

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

January 31, 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) first established the link between GARCH models and diffusions by establishing diffusion limits for GARCH processes. Now there are a number of papers in the literature to forge the gap between the two approaches (e.g. Duan 1997). However, most of these papers study only the weak convergence of discrete-time models to diffusions, so the established link implies 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).)

We consider the observation over a time span,  $T$ , at suitable frequencies of SV models, GARCH models and diffusion models. The SV and GARCH models are mathematically constructed in discrete time, and that construction involves a basic frequency of construction,

$\phi_c$ .

Our main result concerns the asymptotic equivalence (as  $n = \phi_c T \rightarrow \infty$ ) of observations at suitable frequencies,  $\phi_0$ , of these processes. We show that for suitable frequencies  $\phi_0$  the GARCH model is asymptotically equivalent to its diffusion limits. We also show that the SV model is asymptotically equivalent to its diffusion limit at all frequencies  $\phi_0 \leq \phi_c$ .

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.

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.

To describe our results more fully, suppose the processes in time interval  $[0, T]$  are based on a GARCH or SV model constructed at  $t_i = \frac{i}{n} T$ ,  $i = 1, \dots, n$ . Thus,  $T/n$  is the basic time interval for the model and  $\phi_c = n/T$  is the corresponding basic frequency. We will study the statistical equivalence of the processes when observed at different frequencies. Namely, we investigate whether the corresponding processes observed at frequency  $\phi_0 = \phi_c/k$  (i.e.  $t_{k\ell} = (\frac{kT}{n})\ell$ ,  $\ell = 1, \dots, m = [n/k]$ ) are asymptotically equivalent, where  $k$  are integers which may depend on  $n$ .

Wang (2001) 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  $\epsilon$  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  $\phi_0 = \phi_c/k$  with  $k \rightarrow \infty$  but  $k = o(n)$ .

We show that for any choice of  $k$ , including  $k = 1$ , the SV models and their diffusion limits are asymptotically equivalent. Meanwhile the GARCH models are asymptotically equivalent to their diffusion limits only at frequencies  $n/(Tk)$  with  $n^{1/2}/k \rightarrow 0$ . In other words, the SV models are asymptotically equivalent to their diffusion limits at any frequencies up to the basic frequency, while GARCH models match to the diffusion limits only at frequencies lower than the square root of the basic frequency.

To better understand these results consider a market modeled via an SV model constructed at five minute intervals over a year. A diffusion model for this same process would be nearly equivalent when observed at these basic intervals or at greater intervals such as at a daily level. GARCH models constructed at these five minute intervals may have nearly equivalent performance with their corresponding diffusions at the monthly level but not at the daily level and perhaps also not at the weekly level. We are thinking here of a market with an 8 hour trading day, 250 days a year. This involves  $n = 24,000$  5-minute intervals. There are 96 of these intervals in a day. Since  $96 < \sqrt{24,000} = 155$  the GARCH

and diffusion models should be expected to differ slightly at this frequency corresponding to observations at a daily level. But they should surely be nearly equivalent at the monthly level, since a month involves about  $2000 \gg 155$  observations. They should even be quite close to equivalent at the weekly level, since a week involves  $480 > 155$  observations.

There is also an interesting converse implication of these results. Suppose a process is actually an SV model constructed at the basic frequency  $\phi_c = n/T$ . Suppose it is observed at a lower frequency  $\phi_c/k$ . One might wish to construct a model for a process constructed at this lower frequency. If  $n^{1/2}/k \rightarrow 0$ , an appropriate GARCH model with the basic frequency  $\phi_c$  would be a satisfactory approximation at the lower frequency to the original SV model. But a GARCH model with construction frequency  $\phi_c/k$  would not be.

The unexpected 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 logged security price  $S_t = \log P_t$  obeys the following stochastic differential equation

$$dS_t = \mu_t dt + \sigma_t dW_t, \quad t \in [0, T], \quad (1)$$

where  $W_t$  is a standard Wiener process.  $\mu_t$  is called the drift in probability or the mean return in finance, and  $\sigma_t^2$  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  $S_t$  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).) We thus rewrite (1) as

$$dS_t = \sigma_t dW_{1,t} \quad t \in [0, T] \quad (2)$$

where  $W_{1,t}$  is a standard Wiener process.

Many econometric studies have documented that financial time series tend to be highly heteroskedastic. To accommodate this one often allows  $\sigma_t^2$  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

$$d \log \sigma_t^2 = (\beta_0 + \beta_1 \log \sigma_t^2) dt + \beta_2 dW_{2,t}, \quad t \in [0, T], \quad (3)$$

with  $W_{2,t}$  an independent standard Wiener process. Such  $\sigma_t^2$  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  $t_i = i\lambda_n, i = 1, \dots, n$ . Then let  $X_i = S_{t_i}$  and define the corresponding difference process by  $x_i = X_i - X_{i-1}$ .

There is no loss of generality here or later in assuming  $T = 1$ , and we will henceforth do so. Then  $\lambda_n = 1/n$ . We also assume that the initial values  $\sigma_0^2$  and  $X_0 = S_0$  are known constants.

## 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  $\varepsilon_i$  and  $\gamma_i$  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  $\lambda_n = 1/n$ . The observations in this SV model are then described by  $Y_0, \dots, Y_n$  where

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

We take  $Y_0 = X_0$  as a known constant. Redefining the constants to correspond to those in (2) and (3) we write the model as

$$y_i = \rho_i \varepsilon_i / \sqrt{n} \quad \text{and} \tag{4}$$

$$\log \rho_i^2 = \frac{\beta_0}{n} + \left(1 + \frac{\beta_1}{n}\right) \log \rho_{i-1}^2 + \beta_2 \frac{\gamma_i}{\sqrt{n}}. \tag{5}$$

## 2.3 GARCH models

Engle (1982) introduced the ARCH model for a time series of increments,  $z_i = Z_i - Z_{i-1}$ ,  $i = 1, \dots, n$ . He sets the conditional variance,  $\tau_i^2$  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  $\tau_i^2$  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  $\tau_i^2$  remains non-negative, by making  $\log \tau_i^2$  linear in some function of time and lagged  $z_i$ 's. Then

$$z_i = \tau_i \varepsilon_i \text{ and}$$

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

where  $\varepsilon_i$  are independent standard normal random variables and  $\alpha$ '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, the MGARCH(1,1) process on  $[0, 1]$  is defined as follows. For i.i.d. standard normal  $\varepsilon_i$ , let  $z_0 = X_0$ .

$$c_0 = E(\log \varepsilon_i^2), \quad c_1 = \{var(\log \varepsilon_i^2)\}^{1/2}, \quad \xi_i = (\log \varepsilon_i^2 - c_0)/c_1. \quad (6)$$

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

$$z_i = Z_i - Z_{i-1} = \tau_i \varepsilon_i / \sqrt{n}, \quad (7)$$

$$\log \tau_i^2 = \frac{\beta_0}{n} + \left(1 + \frac{\beta_1}{n}\right) \log \tau_{i-1}^2 + \beta_2 \xi_i / \sqrt{n}. \quad (8)$$

A continuous time approximating process  $(Z_{n,t}, \tau_{n,t}^2)$ ,  $t \in [0, T]$ , is given by

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

Nelson (1990) showed that as  $n \rightarrow \infty$ , the normalized partial sum process of  $(\varepsilon_i, \xi_i)$  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  $(S_t, \tau_{n,t}^2)$  described in section 2.1. The diffusion model described by (2)-(3) and the corresponding diffusion process is thus called the diffusion limit of the MGARCH process.

### 3 Statistical equivalence

#### 3.1 Comparison of experiments

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

Consider two statistical experiments with the same parameter space  $\Theta$ ,

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

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

$$\Delta(\mathcal{IE}_1, \mathcal{IE}_2) = \max \left\{ \inf_{\delta_1} \sup_{\delta_2} \sup_{\theta \in \Theta} \sup_{L: \|L\|=1} |R_1(\delta_1, L, \theta) - R_2(\delta_2, L, \theta)|, \right. \\ \left. \inf_{\delta_2} \sup_{\delta_1} \sup_{L: \|L\|=1} |R_1(\delta_1, L, \theta) - R_2(\delta_2, L, \theta)| \right\}.$$

Le Cam (1986) and Le Cam and Yang (2000) provide other useful expressions for  $\Delta$ .

Two experiments  $\mathcal{I}E_1$  and  $\mathcal{I}E_2$  are called equivalent if  $\Delta(\mathcal{I}E_1, \mathcal{I}E_2) = 0$ . The equivalence means that for every procedure  $\delta_1$  in problem  $\mathcal{I}E_1$ , there is a procedure  $\delta_2$  in problem  $\mathcal{I}E_2$  with the same risk, uniformly over  $\theta \in \Theta$  and all  $L$  with  $\|L\| = 1$ , and vice versa. Two sequences of statistical experiments  $\mathcal{I}E_{n,1}$  and  $\mathcal{I}E_{n,2}$  are said to be asymptotically equivalent if  $\Delta(\mathcal{I}E_{n,1}, \mathcal{I}E_{n,2}) \rightarrow 0$ , as  $n \rightarrow \infty$ . Thus, for any sequence of procedures  $\delta_{n,1}$  in problem  $\mathcal{I}E_{n,1}$  there is a sequence of procedures  $\delta_{n,2}$  in problem  $\mathcal{I}E_{n,2}$  with risk differences tending to zero uniformly over  $\theta \in \Theta$  and all  $L$  with  $\|L\| = 1$ , i.e.

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

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  $\Delta(\mathcal{I}E_1, \mathcal{I}E_2)$  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

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

(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,

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

and

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

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

We have the following relation between Hellinger distance and  $L_1$  distance

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

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.

As 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 have the observations  $\{x_{k\ell}\}_\ell$  and  $\{z_{k\ell}\}_\ell$  taken from diffusion and MGARCH processes observed at frequency  $n/k$  where  $k/\sqrt{n} \rightarrow \infty$ . These have joint densities  $(f_{\theta,1}, f_{\theta,2})$ , say, where the dependence on  $n$  is suppressed in this notation. We prove that  $D(f_{\theta,1}, f_{\theta,2}) \rightarrow 0$  uniformly over  $\theta \in \Theta$ . Hence  $\Delta(\{x_{k\ell}\}_\ell, \{y_{k\ell}\}_\ell) \rightarrow 0$  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  $\theta$  to distinguish  $\{x_{k\ell}\}_\ell$  from  $\{y_{k\ell}\}_\ell$  if

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

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

### 3.2 MGARCH, SV and Diffusion Experiments

Let  $\beta = (\beta_0, \beta_1, \beta_2)$  be the vector of parameters for the MGARCH, SV and diffusion models defined in section 2.4, and let the parameter space  $\Theta$  consist of  $\beta_i$  belonging to bounded intervals.

From section 2.4, the SV process  $\{y_i\}_{1 \leq i \leq n}$  and the MGARCH approximating process  $\{z_i\}_{1 \leq i \leq n}$  are defined by the stochastic difference equations (4)-(5) and (6)-(8), respectively. For the diffusion limit,  $X_1, \dots, X_n$  denote the discrete samples at  $t_i = i/n$ ,  $i = 1, \dots, n$ , of the diffusion limit  $X_t$  governed by the stochastic differential equations (2)-(3). Define the corresponding difference process by  $x_i = X_i - X_{i-1}$ ,  $i = 1, \dots, n$ .

In Theorem 1 we establish that the SV process  $\{Y_i\}_{1 \leq i \leq n}$  and the discrete version  $\{X_i\}_{1 \leq i \leq n}$  of the approximating diffusion process are asymptotically equivalent. 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 (2001) 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  $D(\{x_i\}_{1 \leq i \leq n}, \{z_i\}_{1 \leq i \leq n})$  nor  $D(\{y_i\}_{1 \leq i \leq n}, \{z_i\}_{1 \leq i \leq n})$  converge to zero. Thus, at the basic frequency MGARCH is not asymptotically equivalent to the other two models. We will study the equivalence of the processes  $\{x_i\}_{1 \leq i \leq n}$  and  $\{z_i\}_{1 \leq i \leq n}$  at lower observational frequencies. Namely, we investigate whether the processes  $\{x_{k\ell}\}_\ell$  and  $\{z_{k\ell}\}_\ell$ ,  $\ell = 1, \dots, m = [n/k]$ , are

asymptotically equivalent for some integers  $k$ , where  $[n/k]$  denotes the integer part of  $n/k$ . For convenience we give the following general definition corresponding to the above notation.

**Definition 1** *We say processes  $\{a_i\}_{1 \leq i \leq n}$  and  $\{b_i\}_{1 \leq i \leq n}$  are asymptotically equivalent at frequency  $n/k$ , if as  $n \rightarrow \infty$ ,*

$$\Delta(\{a_{k\ell}\}_{1 \leq \ell \leq m}, \{b_{k\ell}\}_{1 \leq \ell \leq m}) \rightarrow 0.$$

## 4 Equivalence of diffusions and SV models

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

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

Remark 1. Theorem 1 implies that the SV models are asymptotically equivalent to their diffusion limits at the frequency of construction. This consequently shows they are asymptotic equivalent also at any lower frequency.

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 (7)-(8), we can easily derive that conditional on  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)$ ,  $y_i$  are independent with  $y_i$  conditionally following a normal distribution with mean zero and variance  $\sigma_i^2$ . Thus,

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

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 (5)-(6) 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  $\gamma$  and the process  $W_2$  can be realized on a common space by writing  $\gamma = \gamma(W_2)$  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

$$|\log \rho_i^2 - \log \sigma_{t_i}^2| = O_p\left(\frac{1}{n}\right) \quad i = 1, \dots, n$$

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

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

## 5 Equivalence of diffusions and MGARCH models

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

$$\Delta(\{x_{k\ell}\}_\ell, \{z_{k\ell}\}_\ell) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Remark 2. Theorem 2 shows that the MGARCH processes are equivalent to their diffusion limits only at frequencies  $n^{1/2}/(T r_n)$  for any  $r_n \rightarrow \infty$ . Taking  $r_n$  diverge arbitrarily slowly,

we have that although the MGARCH models and their diffusion limits are not asymptotically equivalent at the basic frequency  $\phi_c = n/T$ , they are asymptotically equivalent at frequencies lower than the square root of the basic frequency.

Remark 3. Comparing Theorems 1 and 2, we see that the MGARCH models start to be asymptotically equivalent to their diffusion limits at frequencies much lower than the SV models. This is due to the noise propagation systems in their conditional variances. The MGARCH models utilize past observational errors to model their conditional variances, while the conditional variances of the SV models are governed by i.i.d. normal random variables, which is a discrete version of the white noise used by the diffusions to model their conditional variances. Because of the mimicking of white noise by i.i.d. normal errors, the SV models are much closer to the diffusion limits than the MGARCH models. Thus, the SV models can be asymptotically equivalent to the diffusion limits at higher frequencies than the MGARCH models.

Remark 4. The equivalence result in Theorem 2 reveals that MGARCH models at frequencies asymptotically lower than  $n^{1/2}$  are no longer ARCH or GARCH, but instead they behave like SV models. 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) which showed that GARCH structures begin to break down at some lower frequencies. More precisely, our result reveals explicitly that the structures of the MGARCH models at frequencies lower than  $n^{1/2}$  are similar to those of SV models.

Proof. Define a hybrid process as follows,

$$\tilde{z}_i = \bar{\tau}_i \varepsilon_i, \quad i = 1, \dots, n, \quad (16)$$

$$\log \bar{\tau}_{k\ell+1} = \alpha_0 + \alpha_1 \log \bar{\tau}_{k\ell}, \quad \ell = 1, \dots, m, \quad (17)$$

and for  $1 \leq i \leq n$  and  $i \neq k\ell + 1$  with  $1 \leq \ell \leq m$ ,

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

where  $\xi_i$  are defined in (6),  $\alpha_0 = \beta_0 \lambda_n$ ,  $\alpha_1 = 1 + \beta_1 \lambda_n$ , 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  $\ell = 1, \dots, m = [n/k]$ , let

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

and

$$\mathbf{x}^* = (x_1^*, \dots, x_m^*), \quad \mathbf{y}^* = (y_1^*, \dots, y_m^*), \quad \mathbf{z}^* = (z_1^*, \dots, z_m^*), \quad \underline{\mathbf{z}}^* = (\underline{z}_1^*, \dots, \underline{z}_m^*).$$

Let  $\boldsymbol{\varepsilon} = \{\varepsilon_i : 1 \leq i \leq n, i \neq k\ell \ \ell = 1, \dots, m\}$ , that is,  $\boldsymbol{\varepsilon}$  consists of all  $\varepsilon_i$  whose index  $i$  is not a multiple of  $k$ . From the framework of the MGARCH process defined by (7)-(8), we see a one-to-one relationship between  $\{z_1, \dots, z_n\}$  and  $\{\boldsymbol{\varepsilon}, \mathbf{z}^*\}$ , and thus the distribution of  $z_1, \dots, z_n$  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

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

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

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

From the definition of  $\underline{z}_i$  given by (16)-(18), the conditional variance of  $\underline{\mathbf{z}}^* = (\underline{z}_1^*, \dots, \underline{z}_m^*)$  depend on only  $\{\underline{z}_i, 1 \leq i \leq n, i \neq k\ell, \ell = 1, \dots, m\}$ , or equivalently,  $\boldsymbol{\varepsilon}$ . Thus, conditional on

$\boldsymbol{\varepsilon}, \underline{z}_1^*, \dots, \underline{z}_m^*$  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  $D(\mathbf{y}^*, \mathbf{z}^*) \leq D(\mathbf{y}^*, \underline{\mathbf{z}}^*) + D(\mathbf{z}^*, \underline{\mathbf{z}}^*)$ , 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

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

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

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

Applying successive conditional arguments to the GARCH process  $z_i$  defined by (7)-(8), we derive that the joint conditional distribution of  $\mathbf{z}^* = (z_1^*, \dots, z_m^*)$  given  $\boldsymbol{\varepsilon}$  is a product of  $N(0, \tau_{k\ell}^2)$ , where  $\tau_{k\ell}^2$  depends on  $z_1^*, \dots, z_{\ell-1}^*$  and  $\varepsilon_i$  for  $1 \leq i < k\ell$  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  $\varepsilon_i$ , where  $1 \leq i < k\ell$  and  $i$  is not a multiple of  $k$ .

Let

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

and define events

$$\Omega_{j,n} = \left\{ \sup_{1 \leq \ell \leq j-1} M_\ell \leq A_n \right\}, \quad j = 2, \dots, m, \quad (24)$$

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  $\varepsilon_i$  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

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

Denote by  $\phi$  the density of standard normal distribution. Direct calculations and Lemma 1 show

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

where  $\Upsilon$  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  $\varepsilon_{km}$ . Thus

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

$$- \Upsilon(e^{A_n/2}) \int_{\Omega_{m-1,n} \cap [|M_{m-1}| > A_n]} \prod_{\ell=1}^{m-1} |\phi(u_\ell/\tau_{k\ell}) \phi(u_\ell/\bar{\tau}_{k\ell})|^{1/2} du_1 \cdots du_{m-1}, \quad (26)$$

where the third equation is due to the fact that on  $\Omega_{m,n}$ ,  $\tau_{km}/\bar{\tau}_{km}$  is bounded below from  $e^{-A_n}$  and above by  $e^{A_n}$ , and thus by Lemma 1 (b),  $\Upsilon(\tau_{km}/\bar{\tau}_{km})$  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)

$$\begin{aligned} & \int_{\Omega_{m-1,n} \cap [|M_{m-1}| > A_n]} \prod_{\ell=1}^{m-1} |\phi(u_\ell/\tau_{k\ell}) \phi(u_\ell/\bar{\tau}_{k\ell})|^{1/2} du_1 \cdots du_{m-1} \\ & \leq 0.5 \int_{\Omega_{m-1,n} \cap [|M_{m-1}| > A_n]} \prod_{\ell=1}^{m-1} \phi(u_\ell/\tau_{k\ell}) du_1 \cdots du_{m-1} \\ & \quad + 0.5 \int_{\Omega_{m-1,n} \cap [|M_{m-1}| > A_n]} \prod_{\ell=1}^{m-1} \phi(u_\ell/\bar{\tau}_{k\ell}) du_1 \cdots du_{m-1} \\ & = P(\Omega_{m-1,n} \cap [|M_{m-1}| > A_n]), \end{aligned} \quad (27)$$

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  $\varepsilon_i$  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

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

Plugging (28) into (25) we have

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

By Lemma 8,

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

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  $C_1$ ,

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

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

## 6 Technical lemmas

**Lemma 1** *Define function*

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

*Then*

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

(b) For any  $a > 0$ ,

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

(c) As  $a \rightarrow 0$ ,

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

Lemma 1 can be easily verified by direct calculations.

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

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

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

Proof.

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

**Lemma 3**

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

and

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

where  $\sigma_t^2$  and  $\rho_t^2$  are the respective conditional variances of the diffusion process defined by

(2)-(3) and the SV process defined by (4)-(5), and here  $\alpha_0 = \beta_0/n$ ,  $\alpha_1 = 1 + \beta_1/n$ .

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

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

Thus,  $\log \sigma_t^2$  given in (30) is the solution of (3).

We can verify the expression for  $\rho_i^2$  by applying (5) recursively or by an inductive argument. In fact, for  $i = 1$  formulas (5) and (31) agree. And, substituting (31) for  $i - 1$  into (5) yields

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

as desired.

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

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

Proof. Evaluate (31) in terms of  $\beta_0, \beta_1$  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  $\Theta, i, j$ . 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

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

for any bounded function  $h$ .

**Lemma 5**

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

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  $t = t_i$ ,

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

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  $\sigma_t^2$  given

in Lemma 3 we have

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

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

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

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

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

where  $\tau_i^2$  and  $\bar{\tau}_i^2$  are the respective conditional variances of the MGARCH process (7)-(8) 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  $\tau_i^2$  and  $\bar{\tau}_i^2$  can be easily obtained by recursively applying (8) and (17)-(18), respectively.

**Lemma 7**

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

Proof. Applying KMT's strong approximation to the partial sum processes of  $\delta_i$  and  $\xi_i$ , and using the expressions for  $\log \sigma_i^2$  and  $\log \tau_i^2$  in Lemmas 3 and 6, we have

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

**Lemma 8**

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

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

Proof. From the definition of  $M_\ell$  in (23) we have

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

and

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

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  $\ell = 1, \dots, m$ ,

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

Now the lemma is a direct application of Kolomogorov inequality.

**Lemma 9**

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

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

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

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

## 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. (2001). Asymptotic equivalence theory for nonparametric regression with random design. Manuscript.
- 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. *Econometrica* **61**, 909-927.
- 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. *Econo-*

- metric Review **5**, 57-61.
- Greenwood, P. E. and Shiriyayev, 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.
- 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 (2001). Asymptotic nonequivalence of GARCH models and diffusions. To be published  
in *Annals of Statistics*.