On the Statistical Equivalence at Suitable Frequencies of GARCH and Stochastic Volatility Models with the Corresponding Diffusion Model

Lawrence D. Brown
University of Pennsylvania
Yazhen Wang
University of Connecticut
Linda H. Zhao
University of Pennsylvania

September 11, 2002

Abstract

Continuous-time models play a central role in the modern theoretical finance literature, while discrete-time models are often used in the empirical finance literature. The continuous-time models are diffusions governed by stochastic differential equations. Most of the discrete-time models are autoregressive conditionally heteroscedastic (ARCH) models and stochastic volatility (SV) models. The discrete-time models are often regarded as discrete approximations of diffusions because the discrete-time processes weakly converge to the diffusions. It is known that SV models and multiplicative GARCH models share the same diffusion limits in a weak-convergence sense. Here we investigate a much stronger convergence notion. We show that SV models are asymptotically equivalent to their diffusion limits at the basic frequency of their construction, while multiplicative GARCH models match to the diffusion limits only at observed frequencies lower than the square root of the basic frequency of construction. These results also reveal that the structure of the multiplicative GARCH models at frequencies lower than the square root of the basic frequency no longer obey the GARCH framework at the observed frequencies. Instead they behave there like the SV models.

AMS 1991 Subject Classification. Primary: 62B15. Secondary: 90A09, 90A20, 90A16, 62M99.

Key Words: Conditional variance, deficiency distance, financial modeling, frequency, stochastic differential equation, stochastic volatility.

1 Introduction

Since Black and Scholes (1973) derived the price of a call option under the assumption that the underlying stock obeys a geometric Brownian motion, continuous-time models are central to modern finance theory. Currently, much of the theoretical development of contingent claims pricing models has been based on continuous-time models of the sort that can be represented by stochastic differential equations. Application of various "no arbitrage" conditions is most easily accomplished via the Itô differential calculus and requires a continuous-time formulation of the problem. (See Duffie 1992, Hull 1997, Merton 1990.)

In contrast to the stochastic differential equation models so widely used in theoretical finance, in reality virtually all economic time series data are recorded only at discrete intervals, and empiricists have favored discrete-time models. The discrete-time modeling often adopts some stochastic difference equation systems which capture most of the empirical regularities found in financial time series. These regularities include leptokurtosis and skewness in the unconditional distribution of stock returns, volatility clustering, pronounced serial correlation in squared returns but little or no serial dependence in the return process itself. One approach is to express volatility as a deterministic function of lagged residuals. Econometric specifications of this form are known as ARCH models and have achieved widespread popularity in applied empirical research (Bollerslev, Chou and Kroner 1992, Engle 1982, Engle and Bollerslev 1986, Gouriéroux 1997). Alternatively, volatility may be modeled as an unobserved component following some latent stochastic process, such as an autoregression. Models of this kind are known as stochastic volatility (SV) models (Jacquier, Polson and Rossi 1994).

Historically the literature on discrete-time and continuous-time models developed quite independently. Interest in models with stochastic volatility dates back to the early 1970s. Stochastic volatility models naturally arise as discrete approximations to various diffusion processes of interest in the continuous-time asset-pricing literature (Hull and White 1987, Jacquier, Polson and Rossi 1994). The ARCH modeling idea was introduced in 1982 by Robert Engle. Since then, hundreds of research papers applying this modeling strategy to financial time series data have been published, and empirical work with financial time series has been mostly dominated by variants of the ARCH model. Nelson (1990) and Duan (1997) established the link between GARCH models and diffusions by deriving diffusion limits for GARCH processes. Although ARCH modeling was proposed as statistical models and is often viewed as an approximation or a filter tool for diffusion processes, GARCH option pricing model has been developed and shown via the weak convergence linkage to be consistent with option pricing theory based on diffusions (Duan 1995, Kallsen and Taqqu 1998). However, these link study relies solely on the discrete-time models as diffusion approximations in the sense of weak convergence. A precise formulation is described later in this section, and in more detail in Section 2.3. In that formulation weak convergence is satisfactory for studying the limiting distribution of these discrete-time models at separated, fixed time points. It also suffices for studying the distribution of specific linear functionals. Weak convergence is not adequate for studying asymptotic distributions of more complicated functionals or the joint distributions of observations made at converging sets of time points. These issues can be studied by treating GARCH models and their diffusion limits in the statistical paradigm constructed by Le Cam. (See e.g. Le Cam (1986) and Le Cam and Yang (2000).)

The diffusion model is a continuous-time model, while SV and GARCH models are mathematically constructed in discrete time. We consider statistical equivalence for observations from the SV and GARCH models and discrete observations from the corresponding diffusion model over a time span at some frequencies. To describe our results more fully, suppose the processes in time interval $[0, T]$ based on a GARCH or SV model are constructed at $t_i = \frac{i}{n} T$, and the process from the corresponding diffusion model is also discretely observed at t_i, $i = 1, \cdots, n$. Thus, T/n is the basic time interval for the models and $\phi_c = n/T$ is the corresponding basic frequency. We follow Drost and Nijman (1993) to define what we mean lower frequency observations from the observations at basic frequency. Assume x_i, $i = 1, \cdots, n$, is the observations at the basic frequency, the first kind of low frequency observations are assumed to be $x_{k\,\ell}$, $\ell = 1, \cdots, [n/k]$, where k is some integer (which may depend on n in our asymptotic study), $[n/k]$ denotes the integer part of n/k, and for each k, $\phi = \phi_c/k = n/(k\,T)$ is defined to be an associated low frequency. The second kind of low frequency observations are $\bar{x}_{k\,\ell} = \sum_{j=0}^{k-1} x_{k\,\ell-j}$, $\ell = 1, \cdots, [n/k]$, with k as before. Drost and Nijman (1993, section 2) adopted the first kind of low frequency observations for a stock variable and the second kind of low frequency observations for a flow variable. The first case catches the intuition that low frequency observations corresponds to data singled-out at sparse time points and may apply to stationary processes such as currency exchange rate data, while the second case captures the cumulative sum of observations between the spaced-out time points and is naturally appropriate for modeling data like stock returns. This paper will study asymptotic equivalence of low frequency observations from the SV, GARCH and diffusion models at some suitable frequencies. Asymptotic equivalence in this sense can be interpreted in several ways. A basic interpretation is that any sequence of statistical procedures for one model has a corresponding asymptotic-equal-performance sequence for the other model.

We have mainly established asymptotic equivalence for low frequency observations in the sense of the first kind, namely for observations singled out every once a while. Specifically, we are able to show that for any choice of k, including $k = 1$, the SV model and its diffusion limit are asymptotically equivalent, and meanwhile the low frequency observations of the first kind for the GARCH model are asymptotically equivalent to those for its diffusion limit at frequencies $\phi = n/(T\,k)$ with $n^{1/2}/k \to 0$. When $k = 1$, the both kinds of low frequency observations are coincide with the observations at the basic frequency, asymptotic equivalence with $k = 1$ implies that the SV model is asymptotically equivalent to its diffusion limit at any frequencies up to the basic frequency for either kind of low frequency observations. While for the GARCH model, we show only that the sparse observations for the GARCH model match to those for the diffusion limit only at frequencies lower than the square root of the basic frequency. So far we have not succeeded in proving a similar asymptotic equivalence result for the GARCH model with the second kind of low frequency observations, namely aggregated observations between the singled-out sparse time points. However, we conjecture that the same frequency based asymptotic equivalence holds for the second kind of low frequency observations, that is, the low frequency observations of the second kind for the GARCH model match to those for the diffusion limit at frequencies $\phi = n/(T\,k)$ with $n^{1/2}/k \to 0$.

This paper proves only one part of the whole envisioned picture, and the results can also be viewed as a step stone to get the conjecture. In fact, we believe the techniques like bybrid process developed in the proof of Theorem 2 should be very useful for the aggregation case. Also our proofs are actually constructed to show that observations at suitable frequencies of SV or GARCH models asymptotically match in the appropriate distributional sense to observations at the same frequency of their diffusion limit. This establishes somewhat more than asymptotic equivalence in the sense of Le Cam's deficiency distance. It also shows that on the basis of observations at these frequencies it is asymptotically impossible to distinguish whether the observations arose from the SV or GARCH model or the corresponding diffusion model.

Wang (2002) investigated asymptotic equivalence of GARCH and diffusion models when observed at the basic frequency of construction, i.e. when $k = 1$. He showed that these models are not equivalent when observed at that frequency except in the trivial case where the variance term in the GARCH model is non-stochastic. At the other extreme, the choice $k = \epsilon n$ for some fixed ϵ corresponds to observation only at a finite set of time points. In this case a minor elaboration of the weak convergence results of Nelson (1990) shows that the GARCH and diffusion models are asymptotically equivalent when observed at only such a finite set of times. These contrasting results provide motivation for studying asymptotic equivalence for GARCH and SV processes when observed at frequency $\phi = \phi_c/k$ with $k \to \infty$ but $k = o(n)$.

The difference between the equivalence results for the SV models and the GARCH models is due to the fact that these models employ quite different mechanisms to propagate noise in their conditional variances. In the diffusion framework, the conditional variances are governed by an unobservable white noise. However, the GARCH models use past observations to model their conditional variances. The SV models employ an unobservable, i.i.d. normal noise to model their conditional variances, and this closely mimics the diffusion mechanism. This fact has a twofold implication. First, the close mimicking makes the SV models asymptotically equivalent to diffusions at all frequencies. Second, the different noise propagation systems in the GARCH and SV models result in different patterns in equivalence with respect to frequency. It takes much longer for the GARCH framework to make the innovation process (i.e. the square of past observation errors) in the conditional variance close to white noise than it does for the SV models with i.i.d. normal errors. Thus, the GARCH models are asymptotically equivalent to the diffusion limits only when observed at much lower frequencies than the SV models.

The paper is organized as follows. Section 2 reviews diffusions, GARCH and SV models and illustrates the link of the discrete-time models to diffusions. Section 3 presents some basic concepts of statistical equivalence and defines what we mean by equivalence in terms of observational frequency for the GARCH, SV, and diffusion models. The equivalence results for the SV and GARCH models are featured in Sections 4 and 5, respectively. Some technical lemmas are collected in Section 6. Since the GARCH counterpart of an SV model is the multiplicative GARCH, and the multiplicative GARCH and SV models have the same diffusion limits, this paper investigates equivalence only for the multiplicative GARCH models. We believe that the methods and techniques developed in this paper may be adopted for the study of equivalence of other GARCH models and their diffusion limits.

2 Financial models

2.1 Diffusions

Continuous-time financial models frequently assume that a security price S_t obeys the following stochastic differential equation

$$dS_t = \mu_t \, S_t \, dt + \sigma_t \, S_t \, dW_t, \qquad t \in [0, T], \tag{1}$$

where W_t is a standard Wiener process, μ_t is called the drift in probability or the mean return in finance, and σ_t^2 is called the diffusion variance in probability or the (conditional) volatility in finance. The celebrated Black-Scholes model corresponds to (1) with constants $\mu_t = \mu$ and $\sigma_t = \sigma$.

For continuous-time models, the "no arbitrage" (often labeled in plain English as "no free lunch") condition can be elegantly characterized by martingale measure under which $\mu_t = 0$ and the discounted price process is a martingale. Prices of options are then the conditional expectation of a certain functional of S under this measure. These calculations and derivations can be easily manipulated by tools such as Itô's lemma and Girsanov's theorem. (See Duffie (1992), Hull and White (1987), Karatzas and Shreve (1997), Merton (1990).)

Many econometric studies have documented that financial time series tend to be highly heteroskedastic. To accommodate this one often allows σ_t^2 to be random (in place of the assumption that $\sigma_t = \sigma$) and assumes $\log \sigma_t^2$ itself is governed by another stochastic differential equation. Such σ_t^2 is called stochastic volatility.

We will be interested in properties of this continuous time model when observed at regular discrete time intervals. To describe this divide the time interval $[0,\ T]$ into n subintervals of length $\lambda_n = T/n$ and set $t_i = i\lambda_n, i = 1, \ldots, n$. There is no loss of generality in assuming $T = 1$, and we will henceforth do so. Then $\lambda_n = 1/n$.

2.2 Stochastic volatility models

In the general discrete time stochastic volatility model each data point has a conditional variance which is called volatility. The volatilities are unobservable and are assumed to be probabilistically generated. The density of the data is a mixture over the volatility distribution. The widely used stochastic volatility model assumes that the conditional variance of each incremental observation y_i follows a log-AR(p) process

$$y_i = \rho_i\, \varepsilon_i,$$

$$log\, \rho_i^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j\, log\, \rho_{i-j}^2 + \alpha_{p+1}\, \gamma_i,$$

where ε_i and γ_i are independent standard normal random variables. See Ghysels, Harvey and Renault (1996). This paper deals with SV models with AR(1) specification only. In accordance with the previous assumption we take $T = 1$ and $\lambda_n = 1/n$. Redefining the constants to correspond to the diffusion model in (7) and (8) below we write the model as

$$y_i = \rho_i \varepsilon_i / \sqrt{n} \quad and \tag{2}$$

$$\log \rho_i^2 = \frac{\beta_0}{n} + (1 + \frac{\beta_1}{n}) \log \rho_{i-1}^2 + \beta_2 \frac{\gamma_i}{\sqrt{n}}. \tag{3}$$

Denote by $Y_0, \ldots, Y_n$ the partial sum process of y_i, or equivalently,

$$y_i = Y_i - Y_{i-1}, i = 1, \ldots, n.$$

2.3 GARCH models

Engle(1982) introduced the ARCH model by setting the conditional variance, τ_i^2, of a series of prediction errors equal to a linear function of lagged errors. Generalizing ARCH(p), Bollerslev(1986) introduced a linear GARCH specification in which τ_i^2 is an ARMA process with non-negative coefficients and with past z_i^2's as the innovation process. Geweke(1986) and Pantula(1986) adopted a natural device for ensuring that τ_i^2 remains non-negative, by making $log\,\tau_i^2$ linear in some function of time and lagged z_i's. Then

$$z_i = \tau_i \varepsilon_i \quad and$$

$$\log\,\tau_i^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \log\,\tau_{i-j}^2 + \sum_{j=1}^{q} \alpha_{p+j} \log\,\varepsilon_{i-j}^2,$$

where ε_i are independent standard normal random variables and α's are constants. This model is often referred to as multiplicative GARCH (p,q) (MGARCH (p,q)).

In many applications, the MGARCH(1,1) specification has been used and has been found to be adequate. (See Bollerslev, Chou and Kroner 1992, Engle 1982, Duan 1997, Engle and Bollerslev 1986, Gouriéroux 1997.) In the sequel we treat only the case MGARCH (1,1). There are several other variants of ARCH and GARCH models. We believe that the methods of this paper could be successfully applied to many of these variants.

More formally, for i.i.d. standard normal ε_i, let

$$c_0 = E(\log\,\varepsilon_i^2), \qquad c_1 = \{var(\log\,\varepsilon_i^2)\}^{1/2}, \qquad \xi_i = (log\,\varepsilon_i^2 - c_0)/c_1. \tag{4}$$

Then, suppressing in the notation the dependence on n, let

$$z_i = \tau_i\, \varepsilon_i/\sqrt{n}, \tag{5}$$

$$log\,\tau_i^2 = \frac{\beta_0}{n} + (1 + \frac{\beta_1}{n})\, log\,\tau_{i-1}^2 + \beta_2\ \xi_i/\sqrt{n}. \tag{6}$$

Denote by $Z_0, \ldots, Z_n$ the partial sum process of z_i, or equivalently, $z_i = Z_i - Z_{i-1}$, $i = 1, \ldots, n$. A continuous time MGARCH(1,1) approximating process $(Z_{n,t}, \tau_{n,t}^2)$, $t \in [0,1]$, is given by

$$Z_{n,t} = Z_i, \qquad \tau_{n,t}^2 = \tau_{i+1}^2, \qquad \text{for } t \in [t_i, t_{i+1}).$$

Nelson (1990) showed that as $n \to \infty$, the normalized partial sum process of (ε_i, ξ_i) weakly converges to a planar Wiener process and the process $(Z_{n,t}, \tau_{n,t}^2)$ converges in distribution to the bivariate diffusion process (X_t, σ_t^2) satisfying

$$dX_t = \sigma_t\, dW_{1,t} \quad t \in [0,\ 1], \tag{7}$$

$$d\ \log\ \sigma_t^2 = (\beta_0 + \beta_1 \log \sigma_t^2)\, dt + \beta_2 dW_{2,\,t}, \quad t \in [0,1], \tag{8}$$

where $W_{1,\,t}$ and $W_{2,\,t}$ are two independent standard Weiner processes. The diffusion model described by (7)-(8) is thus called the diffusion limit of the MGARCH process. For the diffusion limit, denote its discrete samples at t_i by $X_i = X_{t_i}$, and define the corresponding difference process by $x_i = X_i - X_{i-1}$, $i = 1, \cdots, n$.

We assume that the initial values $X_0 = Y_0 = Z_0$ and $\sigma_0^2 = \tau_0^2 = \rho_0^2$ are known constants. Note that x_i, y_i, z_i are the difference processes of X_i, Y_i, Z_i, respectively, or X_i, Y_i, Z_i are the respective partial sum processes of x_i, y_i, z_i. Also we will refer to z_i as observations from the GARCH model and Z_i as the GARCH approximating process.

3 Statistical equivalence

3.1 Comparison of experiments

A statistical problem $\mathbb{E}$ consists of a sample space Ω, a suitable σ-field $\mathcal{F}$, and a family of distributions P_θ indexed by parameter θ which belongs to some parameter space Θ, that is, $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta, \theta \in \Theta))$.

Consider two statistical experiments with the same parameter space Θ,

$$\mathbb{E}_i = (\Omega_i, \mathcal{F}_i, (P_{i,\theta}, \theta \in \Theta)), \qquad i = 1, 2.$$

Denote by $\mathcal{A}$ a measurable action space, let $L : \Theta \times \mathcal{A} \to [0, \infty)$ be a loss function, and set $\|L\| = sup\{L(\theta, a) : \theta \in \Theta, a \in \mathcal{A}\}$. In the ith problem, let δ_i be a decision procedure and denote by $R_i(\delta_i, L, \theta)$ the risk from using procedure δ_i when L is the loss function and θ is the true value of the parameter. Le Cam's deficiency distance is defined as follows,

$$\begin{aligned} \Delta(\mathbb{E}_1, \mathbb{E}_2) &= max \Bigg\{ \inf_{\delta_1} \sup_{\delta_2} \sup_{\theta \in \Theta} \sup_{L:\|L\|=1} |R_1(\delta_1, L, \theta) - R_2(\delta_2, L, \theta)|, \\ &\qquad \inf_{\delta_2} \sup_{\delta_1} \sup_{L:\|L\|=1} |R_1(\delta_1, L, \theta) - R_2(\delta_2, L, \theta)| \Bigg\}. \end{aligned}$$

Le Cam (1986) and Le Cam and Yang (2000) provide other useful expressions for Δ.

Two experiments $\mathbb{E}_1$ and $\mathbb{E}_2$ are called equivalent if $\Delta(\mathbb{E}_1, \mathbb{E}_2) = 0$. The equivalence means that for every procedure δ_1 in problem $\mathbb{E}_1$, there is a procedure δ_2 in problem $\mathbb{E}_2$ with the same risk, uniformly over $\theta \in \Theta$ and all L with $\|L\| = 1$, and vice versa. Two sequences of statistical experiments $\mathbb{E}_{n,1}$ and $\mathbb{E}_{n,2}$ are said to be asymptotically equivalent if $\Delta(\mathbb{E}_{n,1}, \mathbb{E}_{n,2}) \to 0$, as $n \to \infty$. Thus, for any sequence of procedures $\delta_{n,1}$ in problem $\mathbb{E}_{n,1}$

there is a sequence of procedures $\delta_{n,2}$ in problem $I\!E_{n,2}$ with risk differences tending to zero uniformly over $\theta \in \Theta$ and all L with $\|L\| = 1$, i.e.

$$\sup_{\theta \in \Theta} \sup_{L:\|L\|=1} |R_1(\delta_{n,2}, L, \theta) - R_2(\delta_{n,2}, L, \theta)| \to 0.$$

The procedures $\delta_{n,1}$ and $\delta_{n,2}$ are said to be asymptotically equivalent.

For processes $\mathbf{X}_i$ on $(\Omega_i, \mathcal{F}_i)$ with distributions $P_{\theta,i}$, for convenience we often write $\Delta(I\!E_1, I\!E_2)$ as $\Delta(\mathbf{X}_1, \mathbf{X}_2)$. Suppose $P_{\theta,i}$ have densities $f_{\theta,i}$ with respect to measure $\zeta(d\mathbf{u})$. Define L_1 distance

$$D(f_{\theta,1}, f_{\theta,2}) = \int |f_{\theta,1}(\mathbf{u}) - f_{\theta,2}(\mathbf{u})| \, \zeta(d\mathbf{u}).$$

Then

$$\Delta(\mathbf{X}_1, \mathbf{X}_2) \le \sup_{\theta \in \Theta} D(f_{\theta,1}, f_{\theta,2}). \tag{9}$$

(See Brown and Low 1996 (theorem 3.1), and previously cited references. Define Hellinger distance

$$H^2(f_{\theta,1}, f_{\theta,2}) = \frac{1}{2} \int |f_{\theta,1}^{1/2}(\mathbf{u}) - f_{\theta,2}^{1/2}(\mathbf{u})|^2 \, \zeta(d\mathbf{u}).$$

Hellinger distance can easily handle measures of product forms, as encountered in the study of independent observations and some dependent observations. For example,

$$H^2(\prod_{j=1}^m f_{1,j}, \prod_{j=1}^m f_{2,j}) = 1 - \prod_{j=1}^m \left[1 - H^2(f_{1,j}, f_{2,j})\right] \le \sum_{j=1}^m H^2(f_{1,j}, f_{2,j}), \tag{10}$$

and

$$H^2(N(0, \sigma_1^2), N(0, \sigma_2^2)) = 1 - \left[\frac{2\,\sigma_1\,\sigma_2}{\sigma_1^2 + \sigma_2^2}\right]^{1/2} \le (\frac{\min(\sigma_1^2, \sigma_2^2)}{\max(\sigma_1^2, \sigma_2^2)} - 1)^2. \tag{11}$$

See Brown, et. al. (2001, Lemma 3) for the final inequality.

We have the following relation between Hellinger distance and L_1 distance

$$H^2(f_{\theta,1}, f_{\theta,2}) \le D(f_{\theta,1}, f_{\theta,2}) \le 2\,H(f_{\theta,1}, f_{\theta,2}). \tag{12}$$

For convenience we also use notations $D(\mathbf{X}_1, \mathbf{X}_2)$ and $H(\mathbf{X}_1, \mathbf{X}_2)$ for L_1 and Hellinger distances, respectively.

The above expressions suggest our proofs of asymptotic equivalence of two experiments begin by representing the two relevant series of observations on the same sample space. For example in Theorem 2 we deal with the first kind of low frequency observations $\{x_{k\ell}\}_\ell$ and $\{z_{k\ell}\}_\ell$ for the incremental processes of diffusion and MGARCH processes observed at frequency $\phi = n/(k\,T)$, where $k/\sqrt{n} \to \infty$. These have joint densities $(f_{\theta,1}, f_{\theta,2})$, say, where the dependence on n is suppressed in this notation. We prove that $D(f_{\theta,1},\ f_{\theta,2}) \to 0$ uniformly over $\theta \in \Theta$. Hence $\Delta(\{x_{k\ell}\}_\ell, \{y_{k\ell}\}_\ell) \to 0$ by (9).

Such a proof also verifies the impossibility of constructing an asymptotically informative sequence of tests to determine which of the two experiments produced the observed data. Thus, let $\delta_n(\{w_{k\ell}\}_\ell)$ be any sequence of tests designed to determine which of the two experiments produced the data. Such a sequence is asymptotically non-informative at θ to distinguish $\{x_{k\ell}\}_\ell$ from $\{y_{k\ell}\}_\ell$ if

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta} E_\theta(|\delta_n(\{x_{k\ell}\}_\ell) - \delta_n(\{y_{k\ell}\}_\ell)|) = 0.$$

Since we prove that $\lim_{n\to\infty} \sup_{\theta\in\Theta} D(f_{\theta,1},\ f_{\theta,2}) = 0$ it follows directly that all sequences δ_n are asymptotically non-informative in the above sense.

3.2 MGARCH, SV and Diffusion Experiments

Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ be the vector of parameters for the MGARCH, SV and diffusion models defined in section 2, and let the parameter space Θ consist of β_i belonging to bounded intervals.

From section 2, observations $\{y_i\}_{1\le i\le n}$ from the SV model and observations $\{z_i\}_{1\le i\le n}$ from the MGARCH model are defined by the stochastic difference equations (2)-(3) and (4)-(6), respectively, process $\{x_i\}_{1\le i\le n}$ is the difference proccess of the discrete samples at $t_i = i/n$, $i = 1, \cdots, n$, of the diffusion proccess X_t governed by the stochastic differential equations (7)-(8).

In Theorem 1 we establish that the SV process $\{Y_i\}_{1\le i\le n}$ and the discrete version $\{X_i\}_{1\le i\le n}$ of the approximating diffusion process are asymptotically equivalent (at the basic frequency). The proof proceeds by examining the incremental processes $\{y_i\}$, $\{x_i\}$ and showing these are asymptotically equivalent.

The MGARCH models use past observational errors to propagate their conditional variances, while the diffusion and SV models employ unobservable, white noise and i.i.d. normal random variables to govern their conditional variances, respectively. Because of the different noise propagation systems in the conditional variances, Wang (2002) showed that under stochastic volatility, their likelihood processes have different asymptotic distributions, and consequently the two type of models are not asymptotically equivalent. In other words neither $D(\{x_i\}_{1\le i\le n}, \{z_i\}_{1\le i\le n})$ nor $D(\{y_i\}_{1\le i\le n}, \{z_i\}_{1\le i\le n})$ converge to zero. Thus, at the basic frequency MGARCH is not asymptotically equivalent to the other two models. We will study the asymptotic equivalence of the first kind of low frequency observations for the processes $\{x_i\}_{1\le i\le n}$, $\{y_i\}_{1\le i\le n}$, and $\{z_i\}_{1\le i\le n}$. Namely, we investigate whether the processes $\{x_{k\ell}\}_\ell$, $\{y_{k\ell}\}_\ell$, $\{z_{k\ell}\}_\ell$, $\ell = 1, \cdots, m = [n/k]$, are asymptotically equivalent for some integers k, where $[n/k]$ denotes the integer part of n/k.

For convenience we give a formal definition corresponding to the above notion. For two processes $\{x_i\}_i$ and $\{y_i\}_i$, we say that their low frequency observations, $\{x_{k\ell}\}_\ell$ and $\{y_{k\ell}\}_\ell$, of the first kind are asymptotically equivalent at frequency $\phi = n/(k\,T)$, if as $n \to \infty$,

$$\Delta(\{x_{k\ell}\}_{1\le\ell\le m}, \{y_{k\ell}\}_{1\le\ell\le m}) \to 0.$$

Similarly, we say that their low frequency observations of the second kind are asymptotically equivalent at frequency $\phi = n/(k\,T)$, if as $n \to \infty$,

$$\Delta(\{\bar{x}_{k\ell}\}_{1\le\ell\le m}, \{\bar{y}_{k\ell}\}_{1\le\ell\le m}) \to 0.$$

From the definition in section 1, $\bar{x}_{k\ell}$ and $\bar{y}_{k\ell}$ are the cumulative sum of x_i and y_i for $i = k\,(\ell-1)+1, \cdots, k\,\ell$, respectively, and hence $\bar{x}_{k\ell} = X_{k\ell} - X_{k(\ell-1)}$ and $\bar{y}_{k\ell} = Y_{k\ell} - Y_{k(\ell-1)}$. Therefore, the second kind of low frequency observations for x_i and y_i correspond to the difference of the first kind of low frequency observations for their partial sum processes X_i and Y_i, respectively. As a process is statistically equivalent to its difference process plus initial value, asymptotic equivalence of low frequency observations of the first kind for X's and Y's is the same as that of the second kind for their incremental processes x's and y's. Also, for each kind of low frequency observations, if $k_1 \le k_2$, asymptotic equivalence at frequency $\phi_1 = n/(k_1 T)$ implies asymptotic equivalence at frequency $\phi_2 = n/(k_2 T)$. In particular, asymptotic equivalence at the basic frequency (i.e. $k = 1$) infers asymptotic equivalence at any low frequencies of either kind.

4 Equivalence of diffusions and SV models

Theorem 1 *Let Θ be any bounded subset of $\{\beta_0, \beta_1, \beta_2\}$. As $n \to \infty$,*

$$\Delta(\{x_i\}, \{y_i\}) \to 0.$$

Remark 1. Theorem 1 implies that the SV model is asymptotically equivalent to its diffusion limit at the basic frequency. This consequently shows the asymptotic equivalence of low frequency observations of either kind for the SV and diffusion models.

Proof. We will reserve p and q for the probability densities of processes related to x_i's and y_i's, respectively. From the structure of the SV process defined by (5)-(6), we can easily derive that conditional on $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_n)$, y_i are independent with y_i conditionally following a normal distribution with mean zero and variance σ_i^2. Thus,

$$q(\mathbf{y}) = Eq(\mathbf{y}|\boldsymbol{\gamma}), \tag{13}$$

where $q(\cdot|\boldsymbol{\gamma})$ denotes the conditional normal distribution of $\mathbf{y}$ given $\boldsymbol{\gamma}$. Similarly, the structure of the diffusion process defined in (7)-(8) implies that conditional on W_2 the x_i are independent and follow a normal distribution with mean zero and variance $\tilde{\sigma}_i^2 = \int_{(i-1)/n}^{i/n} \sigma_t^2 \, dt$

$$p(\mathbf{x}) = E(p(\mathbf{x}|W_2)). \tag{14}$$

The normal random variables $\boldsymbol{\gamma}$ and the process W_2 can be realized on a common space by writing $\boldsymbol{\gamma} = \boldsymbol{\gamma}(W_2)$ where $\gamma_i = n^{1/2}(W_{2,t_i} - W_{2,t_{i-1}})$.

Lemma 4 in section 6 shows that on this space

$$|\log \rho_i^2 - \log \sigma_{t_i}^2| = O_p(\frac{1}{n}) \;\; i = 1, \ldots, n$$

uniformly in Θ, i, where $t_i = i/n$. Now we can denote by E_{W_2} the expectation taken with respect to W_2 and write

$$\begin{aligned}
D(\{x_i\}, \{y_i\}) &= \int |p(\mathbf{u}) - q(\mathbf{u})| \, d\mathbf{u} \\
&= \int |E_{W_2} p(\mathbf{u}|W_2) - q(\mathbf{u}|\boldsymbol{\gamma}(W_2))| \, d\mathbf{u} \\
&\leq E_{W_2} \int |p(\mathbf{u}|W_2) - q(\mathbf{u}|\boldsymbol{\gamma}(W_2))| \, d\mathbf{u} \\
&\leq 2\, E_{W_2} H\left(p(\mathbf{u}|W_2), q(\mathbf{u}|\boldsymbol{\gamma}(W_2))\right) \\
&= 2\, E_{W_2} H\left(\prod_{\ell=1}^{n} N(0, \tilde{\sigma}_\ell^2), \prod_{\ell=1}^{n} N(0, \rho_\ell^2)\right) \\
&\leq 2\, E_{W_2}\left(\{\sum_{\ell=1}^{n} (\frac{\min(\tilde{\sigma}_\ell^2, \sigma_\ell^2)}{\max(\tilde{\sigma}_\ell^2, \sigma_\ell^2)} - 1)^2\}^{1/2}\right) \\
&\leq 2 E_{W_2}(\{n\, O_p(\frac{1}{n^2})\}^{1/2}) \\
&= O(\frac{1}{\sqrt{n}}) \to 0 \; as \; n \to \infty.
\end{aligned} \tag{15}$$

5 Equivalence of diffusions and MGARCH models

Theorem 2 *Let Θ be a bounded subset. For any $k = n^{1/2}\, r_n$ with $r_n \to \infty$, we have*

$$\Delta(\{x_{k\ell}\}_\ell, \{z_{k\ell}\}_\ell) \to 0, \qquad as\; n \to \infty.$$

Remark 2. Theorem 2 shows that for the observations z_i from the MGARCH model (i.e. the increments of the MGARCH approximating process Z_i), their low frequency observations of the first kind are asymptotically equivalent to those for its diffusion limit at frequencies $\phi = n^{1/2}/(T\, r_n)$ for any $r_n \to \infty$. Taking r_n to diverge arbitrarily slowly, we have that although the MGARCH model and its diffusion limit are not asymptotically equivalent at the basic frequency $\phi_c = n/T$, their low frequency observations of the first kind are asymptotically equivalent at frequencies lower than the square root of the basic frequency.

Remark 3. We are currently a few step short of obtaining a similar asymptotic equivalence result for the MGARCH model with regard to the second kind of low frequency observations, namely aggregated observations $\bar{z}_{k\,\ell}$ between the singled-out sparse time points (or equivalently, the difference of the first kind of low frequency observations $Z_{k\,\ell}$ for MGARCH approximating process Z_i). The heuristic intuition and our insight lead us to believe that the same frequency based asymptotic equivalence holds for the second kind of low frequency observations, that is, the low frequency observations $\bar{z}_{k\,\ell}$ of the second kind from the GARCH model are asymptotically equivalent to $\bar{x}_{k\,\ell}$ from the diffusion limit at frequencies $\phi = n/(T\,k)$ with $n^{1/2}/k \to 0$. Or equivalently, the low frequency observations $Z_{k\,\ell}$ of the first kind for the GARCH approximating process are asymptotically equivalent to $X_{k\,\ell}$ for the diffusion limit at frequencies $\phi = n/(T\,k)$ with $n^{1/2}/k \to 0$.

Remark 4. Comparing Theorems 1 and 2, we see that observations from the MGARCH model and its diffusion limit start to be asymptotically equivalent at frequencies much lower than those for the SV model case. This is due to the noise propagation systems in their conditional variances. The MGARCH model utilizes past observational errors to model its conditional variance, while the conditional variance of the SV model is governed by i.i.d. normal random variables, which are a discrete version of the white noise used by the diffusion to model its conditional variance. Because of the mimicking of white noise by i.i.d. normal errors, the SV model is much closer to the diffusion limit than the MGARCH model. Thus, observations from the SV model can be asymptotically equivalent to those from the diffusion limit at higher frequencies than those from the MGARCH model.

Remark 5. The equivalence result in Theorem 2 reveals that the first kind of low frequency observations from the MGARCH model at frequencies asymptotically lower than $n^{1/2}$ are no longer ARCH or GARCH, but instead they behave like a SV model. This can be explicitly seen from the introduced hybrid process in the proof of Theorem 2 below. The result is also consistent with Drost and Nijman (1993, examples 1 and 3 in section 3), which showed that for the first kind of low frequency observations, their GARCH structures begin to break down at some lower frequencies. More precisely, our result reveals explicitly that the structures of the MGARCH model at frequencies lower than $n^{1/2}$ are similar to those of a SV model.

Proof. Define a hybrid process as follows,

$$\underline{z}_i = \bar{\tau}_i \, \varepsilon_i, \qquad i = 1, \cdots, n, \tag{16}$$

$$log \, \bar{\tau}_{k\,\ell+1} = \alpha_0 + \alpha_1 \, log \, \bar{\tau}_{k\,\ell}, \qquad \ell = 1, \cdots, m, \tag{17}$$

and for $1 \le i \le n$ and $i \ne k\,\ell + 1$ with $1 \le \ell \le m$,

$$log \, \bar{\tau}_i = \alpha_0 + \alpha_1 \, log \, \bar{\tau}_{i-1} + \alpha_2 \, \xi_{i-1}, \tag{18}$$

where ξ_i are defined in (4), $\alpha_0 = \beta_0 \, \lambda_n$, $\alpha_1 = 1 + \beta_1 \, \lambda_n$, and $\alpha_2 = \beta_2 \, \lambda_n^{1/2}$.

We fix the following convention. Notations h and $\underline{h}$ are reserved for the probability densities of processes relating to z_i's and $\underline{z}_i$'s, respectively, with notations p and q for these of x_i's and y_i's, respectively.

For convenience, for $\ell = 1, \cdots, m = [n/k]$, let

$$x^*_\ell = x_{k\,\ell}, \qquad y^*_\ell = y_{k\,\ell}, \qquad z^*_\ell = z_{k\,\ell}, \qquad \underline{z}^*_\ell = \underline{z}_{k\,\ell},$$

and

$$\mathbf{x}^* = (x^*_1, \cdots, x^*_m), \;\; \mathbf{y}^* = (y^*_1, \cdots, y^*_m), \;\; \mathbf{z}^* = (z^*_1, \cdots, z^*_m), \;\; \underline{\mathbf{z}}^* = (\underline{z}^*_1, \cdots, \underline{z}^*_m).$$

Let $\boldsymbol{\varepsilon} = \{\varepsilon_i : 1 \le i \le n, i \ne k\,\ell \quad \ell = 1, \ldots, m\}$, that is, $\boldsymbol{\varepsilon}$ consists of all ε_i whose index i is not a multiple of k. From the framework of the MGARCH process defined by (5)-(6), we see a one-to-one relationship between $\{z_1, \cdots, z_n\}$ and $\{\boldsymbol{\varepsilon}, \mathbf{z}^*\}$, and thus the distribution of $z_1, \cdots, z_n$ is uniquely determined by $\boldsymbol{\varepsilon}$ and $\mathbf{z}^*$, and vice versa. Denote by $h(\cdot|\boldsymbol{\varepsilon})$ the conditional distribution of $\mathbf{z}^*$ given $\boldsymbol{\varepsilon}$. Then the marginal density of $\mathbf{z}^*$ is given by

$$h(\cdot) = E_{\boldsymbol{\varepsilon}} h(\cdot|\boldsymbol{\varepsilon}), \tag{19}$$

where $E_{\boldsymbol{\varepsilon}}$ denotes the expectation taken with respect to $\boldsymbol{\varepsilon}$. Similarly for the process $\underline{z}_i$'s defined by (16)-(18), denote by $\underline{h}(\cdot|\boldsymbol{\varepsilon})$ the conditional distribution of $\underline{\mathbf{z}}^*$ given $\boldsymbol{\varepsilon}$. Then

$$\underline{h}(\cdot) = E_{\boldsymbol{\varepsilon}} \underline{h}(\cdot|\boldsymbol{\varepsilon}). \tag{20}$$

From the definition of $\underline{z}_i$ given by (16)-(18), the conditional variance of $\underline{\mathbf{z}}^* = (\underline{z}^*_1, \cdots, \underline{z}^*_m)$ depend on only $\{\underline{z}_i, 1 \le i \le n, i \ne k\,\ell, \ell = 1, \cdots, m\}$, or equivalently, $\boldsymbol{\varepsilon}$. Thus, conditional on $\boldsymbol{\varepsilon}$, $\underline{z}^*_1, \cdots, \underline{z}^*_m$ are conditionally independent and have normal distributions with conditional mean zero and conditional variance $\bar{\tau}_{k\,\ell}^2$ for $\underline{z}_\ell^*$. The process $\underline{\mathbf{z}}^*$ behaves like an SV process with conditional variances driven by log normal random variables.

Since $D(\mathbf{y}^*, \mathbf{z}^*) \le D(\mathbf{y}^*, \underline{\mathbf{z}}^*) + D(\mathbf{z}^*, \underline{\mathbf{z}}^*)$, to prove the theorem we need to show that $D(\mathbf{z}^*, \underline{\mathbf{z}}^*)$ and $D(\mathbf{y}^*, \underline{\mathbf{z}}^*)$ both converge to zero for k specified in the theorem.

First, since both $\mathbf{y}^*$ and $\underline{\mathbf{z}}^*$ are SV processes, the same arguments to show (15) in the proof of Theorem 1 lead to

$$D(\mathbf{y}^*, \underline{\mathbf{z}}^*) \le 2\, E_{\boldsymbol{\delta}\boldsymbol{\varepsilon}} \left(1 - \prod_{\ell=1}^{m} \left| \frac{2\,\sigma_{k\,\ell} \bar{\tau}_{k\,\ell}}{\sigma_{k\,\ell}^2 + \bar{\tau}_{k\,\ell}^2} \right|^{1/2} \right). \tag{21}$$

Using Lemmas 1, 7 and 9, and the arguments to prove (15) in the proof of Theorem 1 we can show that the term inside the expectation in (21) is bounded by one and has order

$$\begin{aligned} m\, O_p\left([n^{-1/2}\, log\, n + k^{-1/2}]^2 \right) &= O_p(k^{-1}\, log^2\, n + n\, k^{-2}) \\ &= O_p(n^{-1/2}\, log^2\, n\, r_n^{-1} + r_n^{-2}) = o_p(1). \end{aligned}$$

Now applying the Dominated Convergence Theorem to the right hand side of (21) proves that $D(\mathbf{y}^*, \underline{\mathbf{z}}^*)$ tends to zero.

Second, we will show $D(\mathbf{z}^*, \underline{\mathbf{z}}^*) \to 0$. From (19) and (20) we have

$$\begin{aligned} D(\mathbf{z}^*, \underline{\mathbf{z}}^*) &= \int |h(\mathbf{u}) - \underline{h}(\mathbf{u})|\, d\mathbf{u} \\ &= \int |E_{\boldsymbol{\varepsilon}} h(\mathbf{u}|\boldsymbol{\varepsilon}) - E_{\boldsymbol{\varepsilon}} \underline{h}(\mathbf{u}|\boldsymbol{\varepsilon})|\, d\mathbf{u} \\ &\le E_{\boldsymbol{\varepsilon}} \int |h(\mathbf{u}|\boldsymbol{\varepsilon}) - \underline{h}(\mathbf{u}|\boldsymbol{\varepsilon})|\, d\mathbf{u}. \end{aligned} \tag{22}$$

Applying successive conditional arguments to the GARCH process z_i defined by (5)-(6), we derive that the joint conditional distribution of $\mathbf{z}^* = (z_1^*, \cdots, z_m^*)$ given $\boldsymbol{\varepsilon}$ is a product of $N(0, \tau_{k\,\ell}^2)$, where $\tau_{k\,\ell}^2$ depends on $z_1^*, \cdots, z_{\ell-1}^*$ and ε_i for $1 \le i < k\,\ell$ and i being not a multiple of k. In comparison, the conditional variance $\bar{\tau}_{k\,\ell}^2$ of the SV process $\underline{z}_\ell^*$ depends on only ε_i, where $1 \le i < k\,\ell$ and i is not a multiple of k.

Let

$$M_\ell = log\, \tau_{k\,\ell}^2 - log\, \bar{\tau}_{k\,\ell}^2 = \alpha_2\, \alpha_1^{-1} \sum_{l=1}^{\ell-1} \alpha_1^{k\,\ell - k\,l}\, \xi_{k\,l}, \tag{23}$$

and define events

$$\Omega_{j,n} = \left\{ \sup_{1 \le \ell \le j-1} M_\ell \le A_n \right\}, \qquad j = 2, \cdots, m, \tag{24}$$

where A_n is a constant whose value will be specified later, $\alpha_0 = \beta_0\, \lambda_n$, $\alpha_1 = 1 + \beta_1\, \lambda_n$ and $\alpha_2 = \beta_2\, \lambda_n^{1/2}$.

Since $\Omega_{j,n}^c$ depend on only ε_i whose distributions are the same under both models for z_i's (with density h) and $\underline{z}_i$'s (with density $\underline{h}$), applying Lemma 2 we get

$$\int |h(\mathbf{u}|\boldsymbol{\varepsilon}) - \underline{h}(\mathbf{u}|\boldsymbol{\varepsilon})|\, d\mathbf{u} \le 2P(\Omega_{m,n}^c) + \sqrt{8}\left\{P(\Omega_{m,n}) - \int_{\Omega_{m,n}} |h(\mathbf{u}|\boldsymbol{\varepsilon})\, \underline{h}(\mathbf{u}|\boldsymbol{\varepsilon})|^{1/2}\, d\mathbf{u}\right\}^{1/2}. \quad (25)$$

Denote by ϕ the density of standard normal distribution. Direct calculations and Lemma 1 show

$$\int |\phi(u_m/\tau_{k\,m})\, \phi(u_m/\bar{\tau}_{k\,m})|^{1/2}\, du_m = \left|\frac{2\,\tau_{k\,m}\,\bar{\tau}_{k\,m}}{\tau_{k\,m}^2 + \bar{\tau}_{k\,m}^2}\right|^{1/2} = \Upsilon(\tau_{k\,m}/\bar{\tau}_{k\,m}),$$

where Υ is defined in Lemma 1 in the appendix. Note that $\Omega_{m,n}$ doesn't have any restriction on z_m^*, $\underline{z}_m^*$ or ε_{km}. Thus

$$\begin{aligned}
\int_{\Omega_{m,n}} |h(\mathbf{u}|\boldsymbol{\varepsilon})\, \underline{h}(\mathbf{u}|\boldsymbol{\varepsilon})|^{1/2}\, d\mathbf{u} &= \int_{\Omega_{m,n}} \prod_{\ell=1}^{m-1} |\phi(u_\ell/\tau_{k\,\ell})\, \phi(u_\ell/\bar{\tau}_{k\,\ell})|^{1/2}\, du_1 \cdots du_{m-1} \\
&\quad \int |\phi(u_m/\tau_{k\,m})\, \phi(u_m/\bar{\tau}_{k\,m})|^{1/2}\, du_m \\
&= \int_{\Omega_{m,n}} \prod_{\ell=1}^{m-1} |\phi(u_\ell/\tau_{k\,\ell})\, \phi(u_\ell/\bar{\tau}_{k\,\ell})|^{1/2}\, du_1 \cdots du_{m-1}\; \Upsilon(\tau_{k\,m}/\bar{\tau}_{k\,m}) \\
&\ge \Upsilon(e^{A_n/2}) \int_{\Omega_{m,n}} \prod_{\ell=1}^{m-1} |\phi(u_\ell/\tau_{k\,\ell})\, \phi(u_\ell/\bar{\tau}_{k\,\ell})|^{1/2}\, du_1 \cdots du_{m-1} \\
&= \Upsilon(e^{A_n/2}) \int_{\Omega_{m-1,n}} \prod_{\ell=1}^{m-1} |\phi(u_\ell/\tau_{k\,\ell})\, \phi(u_\ell/\bar{\tau}_{k\,\ell})|^{1/2}\, du_1 \cdots du_{m-1} \\
&\quad - \Upsilon(e^{A_n/2}) \int_{\Omega_{m-1,n}\cap[|M_{m-1}|>A_n]} \prod_{\ell=1}^{m-1} |\phi(u_\ell/\tau_{k\,\ell})\, \phi(u_\ell/\bar{\tau}_{k\,\ell})|^{1/2}\, du_1 \cdots du_{m-1}, \qquad (26)
\end{aligned}$$

where the third equation is due to the fact that on $\Omega_{m,n}$, $\tau_{k\,m}/\bar{\tau}_{k\,m}$ is bounded below from e^{-A_n} and above by e^{A_n}, and thus by Lemma 1 (b), $\Upsilon(\tau_{k\,m}/\bar{\tau}_{k\,m})$ is bounded from below by $\Upsilon(e^{A_n/2})$, and the fourth equation is from the fact that $\Omega_{m,n} = \Omega_{m-1,n} \setminus [|M_{m-1}| > A_n]$. However, the second integral on the right hand side of (26)

$$\begin{aligned}
&\int_{\Omega_{m-1,n}\cap[|M_{m-1}|>A_n]} \prod_{\ell=1}^{m-1} |\phi(u_\ell/\tau_{k\,\ell})\, \phi(u_\ell/\bar{\tau}_{k\,\ell})|^{1/2}\, du_1 \cdots du_{m-1} \\
&\le \left\{\int_{\Omega_{m-1,n}\cap[|M_{m-1}|>A_n]} \prod_{\ell=1}^{m-1} \phi(u_\ell/\tau_{k\,\ell})\; du_1 \cdots du_{m-1}\right\}^{1/2} \\
&\qquad \left\{\int_{\Omega_{m-1,n}\cap[|M_{m-1}|>A_n]} \prod_{\ell=1}^{m-1} \phi(u_\ell/\bar{\tau}_{k\,\ell})\, du_1 \cdots du_{m-1}\right\}^{1/2} \\
&= P(\Omega_{m-1,n} \cap [|M_{m-1}| > A_n]), \qquad (27)
\end{aligned}$$

where the first inequality is from Cauchy-Schwartz inequality, and the second equation is due to the fact that M_{m-1} and $\Omega_{m-1,n}$ depend on ε_i whose distributions are the same under both models for z_i and $\underline{z}_i$. Substituting (27) into (26) and using $\Upsilon(e^{A_n/2}) \leq 1$ implied by Lemma 1, we obtain that

$$\begin{aligned}\int_{\Omega_{m,n}} |h(\mathbf{u}|\boldsymbol{\varepsilon})\, \underline{h}(\mathbf{u}|\boldsymbol{\varepsilon})|^{1/2}\, d\mathbf{u} \;\;&\leq\;\; \Upsilon(e^{A_n/2}) \int_{\Omega_{m-1,n}} \prod_{\ell=1}^{m-1} |\phi(u_\ell/\tau_{k\,\ell})\, \phi(u_\ell/\bar{\tau}_{k\,\ell})|^{1/2}\, du_1 \cdots du_{m-1} \\ &\quad - P(\Omega_{m-1,n} \cap [|M_{m-1}| > A_n]).\end{aligned}$$

Repeatedly applying the above procedure to the successive integrals, we get the following relation

$$\begin{aligned}\int_{\Omega_{m,n}} |h(\mathbf{u}|\boldsymbol{\varepsilon})\, \underline{h}(\mathbf{u}|\boldsymbol{\varepsilon})|^{1/2}\, d\mathbf{u} \;\;&\geq\;\; \left[\Upsilon(e^{A_n/2})\right]^m - \sum_{j=1}^{m-1} P(\Omega_{j,n} \cap [|M_j| > A_n]) \\ &= \left[\Upsilon(e^{A_n/2})\right]^m - P(\sup_{1\leq\ell\leq m-1} M_\ell > A_n) \\ &= \left[\Upsilon(e^{A_n/2})\right]^m - P(\Omega^c_{m,n}). \qquad (28)\end{aligned}$$

Plugging (28) into (25) we have

$$\begin{aligned}\int |h(\mathbf{u}|\boldsymbol{\varepsilon}) - \underline{h}(\mathbf{u}|\boldsymbol{\varepsilon})|\, d\mathbf{u} \;\;&\leq\;\; 2\, P(\Omega^c_{m,n}) + \sqrt{8}\left\{P(\Omega_{m,n}) + P(\Omega^c_{m,n}) - \left[\Upsilon(e^{A_n/2})\right]^m\right\}^{1/2} \\ &= 2\, P(\Omega^c_{m,n}) + \sqrt{8}\left\{1 - \left[\Upsilon(e^{A_n/2})\right]^m\right\}^{1/2} \\ &= 2\, P(\Omega^c_{m,n}) + \sqrt{8}\left\{1 - e^{m\, log\, \Upsilon(e^{A_n/2})}\right\}^{1/2}. \qquad (29)\end{aligned}$$

By Lemma 8,

$$P(\Omega^c_{m,n}) \leq C\, \frac{m}{n\, A_n^2},$$

and from Lemma 1,

$$\left\{1 - e^{m\, log\, \Upsilon(e^{A_n/2})}\right\}^{1/2} \sim m^{1/2}\, A_n/2.$$

Substituting these two results into (29) and taking $A_n \sim n^{-1/3}\, m^{1/6} = n^{-1/4}\, r_n^{-1/6}$, we obtain that for some generic constant C_1,

$$\int |h(\mathbf{u}|\boldsymbol{\varepsilon}) - \underline{h}(\mathbf{u}|\boldsymbol{\varepsilon})|\, d\mathbf{u} \leq C_1\, r_n^{-2/3} \to 0.$$

Finally, applying the Dominated Convergence Theorem to the right hand side of (22) proves that $D(\mathbf{z}^*, \underline{\mathbf{z}}^*)$ converges to zero. This completes the proof.

6 Technical lemmas

Lemma 1 *Define function*

$$\Upsilon(x) = \left|\frac{2\,x}{1+x^2}\right|^{1/2}, \qquad x \in [0,\infty).$$

Then

(a) $0 \le \Upsilon(0) \le 1$, $\Upsilon(0) = \Upsilon(\infty) = 0$, *and* $\Upsilon(x)$ *is increasing for* $x < 1$ *and decreasing for* $x > 1$.

(b) For any $a > 0$,

$$sup\left\{\Upsilon(x) : e^{-a} \le x \le e^{a}\right\} \ge \left|\frac{2\,e^{a}}{1+e^{2\,a}}\right|^{1/2}.$$

(c) As $a \to 0$,

$$log\,\Upsilon(e^{a}) = log\left|1 - \frac{(e^{a}-1)^2}{1+e^{2\,a}}\right|^{1/2} \sim -(e^{a}-1)^2/4 \sim -a^2/4.$$

Lemma 1 can be easily verified by direct calculations.

Lemma 2 *For any* A, *we have*

$$D(f,g) \le P_f(A^c) + P_g(A^c) + \sqrt{8}\left\{\frac{P_f(A)+P_g(A)}{2} - \int_A |f(u)\,g(u)|^{1/2}\,du\right\}^{1/2},$$

where P_f *and* P_g *denote the probability measures with densities* f *and* g, *respectively.*

Proof.

$$\begin{aligned} D(f,g) &= P_f(A^c) + P_g(A^c) + \int_A |f^{1/2}(u) - g^{1/2}(u)|\,|f^{1/2}(u) + g^{1/2}(u)|\,du \\ &\le P_f(A^c) + P_g(A^c) + \left\{\int_A |f^{1/2}(u) - g^{1/2}(u)|^2\,du \int_A |f^{1/2}(u) + g^{1/2}(u)|^2\,du\right\}^{1/2} \\ &\le P_f(A^c) + P_g(A^c) + 2\left\{\int_A |f^{1/2}(u) - g^{1/2}(u)|^2\,du\right\}^{1/2} \\ &= P_f(A^c) + P_g(A^c) + \sqrt{8}\left\{\frac{P_f(A)+P_g(A)}{2} - \int_A |f(u)\,g(u)|^{1/2}\,du\right\}^{1/2}. \end{aligned}$$

Lemma 3

$$log\,\sigma_t^2 = e^{\beta_1\,t}\left\{log\sigma_0^2 + \beta_2 \int_0^t e^{-\beta_1\,s} dW_{2,s} + \frac{\beta_0}{\beta_1}(1 - e^{-\beta_1 t})\right\}, \tag{30}$$

and

$$log\,\rho_i^2 = \alpha_1^i\, log\sigma_0^2 + \beta_2\, \alpha_1^{-1} \sum_{j=1}^{i} \alpha_1^{i-j}\, \gamma_j/\sqrt{n} + \alpha_0\, \alpha_1^{-1} \sum_{j=1}^{i} \alpha_1^{i-j}, \tag{31}$$

where σ_t^2 and ρ_i^2 are the respective conditional variances of the diffusion process defined by (7)-(8) and the SV process defined by (2)-(3), and here $\alpha_0 = \beta_0\,/n$, $\alpha_1 = 1 + \beta_1\,/n$.

Proof. For σ_t^2, applying Itô lemma (Ikeda and Watanabe 1989, Karatzas and Shreve 1997) to the process given by the lemma, we have

$$\begin{aligned} dlog\,\sigma_t^2 &= \beta_1\, e^{\beta_1\, t}\, dt \left\{ log\,\sigma_0^2 + \beta_2 \int_0^t e^{-\beta_1\, s}\, dW_{2,s} + \beta_0 \int_0^t e^{-\beta_1\, s}\, ds \right\} \\ &\quad + e^{\beta_1\, t} \left\{ \beta_2 e^{-\beta_1\, t}\, dW_{2,t} + \beta_0 e^{-\beta_1\, t}\, dt \right\} \\ &= (\beta_0 + \beta_1\, log\,\sigma_t^2)\, dt + \beta_2\, dW_{2,t}. \end{aligned}$$

Thus, $log\,\sigma_t^2$ given in (30) is the solution of (8).

We can verify the expression for ρ_i^2 by applying (3) recursively or by an inductive argument. In fact, for $i = 1$ formulas (3) and (31) agree. And, substituting (31) for $i - 1$ into (3) yields

$$\begin{aligned} \log\rho_i^2 &= \alpha_0 + \alpha_1[\alpha_1^{i-1} \log\sigma_0^2 + \beta_2\alpha_1^{-1} \sum_{j=1}^{i-1} \alpha_1^{i-j}\gamma_j/\sqrt{n} + \alpha_0\alpha_1^{-1} \sum_{j=1}^{i-1} \alpha^{i-j}] + \alpha_2\gamma_i/\sqrt{n}. \\ &= \alpha_0\alpha_1^{-1} \sum_{j=1}^{i} \alpha^{i-j} + \alpha_1^i \log\sigma_0^2 + \alpha_2\alpha_1^{-1} \sum_{j=1}^{i} \alpha_1^{i-j}\gamma_j/\sqrt{n}, \end{aligned}$$

as desired.

Lemma 4 *Let $t_i = i/n, i = 1, \ldots.$ Then*

$$\sup_{1\le i\le n} |log\,\rho_i^2 - log\,\sigma_{t_i}^2| = O_p(\frac{1}{n}),$$

Proof. Evaluate (31) in terms of β_0, β_1 and evaluate sums to get

$$\log\rho_i^2 = e^{\beta_1 i/n} \left\{ \log\sigma_0^2 + \frac{\beta_0}{\beta_1}(1 - e^{-\beta_1 i/n}) + \beta_2 \sum_j [e^{-\beta_1 j/n} + O(\frac{1}{n})] \frac{\gamma_j}{\sqrt{n}} \right\} + O(\frac{1}{n})$$

with the $O(\frac{1}{n})$ terms being uniform over Θ, i, j. Now, as employed in the proof of Theorem 1, let

$$\frac{\gamma_j}{\sqrt{n}} = W_{2,j/n} - W_{2,(j-1)/n} = \int_{(j-1)/n}^{j} dW_{2,s}$$

then the expression for $\log \rho_i^2$ can be rewritten as

$$\log \rho_i^2 = e^{\beta_1 t_i} \left\{ \log \rho_0^2 + \frac{\beta_0}{\beta_1}(1 - e^{-\beta_1 t_i}) + \beta_2 \int_0^t (e^{-\beta s} + O(\frac{1}{n}))dW_{2,s} \right\} + O(\frac{1}{n}).$$

Comparing this to (30) completes the proof of the lemma since

$$\sup_t | \int_0^t h(s) dW_{2,s}| = O_p(1)$$

for any bounded function h.

Lemma 5

$$\sup_{1 \le i \le n} |log\, \sigma_i^2 - log\, \bar{\sigma}_i^2| = O_p(n^{-1/2}\, log^{1/2}\, n),$$

where

$$\bar{\sigma}_i^2 = \bar{\sigma}_{t_i}^2 = \frac{1}{n} \int_{t_{i-1}}^{t_i} \sigma_u^2\, du.$$

Proof. First we show that for $t = t_i$,

$$\bar{\sigma}_t^2 = \sigma_t^2 \int_0^1 exp \left(-\beta_2\, \lambda_n^{1/2} \int_0^u e^{\beta_1 v}\, d\tilde{W}_{2,v} \right) du + O_p(n^{-1}), \tag{32}$$

where $\lambda_n = 1/n$, and

$$\tilde{W}_{2,u} = \lambda_n^{-1/2}\, (W_{2,t} - W_{2,t-\lambda_n u})$$

is the rescaled Brownian motion. From the definition of $\bar{\sigma}^2$ and the expression of σ_t^2 given in Lemma 3 we have

$$\begin{aligned}
\bar{\sigma}_t^2 &= \int_0^1 \sigma_{t-\lambda_n u}^2\, du \\
&= \int_0^1 exp \Big(e^{-\beta_1 \lambda_n u}\, log\, \sigma_t^2 - e^{\beta_1 (t - \lambda_n u)} \Big\{ \beta_2 \int_{t-\lambda_n u}^t e^{-\beta_1 h}\, dW_{2,h} + \\
&\qquad \beta_0 \int_{t-\lambda_n u}^t e^{-\beta_1 h}\, dh \Big\} \Big)\ du \\
&= \sigma_t^2 \int_0^1 exp \left(-e^{-\beta_1 \lambda_n u} \left\{ \beta_2\, \lambda_n^{1/2} \int_0^u e^{\beta_1 v}\, d\tilde{W}_{2,v} + \beta_0\, \lambda_n \int_0^u e^{\beta_1 v}\, dv \right\} \right)\ du + O_p(\lambda_n) \\
&= \sigma_t^2 \int_0^1 exp \left(-\beta_2\, \lambda_n^{1/2} \int_0^u e^{\beta_1 v}\, d\tilde{W}_{2,v} \right)\ du + O_p(\lambda_n).
\end{aligned}$$

As $\tilde{W}_2$ is a Brownian motion, $\int_0^u e^{\beta_1 v}\, d\tilde{W}_{2,v}$ is normally distributed with mean zero and variance

$$\int_0^u e^{2\, \beta_1\, v}\, dv = (2\, \beta_1)^{-1} (e^{2\, \beta_1\, u} - 1).$$

Thus, $\int_0^1 exp\left(-\beta_2\,\lambda_n^{1/2}\int_0^u e^{\beta_1\,v}\,d\tilde{W}_{2,v}\right)\,du$ is of order $1+O_p(n^{-1/2})$. Combing this result with (32) we obtain

$$\bar{\sigma}_t^2=\sigma_t^2\left\{1+O_p(n^{-1/2})\right\}+O_p(n^{-1})=\sigma_t^2+O_p(n^{-1/2}).$$

Now the lemma is a direct consequence of the above relation and Lemma 4.

Lemma 6

$$log\,\tau_i^2=\alpha_1^{i-1}\,log\tau_0+\alpha_2\,\alpha_1^{-1}\sum_{j=1}^{i-1}\alpha_1^{i-j}\,\xi_j+\alpha_0\,\alpha_1^{-1}\sum_{j=1}^{i-1}\alpha_1^{i-j},$$

and

$$log\,\bar{\tau}_i^2=\alpha_1^{i-1}\,log\tau_0+\alpha_2\,\alpha_1^{-1}\sum_{j=1,j\neq k\ell}^{i-1}\alpha_1^{i-j}\,\xi_j+\alpha_0\,\alpha_1^{-1}\sum_{j=1}^{i-1}\alpha_1^{i-j}.$$

where τ_i^2 and $\bar{\tau}_i^2$ are the respective conditional variances of the MGARCH process (5)-(6) and the hybrid process given by (16)-(18), $\alpha_0=\beta_0\,\lambda_n$, $\alpha_1=1+\beta_1\,\lambda_n$ and $\alpha_2=\beta_2\,\lambda_n^{1/2}$.

Proof. The expressions for τ_i^2 and $\bar{\tau}_i^2$ can be easily obtained by recursively applying (6) and (17)-(18), respectively.

Lemma 7

$$\sup_{1\le i\le n}|log\,\sigma_i^2-log\,\tau_i^2|=O_p(n^{-1/2}\,log\,n),$$

Proof. Applying KMT's strong approximation to the partial sum process of δ_i in the formula for $log\,\sigma_i^2$ given by Lemma 3 and the partial sum process of ξ_i in the expression for $log\,\tau_i^2$ in Lemma 6, we can show

$$\sup_{1\le i\le n}|log\,\sigma_i^2-log\,\tau_i^2|=O_p(n^{-1/2}\,log\,n).$$

Lemma 8

$$P(\Omega_{m,n}^c)\le\frac{C\,m}{n\,A_n^2},$$

where C is a generic constant, and M_ℓ and $\Omega_{j,n}$ are defined in (23) and (24), respectively.

Proof. From the definition of M_ℓ in (23) we have

$$M_\ell=log\,\tau_{k\,\ell}^2-log\,\bar{\tau}_{k\,\ell}^2=\alpha_2\,\alpha_1^{-1}\sum_{l=1}^{\ell-1}\alpha_1^{k\,\ell-k\,l}\,\xi_{k\,l}$$

and

$$\Omega_{j,n} = \left[\sup_{1 \le \ell \le j-1} M_\ell \le A_n \right],$$

where $\alpha_0 = \beta_0 \, \lambda_n$, $\alpha_1 = 1 + \beta_1 \, \lambda_n$, $\alpha_2 = \beta_2 \, \lambda_n^{1/2}$, and $\xi_{k\,\ell}$ are i.i.d., direct calculations show that for $\ell = 1, \cdots, m$,

$$\begin{aligned} E(M_\ell^2) &= \alpha_2^2 \, \alpha_1^{-2} \sum_{l=1}^{\ell-1} \alpha_1^{2\,k\,(\ell-l)} \, E(\xi_{k\,l}^2) \\ &= C \, \alpha_2^2 \, \alpha_1^{-2} \sum_{l=1}^{\ell-1} \alpha_1^{2\,k\,(\ell-l)} \\ &\le C/k = C \, m/n. \end{aligned}$$

Now the lemma is a direct application of Kolomogorov inequality.

Lemma 9

$$\sup_{1 \le \ell \le m} |log \, \tau_{k\,\ell}^2 - log \, \bar{\tau}_{k\,\ell}^2| = O_p(k^{-1/2}),$$

Proof. Taking $A_n = B \, k^{-1/2}$ in Lemma 8 we get

$$P(\sup_{1 \le \ell \le m} M_\ell > B \, k^{-1/2}) \le \frac{C \, m \, k}{n \, B^2} = \frac{C}{B^2}.$$

We complete the proof by letting $B \to \infty$.

REFERENCES

Black, F. and Sholes, M. (1973). The pricing of options and corporate liabilities. Journal of Political Economy **81**, 637-659.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics **31**, 307-327.

Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). ARCH modeling in finance. A review of the theory and empirical evidence. Journal of Econometrics **52**, 5-59.

Brown, L., Cai, T.T, Low, M. and Zhang, C. H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. Ann Statis. **30**, no. 3.

Brown, L. and Low, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. Annals of Statistics **24**, 2384-2398.

Drost, F. C. and Nijman, T. E. (1993). Temporal aggregation of GARCH processes. Economtrica **61**, 909-927.

Duan, J. C. (1995). The GARCH option pricing model. Mathematical Finance **5**, 13-32.

Duan, J. C. (1997). Augmented GARCH (p,q) process and its diffusion limits. Journal of Econometrics **79**, 97-127.

Duffie, D. (1992). Dynamic Asset Pricing Theory. Princeton University Press, Princeton, N.J. Second edition.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica **50**, 987-1007.

Engle, R. F. and Bollerslev, T. (1986). Modeling the persistence of conditional variances (with discussion). Econometric Review **5**, 1-50.

Gouriéroux, C. (1997). ARCH models and Financial Applications. Springer.

Geweke, J. (1986). Modeling the persistence of conditional variances: a comment. Econometric Review **5**, 57-61.

Greenwood, P. E. and Shiryayev, A. N. (1985). Contiguity and the Statistical Invariance Principle. Gordon and Breach, London.

Huebner, M. (1997). A characterization of asymptotic behavior of maximum likelihood estimator for stochastic PDEs. Methods of Mathematical Statistics **6**, 395-415.

Hull, J. (1997). Options, Futures, and Other Derivatives. Prentice, N.J. Third edition.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. Journal of Finance **42**, 281-300.

Kallsen, J. and Taqqu, M. (1998). Option pricing in ARCH-type models. Mathematical Finance **8**, 13-26.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion). J. Bus. Econ. Statist. **12**, 371-417.

Karatzas, I. and Shreve, S. E. (1997). Brownian Motion and Stochastic Calculus. Springer. Second edition.

Komlós, J. Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RVs and the sample DF. I. Z. Wahrsch. Verw. Gebiete **32**, 111-131.

Le Cam, L. (1986). Asymptotic Methods in Statistical Decision Theory. Springer.

Le Cam, L. and Yang, G. (2000). Asymptotics in Statistics – Some Basic Concepts. Springer.

Merton, R. (1990). Continuous-Time Finance. Blackwell, Cambridge.

Nelson, D. B. (1990). ARCH models as diffusion approximations. Journal of Econometrics **45**, 7-38.

Nussbaum, M. (1999). Approximation of statistical experiments for ill-posed function estimation problems. Manuscript.

Pantula, S. G. (1986). Comments on "Modeling the persistence of conditional variances". Econometric Review **5**, 71-74.

Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. Ann Statis. **30**, no. 3.