z, \theta_0)[\Pi_N(z, \theta_0) - \Pi(z, \theta_0)]\varpi(dz) + o_p(1) \quad (64)$$

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 (64) 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

$$\begin{aligned} Q_N^\delta(\theta) - Q_N(\theta) = &\int [\Pi_N^\delta(z, \theta) - \Pi_N(z, \theta)]^2 \varpi(dz) \\ &+ 2\int [\Pi_N^\delta(z, \theta) - \Pi_N(z, \theta)][\Pi_N(z, \theta) - \Pi(z, \theta_0)]\varpi(dz). \end{aligned} \quad (65)$$

It follows from Lemma 3.5 that

$$\int [\Pi_N^\delta(z, \theta) - \Pi_N(z, \theta)]^2 \varpi(dz) = o(N^{-1}) \quad (66)$$

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

$$\begin{aligned} &\left| \int [\Pi_N^\delta(z, \theta) - \Pi_N(z, \theta)][\Pi_N(z, \theta) - \Pi(z, \theta_0)]\varpi(dz) \right| \\ &\leq \left(\int [\Pi_N^\delta(z, \theta) - \Pi_N(z, \theta)]^2 \varpi(dz) \right)^{1/2} \left(\int [\Pi_N(z, \theta) - \Pi(z, \theta_0)]^2 \varpi(dz) \right)^{1/2} \\ &= o_p(N^{-1}) \end{aligned} \quad (67)$$

near θ_0. Therefore, we may deduce from (65), (66) and (67) that

$$\left| Q_N^\delta(\theta) - Q_N(\theta) \right| = o_p(N^{-1}) \quad (68)$$

near θ_0. The asymptotic equivalence between $\hat{\theta}_N$ and $\hat{\theta}_N^\delta$ can therefore be seen easily from (63), and the proof is complete. □

Proof of Corollary 3.8 We note that

$$\left|Q_N^\delta(\hat{\theta}_N^\delta) - Q_N(\hat{\theta}_N)\right| \le \left|Q_N^\delta(\hat{\theta}_N^\delta) - Q_N(\hat{\theta}_N^\delta)\right| + \left|Q_N(\hat{\theta}_N^\delta) - Q_N(\theta_0)\right| + \left|Q_N(\hat{\theta}_N) - Q_N(\theta_0)\right|,$$

and that

$$\left|Q_N(\hat{\theta}_N^\delta) - Q_N(\theta_0)\right|,\ \left|Q_N(\hat{\theta}_N) - Q_N(\theta_0)\right| = o_p(N^{-1})$$

due to (63), and

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

due to (68), from which the first part follows immediately. The second part can also be deduced straightforwardly from (63) and (64). □

Proof of Lemma 3.9 To prove the stated result, it suffices to show that

$$\mathbb{E} \max_{1\le i\le N} \sup_{\theta\in\Theta} \left| \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta},\theta) - \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(\bar{X}_{j\delta},\theta) \right| = o(N^{-1/2}) \tag{69}$$

with $\delta M_i = T_{i\Delta}^\delta$. To establish (69), we note that

$$\begin{aligned}\left| \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta},\theta) - \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(\bar{X}_{j\delta},\theta) \right| &\le \delta \sum_{j=M_{i-1}+1}^{M_i} \left|\mu(X_{j\delta},\theta) - \mu(\bar{X}_{j\delta},\theta)\right| \\ &\le \delta \left(\sup_{0\le t\le T} \nu(X_t) \right) \sum_{j=M_{i-1}+1}^{M_i} \left|X_{j\delta} - \bar{X}_{j\delta}\right| \\ &\le O(e_T) \left|T_{i\Delta}^\delta - T_{(i-1)\Delta}^\delta\right| \left(\sup_{0\le t\le T} \left|X_t - \bar{X}_t\right| \right),\end{aligned}$$

from which it follows that

$$\begin{aligned}&\max_{1\le i\le N} \sup_{\theta\in\Theta} \left| \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(X_{j\delta},\theta) - \sum_{j=M_{i-1}+1}^{M_i} \delta\mu(\bar{X}_{j\delta},\theta) \right| \\ &\le O(e_T) \left(\max_{1\le i\le N} \left|T_{i\Delta}^\delta - T_{(i-1)\Delta}^\delta\right| \right) \left(\sup_{0\le t\le T} \left|X_t - \bar{X}_t\right| \right).\end{aligned}$$

Therefore, by Cauchy-Schwarz, we may easily deduce that

$$\begin{aligned}
&\mathbb{E}\max_{1\le i\le N}\sup_{\theta\in\Theta}\left|\sum_{j=M_{i-1}+1}^{M_i}\delta\mu(X_{j\delta},\theta)\ -\sum_{j=M_{i-1}+1}^{M_i}\delta\mu(\bar{X}_{j\delta},\theta)\right| \\
&\le O(e_T)\left[\mathbb{E}\left(\max_{1\le i\le N}\left|T^{\delta}_{i\Delta}-T^{\delta}_{(i-1)\Delta}\right|\right)^2\right]^{1/2}\left[\mathbb{E}\left(\sup_{0\le t\le T}\left|X_t-\bar{X}_t\right|\right)^2\right]^{1/2} \\
&= O(e_T)O\left(\frac{1}{a_T}\right)O(\delta^p T^q)=O\left(\frac{e_T}{a_T}\delta^p T^q\right),
\end{aligned}$$

due, in particular, to (50). The stated result now readily follows upon noticing that

$$\frac{e_T}{a_T}\delta^p T^q = o\left((Tb_T)^{-1/2}\right)$$

under the given condition for δ, and that $N\Delta \le Tb_T$ and Δ is fixed. □

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., P. Mykland and L. Zhang (2005), "A tale of two time scales: Determining integrated volatility with noisy high-frequency data," *Journal of American Statistical Association*, 100, 1394-1411.

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.

Bandi, F.M. and Russell, J.R. (2008), "Microstructure noise, realized variance, and optimal sampling," *Review of Economic Studies*, 75, 339-369.

Barndorff-Nielsen, O.E. 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 (Statistical Methodology)*, 64, 253-280.

Barndorff-Nielsen, O.E. and N. Shephard (2004), "Econometric analysis of realized covariation: High frequency based covariance, regression and correlation in financial economics," *Econometrica*, 72, 885-925.

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.

Jeong, M. and J.Y. Park (2010), "Asymptotic theory of maximum likelihood estimator for diffusion model," Working paper, Department of Economics, Indiana University.

Kanaya, S. (2008), "Uniform convergence of kernel-based nonparametric estimators for continuous and discrete time processes: A damping function approach," Working paper, Department of Economics, University of Oxford.

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 (2008), "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.

Table 2: Simulation Results for Model I

T	$(\alpha,\beta,\gamma^2)=(\alpha_0,\beta_0,\gamma_0^2)$				T	$(\alpha,\beta,\gamma^2)=(\alpha_0/10,\beta_0/10,\gamma_0^2/10)$				T	$(\alpha,\beta,\gamma^2)=(10\alpha_0,10\beta_0,10\gamma_0^2)$			
		MLE	MGE				MLE	MGE				MLE	MGE	
			1D	2D				1D	2D				1D	2D
5	Bias	-0.911	-0.019	-0.014	5	Bias	-0.936	-0.003	-0.005	5	Bias	-0.801	-0.199	-0.206
	Std	0.867	0.876	1.023		Std	0.819	0.817	0.955		Std	1.182	1.579	1.738
	RMSE	1.258	0.876	1.023		RMSE	1.244	0.817	0.955		RMSE	1.427	1.592	1.750
20	Bias	-0.211	-0.025	-0.032	20	Bias	-0.238	-0.005	-0.002	20	Bias	-0.204	-0.121	-0.175
	Std	0.242	0.285	0.320		Std	0.214	0.222	0.251		Std	0.518	0.858	0.855
	RMSE	0.321	0.286	0.322		RMSE	0.320	0.222	0.252		RMSE	0.557	0.867	0.873
50	Bias	-0.081	-0.024	-0.027	50	Bias	-0.091	0.000	0.000	50	Bias	-0.076	-0.071	-0.046
	Std	0.120	0.162	0.177		Std	0.087	0.092	0.107		Std	0.305	0.532	0.499
	RMSE	0.144	0.164	0.179		RMSE	0.126	0.092	0.107		RMSE	0.314	0.537	0.501

The table presents the bias, standard deviation (Std) and root mean squared error (RMSE) of the MLE and MGE of the parameter β in Model I, which are based on $5,000$ iterations. For the MGE, 1D and 2D refer respectively to the 1-dimensional and 2-dimensional MGE's. The parameter values for α, β and γ^2 are set $(\alpha_0,\beta_0,\gamma_0^2)=(0.01579,-0.219,0.06665^2)$ for the benchmark model. The optimal Δ is searched over the 100 equi-spaced grid points between the expected values of the 20 days of quadratic variation and one-twentieth of the total quadratic variation of the error process, which are obtained analytically for each of our simulation models using their parameter values. We choose the value of Δ, which minimizes the estimated standard error of the MGE's.

Table 3: Simulation Results for Model II

$(\mu,\nu,\omega^2)=(\mu_0,\nu_0,\omega_0^2)$					$(\mu,\nu,\omega^2)=(\mu_0/10,\nu_0/10,\omega_0^2/10)$					$(\mu,\nu,\omega^2)=(10\mu_0,10\nu_0,10\omega_0^2)$				
T		MLE	MGE		T		MLE	MGE		T		MLE	MGE	
			1D	2D				1D	2D				1D	2D
5	Bias	5.229	0.551	0.809	5	Bias	8.216	0.575	0.730	5	Bias	9.191	0.069	-0.747
	Std	6.057	7.248	7.786		Std	11.999	11.530	12.938		Std	5.572	10.185	10.019
	RMSE	8.002	7.269	7.828		RMSE	14.542	11.544	12.959		RMSE	10.748	10.185	10.047
20	Bias	2.740	0.337	0.457	20	Bias	0.822	0.290	0.212	20	Bias	8.688	-1.167	-1.415
	Std	2.644	4.311	4.335		Std	2.570	3.969	4.352		Std	2.666	5.165	4.700
	RMSE	3.808	4.324	4.359		RMSE	2.698	3.980	4.357		RMSE	9.088	5.295	4.908
50	Bias	2.247	0.144	0.209	50	Bias	0.064	0.176	0.235	50	Bias	8.559	-1.682	-2.286
	Std	1.711	3.065	2.894		Std	0.606	2.363	2.543		Std	1.672	3.082	2.825
	RMSE	2.825	3.068	2.902		RMSE	0.610	2.370	2.554		RMSE	8.720	3.511	3.634

The table presents the bias, standard deviation (Std) and root mean squared error (RMSE) of the MLE and MGE of the parameter β in Model II, which are based on $5,000$ iterations. For the MGE, 1D and 2D refer respectively to the 1-dimensional and 2-dimensional MGE's. The parameter values for α, β and ρ are set $(\alpha,\beta,\rho)=(0.025,0.94,-0.8)$ for all the results reported here. The parameter values for μ, ν and ω we use for the benchmark model are $(\mu_0,\nu_0,\omega_0^2)=(0.3,-3,0.25^2)$. The optimal Δ is searched over the 100 equi-spaced grid points between the expected values of the 20 days of quadratic variation and one-twentieth of the total quadratic variation of the error process, which are obtained analytically for each of our simulation models using their parameter values. We choose the value of Δ, which minimizes the estimated standard error of the MGE's.

Table 4: Simulation Results for Model III

MGE with 5% Preliminary Jump Test						MGE with 10% Preliminary Jump Test					
T		$\lambda = 1$		$\lambda = 1.5$		T		$\lambda = 1$		$\lambda = 1.5$	
		1D	2D	1D	2D			1D	2D	1D	2D
5	Bias	0.006	0.019	0.001	-0.005	5	Bias	-0.014	0.017	-0.040	-0.012
	Std	0.877	1.019	0.887	1.037		Std	0.873	1.024	0.882	1.040
	RMSE	0.877	1.020	0.887	1.037		RMSE	0.873	1.024	0.883	1.040
20	Bias	-0.048	-0.057	-0.056	-0.062	20	Bias	-0.034	-0.045	-0.049	-0.059
	Std	0.310	0.351	0.342	0.365		Std	0.314	0.341	0.334	0.358
	RMSE	0.314	0.355	0.347	0.370		RMSE	0.316	0.344	0.338	0.363
50	Bias	-0.041	-0.051	-0.060	-0.069	50	Bias	-0.037	-0.050	-0.059	-0.068
	Std	0.174	0.189	0.185	0.199		Std	0.177	0.188	0.185	0.196
	RMSE	0.179	0.196	0.195	0.210		RMSE	0.181	0.195	0.194	0.207

The table presents the bias, standard deviation (Std) and root mean squared error (RMSE) of the MGE of the parameter β in Model III implemented with preliminary jump test. The reported results are based on $5,000$ iterations. As in Tables 2 and 3, 1D and 2D refer respectively to the 1-dimensional and 2-dimensional MGE's. The parameter values are set $(\alpha, \beta, \rho) = (0.025, 0.94, -0.8)$ and $(\mu, \nu, \omega^2) = (0.3, -3, 0.25^2)$, and jumps are generated from uniform distribution on $[-0.015, -0.005] \cup [0.005, 0.015]$. For Δ, we use the same value as in Model I to make our results more directly comparable to those of Model I reported in Table 2.