# Asymptotic Nonequivalence of GARCH Models and Diffusions

Yazhen Wang

University of Connecticut

**Dedicated to the Memory of Lucien Le Cam 1924-2000**

To appear in Annals of Statistics

September, 2000

Running Head: GARCH Models and Diffusions

Yazhen Wang is an associate professor at Department of Statistics, University of Connecticut, Storrs, CT 06269, U.S.A. He would like to thank Peter Bickel, Larry Brown, Lucien Le Cam and Michael Nussbaum, three anonymous referees and Editor Hans R. Künsch for comments and suggestions, which led to significant improvements in both the substance and the style of the paper. The research is partially supported by NSA Grant MDA 904-99-1-0028.

## Abstract

This paper investigates the statistical relationship of the GARCH model and its diffusion limit. Regarding the two types of models as two statistical experiments formed by discrete observations from the models, we study their asymptotic equivalence in terms of Le Cam's deficiency distance. To our surprise, we are able to show that the GARCH model and its diffusion limit are asymptotically equivalent only under deterministic volatility. With stochastic volatility, due to the difference between the structure with respect to noise propagation in their conditional variances, their likelihood processes asymptotically behave quite differently and thus they are not asymptotically equivalent. This stochastic nonequivalence discredits a general belief that the two types of models are asymptotically equivalent in all respects and warns against the common financial practice that applies statistical inferences derived under the GARCH model to its diffusion limit.



Key Words: Black-Scholes, comparison of experiments, conditional variance, deficiency distance, GARCH model, financial modeling, likelihood process, stochastic differential equation, stochastic volatility.

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

# 1 Introduction

There are two relatively independent strands of financial modeling: continuous-time models typically used in theoretical finance and discrete-time models favored for empirical work. The continuous-time models are dominated by the diffusion approach. Most of the discrete-time models are of the autoregressive conditionally heteroscedastic (ARCH) type. Historically, the two literatures on the discrete-time and continuous-time models have developed quite independently. In the early 1990s researchers started to reconcile the two modeling approaches. Nelson (1990) first established the continuous-time diffusion limit for the discrete-time ARCH model by showing that generalized ARCH (GARCH) processes weakly converge to some bivariate diffusions, as the length of the discrete time intervals goes to zero. Duan (1997) proposed an augmented GARCH model to unify various parametric GARCH models and derived its diffusion limit and among others. The existing theory links the two types of models by weak convergence. Given that statistical inference is essential for both types of modeling, this paper investigates the statistical relationship between the GARCH model and its diffusion limit.

Since the GARCH model and its diffusion limit share the same parameters, we naturally treat them as two statistical experiments formed by discrete observations from the two models and study their asymptotic equivalence by Le Cam's deficiency distance. Here equivalence means that each statistical procedure for one model has a corresponding equal-performance statistical procedure for another model. Surprisingly, it is shown that the GARCH model and its diffusion limit are asymptotically equivalent under nonstochastic volatility and not otherwise. Stochastic volatility is the essential characteristic of ARCH modeling. The GARCH model with nonstochastic volatility is nothing but a regression model and its diffusion limit is a white noise model. Thus, the stochastic nonequivalence indicates that Nelson's weak convergence result has no analog in Le Cam's paradigm.

The GARCH and diffusion models are of great interest in finance. It is very common to hear researchers invoke Nelson's result to justify the common belief that both models are "more or less equivalent." Our results send a warning against accepting this belief uncritically. The reconciliation of the two modeling approaches in financial econometrics and financial mathematics is in terms of the weak convergence results for the distributions of price processes. The statistical equivalence considered here is essentially determined by the asymptotic behavior of the likelihoods. Due to the difference between distribution and likelihood, weak convergence results like Nelson's do not necessarily imply asymptotic equivalence of the GARCH and diffusion models viewed as statistical experiments.

There is a great volume of statistical literature on comparison of experiments (Basawa and Prakasa Rao 1980, Brown and Low 1996, Jacod and Shiryaev 1987, Le Cam 1986, Le Cam and Yang 1990, Nussbaum 1996). Most of the work on the equivalence of experiments is about diffusions with known fixed diffusion variance but unknown drift. For the inference of diffusion drift, the Cameron-Martin-Girsanov theorem gives an explicit form for the likelihood process of continuous-time observations from a diffusion. Under Le Cam's deficiency distance, the continuous-time observations are asymptotically equivalent to their discretely sampled versions, and they both, in turn, are asymptotically equivalent to observations from the corresponding discrete-time model. However, the inference for diffusion variance is intrinsically different (Florens-Zmirou 1989, Genon-Catalot 1990, Kessler 1997, Prakasa Rao 1988, Yoshida 1992). For example, singularity may occur for the infinite-dimensional distributions of continuous-time observations from a diffusion under different values for the parameters in its diffusion variance, and the diffusion variance can be perfectly recoverered from continuous-time observations. Thus, it is impossible to have any statistical equivalence between continuous-time observations and discretely sampled observations.

Our explanation for the unexpected phenomenon is as follows. The GARCH and diffusion models employ quite different mechanisms to propagate noise in their conditional variances. In the GARCH framework, the conditional variance is governed by the squares of past observational errors, so the likelihood involves a triple: normal random errors, their squares, and the GARCH conditional variances, where the last two are correlated. While the diffusion model uses an independent, unobservable white noise to govern its conditional variance, and thus its likelihood behaves like the conditional expectation with respect to the unobservable white noise of the GARCH likelihood with the correlated triple replaced by three uncorrelated random components: normal random errors, their squares, and the diffusion conditional variances. Hence, with stochastic volatility, the different noise propagation systems in their conditional variances result in quite different asymptotic likelihoods for the two types of models, and that, in turn, causes the nonequivalence. With nonstochastic volatility, the two models are Gaussian and their deterministic conditional variances approach the same limit. Thus, they are asymptotically equivalent.

The stochastic nonequivalence has an important consequence for statistical inference in the GARCH and diffusion models, and in particular provides some theoretical evidence against the practice that applies statistical procedures derived under the GARCH model to its diffusion limit. In a diffusion model, the conditional volatility is not observable, the likelihood is extremely hard to obtain, and parameter estimation can be very difficult; while a GARCH model uses past observations to model the conditional variance, the likelihood has an explicit expression, and parameters can easily be estimated. This makes the GARCH approach more attractive for estimation and subsequent statistical inference. With diffusion modeling being favored over GARCH modeling for option pricing, one may be tempted to use a diffusion model with parameters estimated by fitting its corresponding GARCH model, to apply the statistical inference procedures developed under the GARCH model to the diffusion model, and to plug the set of estimated parameter values from the GARCH model into formulas for option pricing obtained from the diffusion model. In fact, this naive approach is enthusiastically advocated in the finance literature and widely used in financial practice. The usual justification is that weak convergence results like Nelson's suggest that parameter estimators obtained by fitting a GARCH model can consistently estimate the parameters in its diffusion limit. However, due to the stochastic nonequivalence of the two types of models, estimators and tests derived under the GARCH model may behave asymptotically quite differently from those derived under its diffusion limit and can have inferior performance when being applied to observations coming from its diffusion limit. For modeling stochastic volatility, if a diffusion model is preferred, it is statistically more efficient to fit data directly to the diffusion model and carry out the inference (Aït-Sahalia 1996, Danielsson 1994, Gallant, Hsieh and Tauchen 1997, Gallant and Long 1997, Gallant and Tauchen 1998, Jacquier, Polson and Rossi 1994).

Our approach to proving the results is to follow Le Cam's principle to study likelihood processes and derive the limit of the deficiency distance between the two types of models. The rest of the paper is organized as follows. Section 2 introduces ARCH models and relates them to diffusions. Section 3 briefly reviews comparison of experiments and then presents our main results. Sections 4-6 are devoted to proving the results. The likelihood processes for the models are studied in sections 4 and 5, and theorems are proved in section 6.

# 2 ARCH models and diffusions

## 2.1 ARCH models

Probably the most important innovation in discrete-time modeling of financial time series is the introduction by Engle (1982) of the ARCH model. The model makes the conditional variance of a series of prediction errors equal to some function of lagged errors, time, parameters, and predetermined exogenous and lagged endogenous variables. Specifically, the observed time series $x_k$, $k = 1, \cdots, n$, is assumed to follow the model

$$x_k = \mu_k + y_k, \qquad y_k = \sigma_k \, \varepsilon_k, \tag{1}$$

$$\sigma_k^2 = \sigma^2(y_{k-1}, y_{k-2}, \cdots, k, A_k, \boldsymbol{\alpha}) \tag{2}$$

where $\varepsilon_k$ is a sequence of i.i.d. standard normal random variables, $\sigma_k^2$ is the conditional variance of $x_k$ given the information at time $k$, $\mu_k$ is the drift term which may depend on $k$, $\sigma_k^2$, and $x_{k-1}, x_{k-2}, \cdots$, $\boldsymbol{\alpha}$ is a vector of parameters, $A_k$ is a vector of exogenous and lagged endogenous variables. Any model of the form (1)-(2) is referred to as an ARCH model. The existing ARCH models differ in their specification for $\sigma_k^2$.

Engle (1982) chose the following function form for $\sigma_k^2$,

$$\sigma_k^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \, y_{k-j}^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \, \sigma_{k-j}^2 \, \varepsilon_{k-j}^2, \tag{3}$$

where $\alpha_i$'s are nonnegative constants. The model specified by (3) is often called ARCH(p). The appeal of this model lies in the fact that it can capture the tendency for volatility clustering: large (or small) price changes tend to be followed by other large (or small) price changes, but of unpredictable sign. In other words, a high value of $y_k^2$ drives up $\sigma_{k+1}^2$, which in turn increases the expectation of $y_{k+1}^2$, and so on.

Bollerslev (1986) and Bollerslev and Engle (1986) generalized ARCH(p) by introducing the following specification

$$\sigma_k^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, \sigma_{k-i}^2 + \sum_{j=1}^{q} \alpha_{p+j} \, y_{k-j}^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, \sigma_{k-i}^2 + \sum_{j=1}^{q} \alpha_{p+j} \, \sigma_{k-j}^2 \, \varepsilon_{k-j}^2, \tag{4}$$

where $\alpha_i$'s are nonnegative constants. This model is referred to as linear GARCH (p,q). For real financial data it often yields a more parsimonious representation for $\sigma_k^2$ as a function of lagged values of $\sigma_k^2$'s and $y_k^2$'s.

Since $\sigma_k^2$ is the conditional variance, it clearly must be nonnegative with probability one. Linear GARCH models guarantee this by making $\sigma_k^2$ a positive linear combination of positive random variables. To ensure non-negativity of $\sigma_k^2$, Geweke (1986) and Pantula (1986) adopted the following natural device by making $log\, \sigma_k^2$ a linear function of lagged values of $log\, \sigma_k^2$'s and $log\, y_k^2$'s,

$$\log\,\sigma_k^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \,\log\,\sigma_{k-i}^2 + \sum_{j=1}^{q} \alpha_{p+j}\,\log\,\varepsilon_{k-j}^2, \tag{5}$$

where $\alpha_i$'s are constants, and because of $\log\,\varepsilon_{k-j}^2 = \log\,y_{k-j}^2 - \log\,\sigma_{k-j}^2$, the right hand side of (5) is a linear combination of the lagged values of $\log\,\sigma_k^2$'s and $\log\,y_k^2$'s. We refer to this model as multiplicative GARCH (p,q).

Since the GARCH(1,1) specification has been found to be adequate in most applications, this paper will confine the analysis to the model with the GARCH(1,1) specification and the common financial parameterization of the drift

$$\mu_k = c_0 + c_1\,\sigma_k^2, \tag{6}$$

and leave the generalization to general GARCH(p,q) specifications and other forms of $\mu_k$ to readers.

For a GARCH model, by a conditional argument we can easily derive its likelihood function

$$\prod_{k=1}^{n} \left[\sigma_k^{-1}\,\phi\left(\{x_k - \mu_k\}/\sigma_k\right)\right], \tag{7}$$

where $\phi$ is the density of the standard normal distribution. Because of the relatively simple likelihood function, statistical inference for the GARCH model can be carried out. (See Bollerslev, Chou and Kroner 1992, Gouriéroux 1997.)

## 2.2 Diffusion processes

In contrast to stochastic difference equations used in discrete-time models, stochastic differential equations are widely used to describe continuous-time models in the theoretical finance literature. The stochastic processes characterized by the stochastic differential equations are continuous-time diffusions (also referred to as Itô processes), and continuous-time modeling has made extensive use of Itô stochastic calculus, which provides an elegant means to analyze the diffusions. Specifically, a continuous-time model assumes that a security price $S_t$ obeys the following stochastic differential equation

$$S_t^{-1}\, dS_t = \nu_t\, dt + \sigma_t\, dW_t, \qquad t \in [0, T], \tag{8}$$

where $W_t$ is a standard Wiener process, $\nu_t$ is called (diffusion) drift in probability or instantaneous mean rate of return in finance, and $\sigma_t^2$ is called diffusion variance in probability or instantaneous conditional variance (or volatility) in finance. The celebrated Black-Scholes model corresponds to (8) with constants $\nu_t$ and $\sigma_t$ (Black and Sholes 1973, Merton 1973). However, many econometric studies have documented that financial time series tend to be highly heteroskedastic. To accommodate this, we allow $\sigma_t^2$ to be random and often assume $\sigma_t^2$ itself is governed by another stochastic differential equation [see (15) and (17) in section 2.3 below]. Such $\sigma_t^2$ is called a stochastic volatility and the corresponding model is termed a continuous-time stochastic volatility model. For the continuous-time model, the "no arbitrage" condition (the fundamental concept in finance, which is often labeled in plain English as the "no free lunch" condition) can be beautifully characterized by a martingale measure (or risk-neutral measure in finance), i.e. a probability law under which $S_t$ is a martingale. Prices of options and derivatives are then the conditional expectation of certain functionals of S under this measure. [A derivative is a financial instrument whose value depends on the values of other, more basic underlying variables such as stocks, interest rates, and currency exchange rates. A call (or put) option is a special derivative which gives its holder a right not obligation to buy (or sell) a security at certain price by future time. See Hull (1997).] The calculations and derivations can be manipulated by tools such as the Itô lemma and the Girsanov theorem. (See Dothan 1990, Duffie 1992, Harrison and Kreps 1979, Harrison and Pliska 1981, Hull and White 1987, Ikeda and Watanabe 1989, Karatzas and Shreve 1997, Merton 1990, Stroock and Varadhan 1979.)

The log price process $X_t = log\, S_t$ is often used. By Itô lemma and from (8) we obtain the diffusion model for $X_t$,

$$dX_t = (\nu_t + \sigma_t^2/2)\, dt + \sigma_t\, dW_t,$$

where the drift for $X_t$ has a term $\sigma_t^2$. GARCH models are used to statistically model the increments of the log price process, so from the diffusion point of view, (6) is also a natural parameterization of the GARCH drift $\mu_k$.

Since the likelihood for discretely sampled observations from a diffusion defined by nonlinear stochastic differential equations is not available, the statistical inference for the diffusion model is usually much harder than that for a GARCH model. Recently some inference methods have been developed for the diffusion model (Aït-Sahalia 1996, Gallant, Hsieh and Tauchen 1997, Gallant and Long 1997, Gallant and Tauchen 1998).

## 2.3 Diffusion limits of GARCH models

Divide the time interval $[0, T]$ into $n$ subinterval of length $s_n = T/n$ and set $t_k = k\, s_n$, $k = 0, 1, \cdots, n$. For i.i.d. standard normal random variables $\{\varepsilon_k\}$, let

$$\xi_k = \kappa_1 (\log \varepsilon_k^2 - \kappa_0), \qquad \zeta_k = 2^{-1/2} (\varepsilon_k^2 - 1), \tag{9}$$

where the generic constants $\kappa_0$ and $\kappa_1$ are

$$\kappa_0 = E log\, \varepsilon_1^2 \approx -1.27, \qquad \kappa_1 = \{var(log\, \varepsilon_1^2)\}^{-1/2} \approx 0.45. \tag{10}$$

The multiplicative GARCH(1,1) approximating process is defined as follows. For $k = 1, \cdots, n$, let

$$X_{n,k} - X_{n,k-1} = (\gamma_0 + \gamma_1\, \sigma_{n,k}^2)\, s_n + \sigma_{n,k}\, s_n^{1/2}\, \varepsilon_k, \tag{11}$$

$$log\, \sigma_{n,k}^2 = \beta_0\, s_n + (1 + \beta_1\, s_n)\, log\, \sigma_{n,k-1}^2 + \beta_2\, s_n^{1/2}\, \xi_{k-1}. \tag{12}$$

The approximating process $(X_{n,t}, \sigma_{n,t}^2)$, $t \in [0, T]$, is given by

$$X_{n,t} = X_{n,k}, \qquad \sigma_{n,t}^2 = \sigma_{n,k}^2, \qquad \text{for } t \in [t_k, t_{k+1}), \;\; k = 0, \cdots, n. \tag{13}$$

Nelson (1990) showed that as $n \to \infty$, the normalized partial sum process of $(\varepsilon_k, \xi_k)$ weakly converges to a planar Wiener process and thus the process $(X_{n,t}, \sigma_{n,t}^2)$ converges in distribution to bivariate diffusion process $(X_t, \sigma_t^2)$ governed by the following stochastic differential equation system

$$dX_t = (\gamma_0 + \gamma_1\, \sigma_t^2)\, dt + \sigma_t\, dW_{1,t}, \tag{14}$$

$$dlog\, \sigma_t^2 = (\beta_0 + \beta_1\, log\, \sigma_t^2)\, dt + \beta_2\, dW_{2,t}, \qquad \sigma_0^2 = e^{\beta_3}, \tag{15}$$

where $W_{1,t}$ and $W_{2,t}$ are two independent standard Wiener processes. The diffusion model described by (14)-(15) [or the process $(X_t, \sigma_t^2)$] is referred to as the diffusion limit of the multiplicative GARCH model (11)-(12).

For the linear GARCH model, everything is the same except for replacing (12) and (15), respectively, by

$$\sigma_{n,k}^2 = \beta_0 \, s_n + \sigma_{n,k-1}^2 \, (1 + \beta_1 \, s_n + \beta_2 \, s_n^{1/2} \, \zeta_{k-1}), \tag{16}$$

$$d\sigma_t^2 = (\beta_0 + \beta_1 \, \sigma_t^2) \, dt + \beta_2 \, \sigma_t^2 \, dW_{3,t}, \qquad \sigma_0^2 = e^{\beta_3}, \tag{17}$$

where $W_{3,t}$ is a standard Wiener process, $corr(W_{2,t}, W_{3,t}) = corr(\varepsilon_1^2, log\, \varepsilon_1^2) \approx 0.64$, and $W_{1,t}$ is independent of $W_{2,t}$ and $W_{3,t}$. (See Duan 1995, Nelson 1990, Rossi 1996.)

For simplicity, throughout this paper we assume initial values $X_{n,0} = X_0$ and $\sigma_{n,0}^2 = \sigma_0^2 = e^{\beta_3}$.

Note that the diffusion model is for the log price process and the GARCH model is for the increments of the log price process. The increments of the GARCH approximating processes defined by (11)-(9) and (11)-(16) obey the GARCH structure of section 2.1. Indeed, in equation (11), the term $(\gamma_0 + \gamma_1 \, \sigma_{n,k}^2) \, s_n$ is the drift $\mu_k$ [see (1)] of the form specified in (6) with $\gamma_0$ and $\gamma_1$ the rescaled versions of $c_0$ and $c_1$, namely, $c_0 = \gamma_0 \, s_n$, $c_1 = \gamma_1 \, s_n$. Using the relationship between $(\xi_k, \zeta_k)$ and $\varepsilon_k^2$ given by (9) and (10), we rewrite equations (12) and (16), respectively, as

$$log\, \sigma_{n,k}^2 = \alpha_0 + \alpha_1 \, log\, \sigma_{n,k-1}^2 + \alpha_2 \, log\, \varepsilon_{k-1}^2, \tag{18}$$

with

$$\alpha_0 = \beta_0 \, s_n - \beta_2 \, s_n^{1/2} \, \kappa_0 \, \kappa_1, \qquad \alpha_1 = 1 + \beta_1 \, s_n, \qquad \alpha_2 = \beta_2 \, s_n^{1/2} \, \kappa_1;$$

and

$$\sigma_{n,k}^2 = \alpha_0 + \alpha_1 \, \sigma_{n,k-1}^2 + \alpha_2 \, \sigma_{n,k-1}^2 \, \varepsilon_{k-1}^2, \tag{19}$$

with

$$\alpha_0 = \beta_0 \, s_n, \qquad \alpha_1 = 1 + \beta_1 \, s_n - \beta_2 \, s_n^{1/2}/2^{1/2}, \qquad \alpha_2 = \beta_2 \, s_n^{1/2}/2^{1/2}.$$

Comparing (18) and (19) with GARCH specifications (4) and (5), we clearly demonstrate that (11)-(12) and (11)-(16) are multiplicative and linear GARCH(1,1), respectively. The parameters $\gamma_i$'s and $\beta_i$'s are, respectively, the rescaled versions of the drift parameters $c_i$'s [in (6)] and local reparameterization of the volatility parameters $\alpha_i$'s [in (4) and (5)] so that the diffusion limits can be obtained. As option pricing depends on security prices, these weak convergence results can be used to show that option pricing formulas for GARCH models agree in the limit with those for their diffusion limits. Also, the results are very useful for the cases where one may find the distributional results are available for continuous-time models that are not available for the GARCH models. (See Duan 1995, Nelson 1990, Rossi 1996.)

# 3 Statistical equivalence and nonequivalence

## 3.1 Comparison of experiments

A statistical problem $\mathbb{E}$ 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, $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta, \theta \in \Theta))$. $\mathbb{E}$ is referred to as a statistical experiment. Le Cam's deficiency distance is often used to compare statistical experiments.

Consider two statistical experiments with the same parameter space $\Theta$, $\mathbb{E}_i = (\Omega_i, \mathcal{F}_i, (P_{i,\theta}, \theta \in \Theta))$, $i = 1, 2$. 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 $i$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 $\Delta(\mathbb{E}_1, \mathbb{E}_2)$ between $\mathbb{E}_1$ and $\mathbb{E}_2$ is the maximum of $\delta(\mathbb{E}_1, \mathbb{E}_2)$ and $\delta(\mathbb{E}_2, \mathbb{E}_1)$, where

$$\delta(\mathbb{E}_1, \mathbb{E}_2) = \inf_{\delta_1} \sup_{\delta_2} \sup_{\theta \in \Theta} \sup_{L:\|L\|=1} |R_1(\delta_1, L, \theta) - R_2(\delta_2, L, \theta)|$$

is called the deficiency of $\mathbb{E}_1$ with respect to $\mathbb{E}_2$. Two experiments $\mathbb{E}_1$ and $\mathbb{E}_2$ are called equivalent if $\Delta(\mathbb{E}_1, \mathbb{E}_2) = 0$. Equivalence means that each procedure $\delta_1$ in problem $\mathbb{E}_1$ has a corresponding procedure $\delta_2$ 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, any sequence of procedures $\delta_{n,1}$ in problem $\mathbb{E}_{n,1}$ has a corresponding sequence of procedures $\delta_{n,2}$ in problem $\mathbb{E}_{n,2}$ with risk differences tending to zero uniformly over $\theta \in \Theta$ and all $L$ with $\|L\| = 1$, The procedures $\delta_{n,1}$ and $\delta_{n,2}$ are said to be asymptotically equivalent. (See Le Cam 1986, Le Cam and Yang 1990, Strasser 1985.)

## 3.2 GARCH and Diffusion Experiments

Denote by $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)$ the vectors of variance (or volatility) parameters and drift parameters in the GARCH and diffusion models defined in section 2.3, respectively. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}) = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1)$ be the vector of all six parameters and the parameter space $\Theta$ consist of $\boldsymbol{\theta}$ with $\gamma_i$ and $\beta_i$ belonging to bounded intervals. Denote by $P_{n,\boldsymbol{\theta}}$ the distribution of the approximating process $X_{n,t_k}$, $k = 1, \cdots, n$, defined by the stochastic difference equations (11)-(12) for the multiplicative GARCH model [or (11) and (16) for the linear GARCH model], and denote by $Q_{n,\boldsymbol{\theta}}$ the distribution of the discrete samples at $t_k$, $k = 1, \cdots, n$, of the diffusion limit $X_t$ governed by the stochastic differential equations (14)-(15) for the multiplicative GARCH case [or (14) and (17) for the linear GARCH case, respectively]. Define the GARCH and diffusion experiments, respectively, by

$$\mathbb{E}_{n,1} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{n,\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta)\right), \qquad \mathbb{E}_{n,2} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (Q_{n,\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta)\right). \tag{20}$$

We like to emphasize that as in the standard financial setting, volatility processes $\sigma^2$'s are latent and processes $X$'s are observable, and samples in both GARCH and diffusion experiments are taken from processes $X$'s only. Also as pointed out in Section 1, paragraph 4, the infinite-dimensional distributions of continuous-time observations, $\{X_t, t \in [0,T]\}$, from the diffusion limit under different values for the variance parameter $\boldsymbol{\beta}$ are mutually singular, and the volatility process $\sigma_t^2$ can be exactly recoverered by the quadratic variation process of $\{X_t, t \in [0,T]\}$. For these reasons, the diffusion experiment $\mathbb{E}_{n,2}$ considers discretely sampled observations, with $Q_{n,\boldsymbol{\theta}}$ for the finite-dimensional distribution of the discrete samples.

We have the following theorems whose proofs are given in sections 4-6.

**Theorem 1** *The experiments $\mathbb{E}_{n,1}$ and $\mathbb{E}_{n,2}$ are not asymptotically equivalent.*

As a comparison, we define corresponding experiments with nonstochastic volatility which corresponds to $\beta_2 = 0$. Denote by $\Theta'$ the subset of $\Theta$ consisting all $\boldsymbol{\theta}$ with $\beta_2 = 0$. Treating $\Theta'$ as the parameter space for $(\beta_0, \beta_1, \beta_3, \gamma_0, \gamma_1)$ in five dimensions, we define

$$\mathbb{E}'_{n,1} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{n,\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta')\right), \qquad \mathbb{E}'_{n,2} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (Q_{n,\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta')\right).$$

**Theorem 2** *The experiments $\mathbb{E}'_{n,1}$ and $\mathbb{E}'_{n,2}$ are asymptotically equivalent.*

As we explained in section 1, paragraph 5, Theorem 1 is due to the different noise propagation systems that the GARCH model and its diffusion limit utilize in their conditional variances. The following heuristic comparison of two simple models offers genuine insight into the phenomenon.

For each GARCH model we discretize its diffusion limit and obtain a corresponding discrete model, which is called a discrete stochastic volatility model in financial econometrics. The multiplicative GARCH(1,1) model

$$y_k = \mu + \sigma_k \varepsilon_k, \qquad log\, \sigma_k^2 = \alpha_0 + \alpha_1\, log\, \sigma_{k-1}^2 + \alpha_2\, log\, \varepsilon_{k-1}^2, \qquad k = 1, \cdots, n \tag{21}$$

has the following discrete counterpart:

$$y_k = \mu + \sigma_k z_k, \qquad log\, \sigma_k^2 = \alpha_0 + \alpha_1\, log\, \sigma_{k-1}^2 + \alpha_2\, \delta_k, \qquad k = 1, \cdots, n, \tag{22}$$

where $\varepsilon_k$, $z_k$ and $\delta_k$ are independent standard normal random variables. Obviously model (22) has the same diffusion limit as model (21), but unlike the GARCH model, its likelihood also mimics that of the diffusion limit.

For model (21), the distribution of $(y_k, \sigma_k^2)_{k\geq 1}$ is determined by $(\varepsilon_k, log\, \varepsilon_k^2)_{k\geq 1}$, while its likelihood function given by (7) depends on $(\varepsilon_k, log\, \varepsilon_k^2, \varepsilon_k^2)_{k\geq 1}$. The corresponding distribution and likelihood for model (22) are obtained, respectively, by first replacing $(\varepsilon_k, log\, \varepsilon_k^2)_{k\geq 1}$ with $(z_k, \delta_k)_{k\geq 1}$ in the GARCH distribution and substituting $(\varepsilon_k, log\, \varepsilon_k^2, \varepsilon_k^2)_{k\geq 1}$ with $(z_k, \delta_k, z_k^2)_{k\geq 1}$ in the GARCH likelihood, and then taking the conditional expectation with respect to $\delta_k$'s.

As $\varepsilon_k^2$ and $log\, \varepsilon_k^2$ are correlated but both are uncorrelated with $\varepsilon_k$, the normalized partial sum processes for $(\varepsilon_k, log\, \varepsilon_k^2, \varepsilon_k^2)$ converge in distribution to three dimensional Brownian motion $(W_{1,t}, W_{2,t}, W_{3,t})$, where $W_{2,t}$ and $W_{3,t}$ are correlated but both are independent of $W_{1,t}$. With the independence between $z_k$ and $\delta_k$, the weak limit of the normalized partial sum processes for $(z_k,\, \delta_k,\, z_k^2)$ is three independent Brownian motions $(W_{1,t}, W_{2,t}, W_{4,t})$. Thus for both models, the asymptotic distributions of the partial sum processes for $(y_k, \sigma_k^2)$ are determined by the same stochastic differential equation system governed by two independent Brownian motions $(W_{1,t}, W_{2,t})$, but the asymptotic likelihoods for models (21) and (22) are related in a similar fashion to three correlated Brownian motions $(W_{1,t}, W_{2,t}, W_{3,t})$ and three independent Brownian motions $(W_{1,t}, W_{2,t}, W_{4,t})$, respectively. As a result, under stochastic volatility, the two models have the same asymptotic distributions but different asymptotic likelihoods, and consequently they are not asymptotically equivalent (Le Cam and Yang 1990, section 2.2 of chapter 2 or Le Cam 1986, Proposition 8 and its remark in section 4 of chapter 6, pp. 93-95). With nonstochastic volatility, the deterministic conditional variances don't depend on $log\, \varepsilon_k^2$ or $\delta_k$, and the two likelihood processes asymptotically depend on only either $(W_{1,t}, W_{3,t})$ or $(W_{1,t}, W_{4,t})$. As $(W_{1,t}, W_{3,t})$ and $(W_{1,t}, W_{4,t})$ are identically distributed, asymptotically the two models have the same statistical behavior and thus are equivalent.

Empirical work suggests great differences in statistical inference between the GARCH and discrete stochastic volatility models. Hsieh (1991) and Jacquier, Polson and Rossi (1994) reported substantial differences in likelihood, parameter estimation, forecasting and auto-correlation when fitting the two models to simulated data and real financial data. For each model their simulations indicate that the likelihood based methods can very accurately estimate parameters (with root mean squared error ranging from 0.02 to 0.05 for sample size 2000). Here we conducted a simulation to evaluate the numerical performance of the MLE derived under the linear GARCH model when being fed with data coming from its corresponding discrete stochastic volatility model. In the simulation we took $n = 2000$, $\mu = 0$, $\alpha_0 = 0.5$, $\alpha_1 = 0.3$ and $\alpha_2 = 0.6$. A sample from each model was simulated. Using S-plus GARCH module we calculated the MLE and its estimated asymptotic standard error (a.se) by fitting the GARCH model to the sample generated from the GARCH model and then repeated the calculation by replacing the GARCH sample with data coming from the volatility model. The whole procedure was repeated 100 times. From the 100 repetitions, we calculated the average estimated values and a.se, and their standard errors (se). Results for $\alpha_1$ and $\alpha_2$ are listed in Table 1.

Table 1. The average values of the GARCH MLE of $(\alpha_1, \alpha_2)$ and its a.se, with their se (in parentheses) for GARCH and volatility data

| | GARCH sample | | Volatility sample | |
|---|---|---|---|---|
| Parameter | Estimate | a.se | Estimate | a.se |
| $\alpha_1$ | 0.2958 (0.0371) | 0.0348 (0.0037) | 0.5292 (0.1015) | 0.0281 (0.0093) |
| $\alpha_2$ | 0.5998 (0.0475) | 0.0476 (0.0028) | 0.2837 (0.0782) | 0.0200 (0.0049) |

The simulation results show dramatical differences between the GARCH and volatility samples. When fitting the GARCH model to the GARCH data, the MLE performs extremely well, with negligible bias, small se, and a.se close to the actual se. While for the volatility samples, the same MLE produces a very poor estimator, with huge bias, large se and a.se far off the target. To rule out the possibility that the bias is caused by the fact that the gradient search algorithm used by Splus in the MLE computation was stuck by local maxima, we have set the true parameter values as initial values and found little difference in the outcome values for the MLE. We have tried to reduce the huge bias by increasing sample size up to $10^5$ but failed to achieve even a small amount of reduction. Also we have tested various values for $\alpha_i$ and found out that for volatility samples, the MLE tends to overestimate $\alpha_1$ and underestimate $\alpha_2$, and the amounts over and under estimated tend to get smaller as $\alpha_2$ decreases. As $\alpha_2$ gets close to zero, the bias approaches to the level for GARCH data. The numerical findings are very much in agreement with the theoretical results described in Theorems 1 and 2 and reinforce the point in paragraph 6 of Section 1 that the nonequivalence result has an important consequence for the statistical inference of the GARCH and diffusion models.

Our approach to proving the theorems is based on the following principle described in Le Cam (1986), Le Cam and Yang (1990) and Nussbaum (1996). For two experiments $\mathbb{E}_i = (\Omega_i, \mathcal{F}_i, (P_{i,\theta}, \theta \in \Theta))$, $i = 1, 2$, assume there is some $\theta^* \in \Theta$ such that all the $P_{i,\theta}$ are dominated by $P_{i,\theta^*}$, $i = 1, 2$, and form

$$\Lambda_i(\theta) = \frac{dP_{i,\theta}}{dP_{i,\theta^*}}.$$

Treating $\Lambda_i = (\Lambda_i(\theta), \theta \in \Theta)$ as stochastic processes indexed by $\theta$ given on the probability space $(\Omega_i, \mathcal{F}_i, P_{i,\theta^*})$, we call them the likelihood processes of the experiments $\mathbb{E}_i$. If there are versions, $\Lambda_i^*$, of $\Lambda_i$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then on the common probability space $\Lambda_i^*$ generate equivalent versions of the experiments, and

$$\Delta(\mathbb{E}_1, \mathbb{E}_2) \leq \sup_{\theta \in \Theta} E_{\mathbb{P}} |\Lambda_1^* - \Lambda_2^*| \leq 2 \sup_{\theta \in \Theta} H(\Lambda_1^*(\theta), \Lambda_2^*(\theta)), \tag{23}$$

where $H(\Lambda_1^*(\theta), \Lambda_2^*(\theta)) = E_{\mathbb{P}} \left( [\Lambda_1^*(\theta)]^{1/2} - [\Lambda_2^*(\theta)]^{1/2} \right)^2$ is Hellinger distance. Hellinger distance can easily handle normal distributions and distributions of product forms.

Specifically, sections 4 and 5 derive likelihood processes for the GARCH model and its diffusion limit and study their asymptotic distributions in a local neighborhood. Section 6 proves nonequivalence under stochastic volatility by showing different limiting distributions for the two likelihood processes and equivalence under nonstochastic volatility by proving the convergence of Hellinger distance to zero.

# 4 Likelihood processes for GARCH models

## 4.1 Notations and conventions

To track complex processes under different circumstances and manage long technical arguments, we fix the following notations and conventions.

**Convention 1** It is often necessary to put processes and random variables on some common probability spaces. At such occasions, we often automatically change probability spaces and consider versions of the processes and the random variables on new probability spaces, without altering notations. Because of this convention and Skorohod's theorem, we often switch between "convergence in probability" and "convergence in distribution". Also because of the convention, when no confusion occurs, we try to use the same notation for random variables or processes with identical distribution. For the sake of simplicity, we take $T = 1$ and $s_n = 1/n$. All $O$'s and $o$'s hold uniformly over $t \in [0, 1]$.

**Notation 2**. For a fixed $\boldsymbol{\beta^*}$, let $\boldsymbol{\theta}^* = (\boldsymbol{\beta^*, 0})$ and define a local neighborhood around $\boldsymbol{\theta^*}$

$$\Theta_{n,c}(\boldsymbol{\beta}^*) = \left\{ \boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}) \in \Theta : \boldsymbol{\beta} = \boldsymbol{\beta}^* + n^{-1/2}\, \boldsymbol{\varphi}, |\boldsymbol{\varphi}| \le c \right\} \subset \Theta, \tag{24}$$

and the corresponding local experiments

$$\mathbb{E}_{n,1}(\boldsymbol{\beta}^*) = \left( \mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{n,\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_{n,c}) \right), \quad \mathbb{E}_{n,2}(\boldsymbol{\beta}^*) = \left( \mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (Q_{n,\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_{n,c}) \right). \tag{25}$$

Also introduce notation $\boldsymbol{\theta}^+ = (\boldsymbol{\beta}^* + n^{-1/2}\, \boldsymbol{\varphi}, \mathbf{0})$ to denote parameter $\boldsymbol{\theta} \in \Theta_{n,c}(\boldsymbol{\beta}^*)$ with drift component $\boldsymbol{\gamma} = \mathbf{0}$.

The shrinking in $\Theta_{n,c}$ is only for $\boldsymbol{\beta}$, because $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are the drift and variance parameters, respectively. As discussed in paragraph 4 of Section 1, for drift, the likelihood processes have non-degenerate limiting distributions over all $\boldsymbol{\gamma}$; while for variance, we need to localize $\boldsymbol{\beta}$ in a $n^{-1/2}$-shrinking neighborhood and derive non-degenerate limiting distributions for the likelihood processes.

**Convention 3**. For $\boldsymbol{\theta} = (\boldsymbol{\beta}^* + n^{-1/2}\,\boldsymbol{\varphi}, \boldsymbol{\gamma}) \in \Theta_{n,c}(\boldsymbol{\beta}^*)$, define $\boldsymbol{\vartheta} = (\boldsymbol{\varphi},\, \boldsymbol{\gamma})$. Then there is a one-to-one correspondence between $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$. When no confusion occurs, for convenience we use $\boldsymbol{\vartheta}$ to index $\Theta_{n,c}(\boldsymbol{\beta}^*)$ and write $\Theta_{n,c}(\boldsymbol{\beta}^*)$ as $\Theta_{n,c}$ for short. For example, under this convention, $\boldsymbol{\vartheta}^* = (\mathbf{0}, \mathbf{0})$ and $\boldsymbol{\vartheta}^+ = (\boldsymbol{\varphi}, \mathbf{0})$ correspond to $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^+$, respectively.

**Convention 4**. Our notations must keep track of three distinct kinds of processes defined in section 2.3 under two circumstances.

(a) The sequence of the discrete-time processes $X_{n,k}$ and $\sigma^2_{n,k}$ that depend both on $n$ and on the discrete time index $k$, $k = 0, \cdots, n$,

(b) the sequence of the continuous-time processes $X_{n,t}$ and $\sigma^2_{n,t}$ formed as random step functions from the discrete time processes in (a) using (13), and

(c) the limiting diffusion process $(X_t,\, \sigma^2_t)$ with initial values $X_{n,0} = X_0$ and $\sigma^2_{n,0} = \sigma^2_0$, and as $n \to \infty$, $(X_{n,t}, \sigma^2_{n,t}) \Longrightarrow (X_t, \sigma^2_t)$.

Each kind of the processes are under two circumstances that the parameter $\boldsymbol{\vartheta} = (\boldsymbol{\varphi}, \boldsymbol{\gamma})$ [or $\boldsymbol{\theta} = (\boldsymbol{\beta}^* + n^{-1/2}\,\boldsymbol{\varphi}, \boldsymbol{\gamma})$] and $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^* \equiv (\mathbf{0}, \mathbf{0})$ [or $\boldsymbol{\theta} = \boldsymbol{\theta}^* \equiv (\boldsymbol{\beta}^*, \mathbf{0})$]. To distinguish the latter from the former, we add a subscript "0" to the processes to denote that they are under the condition $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^*$. For example, $\sigma^2_{n,t}$ and $\sigma^2_{n,t,0}$ denote the GARCH volatility process under the conditions that $\boldsymbol{\vartheta} = (\boldsymbol{\varphi}, \boldsymbol{\gamma})$ and $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^*$, respectively; and $\sigma^2_{t,0}$ denotes the diffusion volatility process with $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^*$.

**Notation 5**. Define

$$V_{n,t} = n^{1/2}\left(1 - \frac{\sigma^2_{n,t,0}}{\sigma^2_{n,t}}\right), \qquad H_{n,t} = n\left(1 + log\,\frac{\sigma^2_{n,t,0}}{\sigma^2_{n,t}} - \frac{\sigma^2_{n,t,0}}{\sigma^2_{n,t}}\right), \tag{26}$$

$$V_t = \sum_{i=0}^{3} \varphi_i \frac{\partial log\, \sigma^2_{t,0}}{\partial \beta^*_i}. \tag{27}$$

All processes depend on parameter $\boldsymbol{\vartheta}$ and the dependence is not often explicitly given in process notations. For example, we may write $V_t$ as $V_t(\boldsymbol{\vartheta})$ to mark the dependence clearly.

**Notation 6**. Define partial sum processes

$$W^{(n)}_{1,t} = n^{-1/2}\sum_{j=1}^{[n\,t]} \varepsilon_j, \qquad W^{(n)}_{2,t} = n^{-1/2}\sum_{j=1}^{[n\,t]} \xi_j, \qquad W^{(n)}_{3,t} = n^{-1/2}\sum_{j=1}^{[n\,t]} \zeta_j, \tag{28}$$

where $\varepsilon_j$ are standard normal random errors, and $\xi_j$ and $\zeta_j$ are defined in (9). From the modulus of continuity of Wiener process (Karatzas and Shreve 1997, chapter 2) and strong approximation (Komlós, Major and Tusnády 1975, Tusnády 1977) we have that on some probability spaces there exist three standard Wiener processes $W_{1,t}, W_{2,t}, W_{3,t}$ with $W_{1,t}$ independent of $W_{2,t}$ and $W_{3,t}$, and $corr(W_{2,t}, W_{3,t}) = corr(\varepsilon_1^2, log\, \varepsilon_1^2) \approx 0.64$ (see also section 2.3), such that

$$\sup_{0\le t\le 1} \left\{ |W_{1,t}^{(n)} - W_{1,t}| + |W_{2,t}^{(n)} - W_{2,t}| + |W_{3,t}^{(n)} - W_{3,t}| \right\} = O_p\left( n^{-1/2} log^2\, n \right). \tag{29}$$

## 4.2 Asymptotics of likelihood processes

Denote by $L_{n,1}(\boldsymbol{\vartheta})$ the likelihood function of the GARCH approximating process $X_{n,t_k}$ and let $\Lambda_{n,1}(\boldsymbol{\vartheta}) = L_{n,1}(\boldsymbol{\vartheta})/L_{n,1}(\boldsymbol{\vartheta}^*)$ be its likelihood process under $P_{n,\boldsymbol{\vartheta}^*}$. We will show below that $\Lambda_{n,1}(\boldsymbol{\vartheta})$ has a limit $\Lambda_1(\boldsymbol{\vartheta})$ defined by

$$\begin{aligned} log\, \Lambda_1(\boldsymbol{\vartheta}) &= \frac{1}{\sqrt{2}} \int_0^1 V_t\, dW_{3,t} - \frac{1}{4}\int_0^1 V_t^2\, dt + \int_0^1 \sigma_{t,0}^{-1} \left(\gamma_0 + \gamma_1\, \sigma_{t,0}^2\right)\, dW_{1,t} \\ &\quad - \frac{1}{2}\int_0^1 \sigma_{t,0}^{-2} \left(\gamma_0 + \gamma_1\, \sigma_{t,0}^2\right)^2\, dt. \end{aligned} \tag{30}$$

From (6), (7), (11), (12), (13) and (16), we obtain that

$$\begin{aligned} \log L_{n,1}(\boldsymbol{\vartheta}) &= -\sum_{j=1}^n \left( X_{n,t_j} - X_{n,t_{j-1}} - s_n\, (\gamma_0 + \gamma_1\, \sigma_{n,t_j}^2) \right)^2\, (2\, s_n\, \sigma_{n,t_j}^2)^{-1} \\ &\quad - \sum_{j=1}^n log\, \sigma_{n,t_j} - (n/2)\, log(2\, \pi\, s_n), \end{aligned}$$

and thus

$$\begin{aligned} log\, \Lambda_{n,1}(\boldsymbol{\vartheta}) &= \sum_{j=1}^n (log\, \sigma_{n,t_j,0} - log\, \sigma_{n,t_j}) + (2\, s_n)^{-1} \sum_{j=1}^n \sigma_{n,t_j,0}^{-2}\, (X_{n,t_j,0} - X_{n,t_{j-1},0})^2 \\ &\quad - (2\, s_n)^{-1} \sum_{j=1}^n \sigma_{n,t_j}^{-2}\, \left\{ X_{n,t_j,0} - X_{n,t_{j-1},0} - s_n\, (\gamma_0 + \gamma_1\, \sigma_{n,t_j}^2) \right\}^2 \\ &= \frac{1}{2}\sum_{j=1}^n \left( 1 - \frac{\sigma_{n,t_j,0}^2}{\sigma_{n,t_j}^2} \right)\, (\varepsilon_j^2 - 1) + \frac{1}{2}\sum_{j=1}^n \left( 1 + log \frac{\sigma_{n,t_j,0}^2}{\sigma_{n,t_j}^2} - \frac{\sigma_{n,t_j,0}^2}{\sigma_{n,t_j}^2} \right) \\ &\quad + s_n^{1/2} \sum_{j=1}^n \left( \gamma_0 + \gamma_1\, \sigma_{n,t_j}^2 \right)\, \sigma_{n,t_j,0}\, \sigma_{n,t_j}^{-2}\, \varepsilon_j - \frac{s_n}{2} \sum_{j=1}^n \left( \gamma_0 + \gamma_1\, \sigma_{n,t_j}^2 \right)^2\, \sigma_{n,t_j}^{-2} \\ &= \frac{1}{\sqrt{2}} \int_0^1 V_{n,t}\, dW_{3,t}^{(n)} + \frac{1}{2}\int_0^1 H_{n,t}\, dt + \int_0^1 \left( \gamma_0 + \gamma_1\, \sigma_{n,t}^2 \right)\, \sigma_{n,t,0}\, \sigma_{n,t}^{-2}\, dW_{1,t}^{(n)} \\ &\quad - \frac{1}{2}\int_0^1 \left( \gamma_0 + \gamma_1\, \sigma_{n,t}^2 \right)^2\, \sigma_{n,t}^{-2}\, dt, \end{aligned} \tag{31}$$

where the second equality is from (11), and the third equality is due to the definitions of $V_{n,t}$ and $H_{n,t}$ in (26), and the piecewiseness of $\sigma^2_{n,t}$ and $W^{(n)}_{i,t}$ defined in (13) and (28), respectively.

**Proposition 1**

$$V_{n,t} = V_t + O_p(n^{-1/2}\, log^2\, n).$$

*where $V_{n,t}$ and $V_t$ are defined in (26) and (27), respectively.*

Proposition 1 will be proved in sections 4.3 and 4.4 where Lemmas 1 and 4 indicate that $V_t$ smoothly depends on $W_{2,t}$ or $W_{3,t}$.

**Proposition 2**

$$H_{n,t} = -V_t^2/2 + O_p(n^{-1/2}\, log^2\, n),$$

*where $H_{n,t}$ and $V_t$ are defined in (26) and (27), respectively.*

Proof. Note that $H_{n,t} = n\left\{n^{-1/2}\, V_{n,t} + log\left(1 - n^{-1/2}\, V_{n,t}\right)\right\}$ and as $x \to 0$, $x + log\,(1-x) = -x^2/2 + O(x^3)$. In view of Proposition 1, we have that

$$H_{n,t} \;=\; n\left\{-(n^{-1/2}\, V_{n,t})^2/2 + O_p(n^{-3/2})\right\} = -V_t^2/2 + O_p(n^{-1/2}\, log^2\, n). \qquad \square$$

**Proposition 3**

$$\Lambda_{n,1}(\boldsymbol{\vartheta}) = \Lambda_1(\boldsymbol{\vartheta}) + O_p(n^{-1/2}\, log^2\, n),$$

*where $\Lambda_1(\boldsymbol{\vartheta})$ is defined in (30).*

Proof. Comparing Lemma 3 (or Lemma 6) with Lemma 1 (or Lemma 4, respectively,) [in sections 4.3 and 4.4 below] and using the strong approximation (29) we easily conclude that both $\sigma^2_{n,t}$ and $\sigma^2_{n,t,0}$ converge in probability to $\sigma^2_{t,0}$ with error rate $n^{-1/2}\, log^2\, n$. Then by (26), (27), the strong approximation (29), and Propositions 1 and 2, we can show that the integrals in $\Lambda_{n,1}(\boldsymbol{\vartheta})$ given by (31) converge in probability to the corresponding integrals in $\Lambda_1(\boldsymbol{\vartheta})$ defined in (30) with errors of order $n^{-1/2}\, log^2\, n$. This completes the proof. $\square$

**Proposition 4** *Assume that $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^+ \equiv (\boldsymbol{\varphi}, \mathbf{0})$ (i.e. $\boldsymbol{\gamma} = 0$) and Novikov's condition [i.e. $E exp\left\{\frac{1}{4}\int_0^1 V_t^2\, dt\right\} < \infty$ ] is satisfied. Then as $n \to \infty$, $E|\Lambda_{n,1}(\boldsymbol{\vartheta}^+) - \Lambda_1(\boldsymbol{\vartheta}^+)| \to 0$.*

Proof. Since $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^+$ means that the drift parameter $\boldsymbol{\gamma} = 0$, substituting $\boldsymbol{\gamma}$ by zero in $\Lambda_1(\boldsymbol{\vartheta})$ defined in (30), we obtain that Novikov's condition ensures $E\Lambda_1(\boldsymbol{\vartheta}^+) = 1$ and thus $E[\Lambda_1(\boldsymbol{\vartheta}^+)|W_{2,u}, W_{3,u}, u \le t] = \exp\left\{\frac{1}{\sqrt{2}}\int_0^t V_s\, d\, W_{3,s} - \frac{1}{4}\int_0^t V_s^2\, d\, s\right\}$ is a martingale (Ikeda and Watanabe 1989, Karatzas and Shreve 1997, section 3.5). As $1 - \sigma^2_{n,j,0}/\sigma^2_{n,j} < 1$, it is easy to check that as a likelihood process, $E\Lambda_{n,1}(\boldsymbol{\vartheta}^+) = 1$. Now the proposition is a consequence of Proposition 3 and Scheffé theorem. $\square$

**Remark 4.1**. Since $V_t^2$ is of order $W^2_{2,t}$ or $W^2_{3,t}$, we can show that there exists constant $\delta > 0$ depending only on time interval $[0, 1]$ such that Novikov's condition holds for all $\boldsymbol{\varphi}$ with $\varphi_2^2 \le \delta$. For example, for the multiplicative GARCH case with $\beta_0^* = \beta_1^* = 0$, $V_t = \varphi_2\, W_{2,t}$. the Karhunen-Loève expansion of $W_{2,t}$, $t \in [0, 1]$, is given by

$$W_{2,t} = \sum_{j=0}^{\infty} 2^{1/2}\, \pi^{-1}\, (j + 1/2)^{-1}\, sin\, \{\pi\, (j + 1/2)\, t/T\}\; z_j,$$

where $z_j$ are i.i.d. standard normal random variables. Then we have $\int_0^1 W^2_{2,t}\, dt = \sum_{j=0}^{\infty} \pi^{-2}$ $(j + 1/2)^{-2}\, z_j^2$, and for $\varphi_2^2 < \delta = \pi^2/4$,

$$Eexp\left(\frac{1}{4}\int_0^1 V_t^2\, dt\right) = \prod_{j=0}^{\infty} E\, exp\left\{\pi^{-2}\, \varphi_2^2\, (2\, j + 1)^{-2}\, z_j^2\right\}$$

$$= \prod_{j=0}^{\infty} \left\{1 - 2\, \pi^{-2}\, \varphi_2^2\, (2\, j + 1)^{-2}\right\}^{-1/2} \sim exp\left\{\pi^{-2}\, \varphi_2^2\, \sum_{j=0}^{\infty} (2\, j + 1)^{-2}\right\} < \infty.$$

## 4.3 Proof of Proposition 1 for the multiplicative GARCH model

**Lemma 1** *The solution of (15) is given by*

$$log\, \sigma_t^2 = e^{\beta_1\, t} \left\{\beta_3 + \beta_2 \int_0^t e^{-\beta_1\, s}\, dW_{2,s} + \beta_0 \int_0^t e^{-\beta_1\, s}\, ds\right\}.$$

Proof. Applying Itô lemma (Ikeda and Watanabe 1989, Karatzas and Shreve 1997) to the process given by the lemma, we have

$$\begin{aligned} dlog\, \sigma_t^2 &= \beta_1\, e^{\beta_1\, t}\, dt\; e^{-\beta_1\, t}\, log\, \sigma_t^2 + e^{\beta_1\, t} \left\{\beta_2 e^{-\beta_1\, t}\, dW_{2,t} + \beta_0 e^{-\beta_1\, t}\, dt\right\} \\ &= (\beta_0 + \beta_1\, log\, \sigma_t^2)\, dt + \beta_2\, dW_{2,t}. \qquad \square \end{aligned}$$

**Lemma 2** *The process defined in (12) has the expression*

$$log\, \sigma^2_{n,k} = \alpha_1^{k-1}\, log\, \sigma_0^2 + (\alpha_2/\alpha_1) \sum_{j=1}^{k-1} \alpha_1^{k-j}\, \xi_j + (\alpha_0/\alpha_1) \sum_{j=1}^{k-1} \alpha_1^{k-j},$$

*with* $log\, \sigma_0^2 = \beta_3$, $\alpha_0 = s_n\, \beta_0$, $\alpha_1 = 1 + s_n\, \beta_1$, *and* $\alpha_2 = s_n^{1/2}\, \beta_2$.

Proof. The lemma is easily proved by applying (12) recursively. □

**Lemma 3**

$$log\,\sigma^2_{n,t} = e^{\beta_1 t}\left\{\beta_3 + \beta_2 \int_0^t e^{-\beta_1 s}\, dW^{(n)}_{2,s} + \beta_0 \int_0^t e^{-\beta_1 s}\, ds\right\} + O_p(n^{-1}).$$

**Remark 4.2**. As $V_{n,t}$ relates to $V_t$ through the $n^{-1/2}$ order term in the expansion of $log\,\sigma^2_{n,t} - log\,\sigma^2_{n,t,0}$, in order to prove Proposition 1 we need to keep the approximation to be of order higher than $n^{-1/2}$. Since the strong approximation has error of order lower than $n^{-1/2}$, we can't replace $W_2^{(n)}$ by $W_2$ at this point.

Proof of Lemma 3. First note that uniformly for $k = 1, \cdots, n$,

$$(1+\beta_1/n)^k = exp\{k\, log(1+\beta_1/n)\} = exp\{k\,[\beta_1/n + O(n^{-2})]\} = e^{\beta_1\, k/n} + O(n^{-1}).$$

Now Lemma 2 implies that

$$\begin{aligned} log\,\sigma^2_{n,k} &= (1+\beta_1/n)^{k-1}\,\beta_3 + \beta_2\,(1+\beta_1/n)^{-1}\sum_{j=1}^{k-1}(1+\beta_1/n)^{k-j}\,[W^{(n)}_{2,j/n} - W^{(n)}_{2,(j-1)/n}] \\ &\quad +\beta_0\,(1+\beta_1/n)^{-1}\sum_{j=1}^{k-1}(1+\beta_1/n)^{k-j}/n \\ &= e^{\beta_1\, k/n}\,\beta_3 + \beta_2 \int_0^{k/n} e^{\beta_1\,(k/n-s)}\, dW^{(n)}_{2,s} + \beta_0 \int_0^{k/n} e^{\beta_1\,(k/n-s)}\, ds + O_p(n^{-1}). \end{aligned}$$

The lemma is easily proved by combining above result with (13).

**Proof of Proposition 1**. Note from Convention 4 in section 4.1 that for $\sigma^2_{n,t}$ and $\sigma^2_t$, $\boldsymbol{\beta} = \boldsymbol{\beta}^* + n^{-1/2}\,\boldsymbol{\varphi}$, $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*)$ and $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \varphi_2, \varphi_3)$; and for $\sigma^2_{n,t,0}$ and $\sigma^2_{t,0}$, $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. Then by Lemmas 1 and 3 we obtain

$$\begin{aligned} log\,\frac{\sigma^2_{n,t}}{\sigma^2_{n,t,0}} &= n^{-1/2}\,\varphi_3\, e^{\beta_1^*\, t} + n^{-1/2}\,\varphi_1\, t\, e^{\beta_1^*\, t}\,\beta_3^* + n^{-1/2}\int_0^t e^{\beta_1^*\,(t-s)}\,\{\varphi_2 + \varphi_1\,\beta_2^*\,(t-s)\}\, dW^{(n)}_{2,s} \\ &\quad + n^{-1/2}\int_0^t e^{\beta_1^*\,(t-s)}\{\varphi_0 + \beta_0^*\,\varphi_1\,(t-s)\}\, ds + O_p(n^{-1}) \\ &= n^{-1/2}\, V_t + O_p(n^{-1}\, log^2 n), \end{aligned}$$

and thus

$$V_{n,t} = n^{1/2}\left\{1 - exp\left(-n^{-1/2}\, V_t + O_p(n^{-1}\, log^2 n)\right)\right\} \;=\; V_t + O_p(n^{-1/2}\, log^2 n). \qquad \square$$

## 4.4 Proof of Proposition 1 for the linear GARCH model

**Lemma 4** *The solution of (17) is given by*

$$\sigma_t^2 = exp\left\{\beta_1\, t + \beta_2\, W_{3,t} - \beta_2^2\, t/2\right\} \left\{e^{\beta_3} + \beta_0 \int_0^t exp\left(-\beta_1\, s - \beta_2\, W_{3,s} + \beta_2^2\, s/2\right)\, ds\right\}.$$

Proof. Applying Itô lemma to the process given by the lemma, we have

$$\begin{aligned} d\sigma_t^2 &= \sigma_t^2\, d(\beta_1\, t + \beta_2\, W_{3,t} - \beta_2^2\, t/2) + (\beta_2^2/2)\, \sigma_t^2\, dt + \beta_0\, dt \\ &= (\beta_0 + \beta_1\, \sigma_t^2)\, dt + \beta_2\, \sigma_t^2\, dW_{3,t}. \qquad \square \end{aligned}$$

**Lemma 5** *The process defined by (16) has the expression*

$$\sigma_{n,k}^2 = \prod_{j=1}^{k-1} (\alpha_1 + \alpha_2\, \zeta_j) \left\{\sigma_0^2 + \alpha_0 \sum_{i=1}^{k-1} \prod_{j=1}^{i-1} (\alpha_1 + \alpha_2\, \zeta_j)^{-1}\right\},$$

*with* $\sigma_0^2 = e^{\beta_3}$, $\alpha_0 = s_n\, \beta_0$, $\alpha_1 = 1 + s_n\, \beta_1$, *and* $\alpha_2 = s_n^{1/2}\, \beta_2$.

Proof. The lemma is proved by using (16) recursively. $\square$

**Lemma 6**

$$\sigma_{n,t}^2 = exp\left\{\beta_1\, t + \beta_2\, W_{3,t}^{(n)} - \frac{\beta_2^2\, t}{2}\right\} \left\{e^{\beta_3} + \beta_0 \int_0^t exp\left(-\beta_1\, s - \beta_2\, W_{3,s}^{(n)} + \frac{\beta_2^2\, s}{2}\right)\, ds\right\} + O_p(n^{-1}).$$

**Remark 4.3**. For the reason discussed in Remark 4.2, we can't substitute $W_3^{(n)}$ by $W_3$ now.

Proof of Lemma 6. An application of Doléans-Dade formula (Jacod and Shiryaev 1987) implies that $\prod_{j=1}^{[n\,t]} (\alpha_1 + \alpha_2\, \zeta_j)$ is an exponential semi-martingale, which converges weakly to the exponential semi-martingale $exp\left\{\beta_1\, t + \beta_2\, W_{3,t} - \beta_2^2\, t/2\right\}$. Moreover,

$$\begin{aligned} &\prod_{j=1}^{[n\,t]} (\alpha_1 + \alpha_2\, \zeta_j) = exp\left\{\sum_{j=1}^{[n\,t]} log\, \left(1 + s_n\, \beta_1 + s_n^{1/2}\, \beta_2\, \zeta_j\right)\right\} \\ &= exp\left(\beta_1\, t + \beta_2\, W_{3,t}^{(n)} - \beta_2^2\, t/2\right) + O_p(n^{-1}), \end{aligned}$$

and hence

$$\begin{aligned} \alpha_0 \sum_{i=1}^{[n\,t]} \prod_{j=1}^{i-1} (\alpha_1 + \alpha_2\, \zeta_j)^{-1} &= s_n\, \beta_0 \sum_{i=1}^{[n\,t]} exp\left\{-\beta_1\, (i-1)/n - \beta_2\, W_{3,(i-1)/n}^{(n)} \right. \\ &\qquad \left. + \beta_2^2\, (i-1)/(2\, n)\right\} + O_p(n^{-1}) \\ &= \int_0^t exp\left(-\beta_1\, s - \beta_2\, W_{3,s}^{(n)} - \beta_2^2\, s/2\right)\, ds + O_p(n^{-1}). \end{aligned}$$

Now the lemma is a consequence of Lemmas 4 and 5 □

**Proof of Proposition 1**. Note from Convention 4 in section 4.1 that for $\sigma^2_{n,t}$ and $\sigma^2_t$, $\boldsymbol{\beta} = \boldsymbol{\beta}^* + n^{-1/2}\,\boldsymbol{\varphi}$, $\boldsymbol{\beta}^* = (\beta^*_0, \beta^*_1, \beta^*_2, \beta^*_3)$ and $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \varphi_2, \varphi_3)$; and for $\sigma^2_{n,t,0}$ and $\sigma^2_{t,0}$, $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. By Lemmas 4 and 6 we obtain $\sigma^2_{n,t} - \sigma^2_{n,t,0} = D_1 + D_2$, where

$$D_1 = n^{-1/2}\, exp\left\{\beta^*_3 + \beta^*_1\, t + \beta^*_2\, W^{(n)}_{3,t} - (\beta^*_2)^2\, t/2\right\} \left\{\varphi_3 + \varphi_1\, t + \varphi_2\, W^{(n)}_{3,t} - \varphi_2\, \beta^*_2\, t\right\} + O_p(n^{-1})$$
$$= n^{-1/2}\, exp\left\{\beta^*_3 + \beta^*_1\, t + \beta^*_2\, W_{3,t} - (\beta^*_2)^2\, t/2\right\} \{\varphi_3 + \varphi_1\, t + \varphi_2\, W_{3,t} - \varphi_2\, \beta^*_2\, t\} + O_p(n^{-1}\, log^2\, n),$$

$$\begin{aligned} D_2 &= n^{-1/2} \int_0^t exp\left\{\beta^*_1\,(t-s) + \beta^*_2\,(W^{(n)}_{3,t} - W^{(n)}_{3,s}) - (\beta^*_2)^2\,(t-s)/2\right\} \\ &\quad \left\{\varphi_0 + \varphi_1\, \beta^*_0\,(t-s) + \varphi_2\, \beta^*_0\,(W^{(n)}_{3,t} - W^{(n)}_{3,s}) - \varphi_2\, \beta^*_0\, \beta^*_2\,(t-s)\right\}\, ds + O_p(n^{-1}) \\ &= n^{-1/2} \int_0^t exp\left\{\beta^*_1\,(t-s) + \beta^*_2\,(W_{3,t} - W_{3,s}) - (\beta^*_2)^2\,(t-s)/2\right\} \\ &\quad \{\varphi_0 + \varphi_1\, \beta^*_0\,(t-s) + \varphi_2\, \beta^*_0\,(W_{3,t} - W_{3,s}) - \varphi_2\, \beta^*_0\, \beta^*_2\,(t-s)\}\, ds + O_p(n^{-1}\, log^2\, n). \end{aligned}$$

Finally,

$$\begin{aligned} V_{n,t} &= n^{1/2}\left\{1 - (\sigma^2_{n,t,0} + D_1 + D_2)^{-1}\, \sigma^2_{n,t,0}\right\} \\ &= n^{1/2}\,(D_1 + D_2)\, \sigma^{-2}_{t,0} + O_p(n^{-1/2}\, log^2\, n) = V_t + O_p(n^{-1/2}\, log^2\, n). \qquad \Box \end{aligned}$$

# 5 Likelihood process for diffusion models

For discrete observations $X_{t_k}$ from (14), $k = 1, \cdots, n$, we have

$$\begin{aligned} X_{t_k} - X_{t_{k-1}} &= \int_{t_{k-1}}^{t_k} (\gamma_0 + \gamma_1\, \sigma^2_s)\, ds + \int_{t_{k-1}}^{t_k} \sigma_u\, dW_{1,u} \\ &= \gamma_0\, s_n + \gamma_1\, s_n\, \bar{\sigma}^2_{n,t_k} + \{s_n\, \bar{\sigma}^2_{n,t_k}\}^{1/2}\, z_k, \end{aligned} \tag{32}$$

where

$$\bar{\sigma}^2_{n,t} = s_n^{-1} \int_{t-s_n}^t \sigma^2_u\, du, \qquad z_k = \{s_n\, \bar{\sigma}^2_{n,t_k}\}^{-1/2} \int_{t_{k-1}}^{t_k} \sigma_u\, dW_{1,u}. \tag{33}$$

From (14), (15) and (17), we have that conditioning on $W_2$ and $W_3$, $\int_{t_{k-1}}^{t_k} \sigma_u\, dW_{1,u}$ are independent and follow normal distributions with mean zero and variances $s_n\, \bar{\sigma}^2_{n,t_k}$, and thus $z_k$ are i.i.d. standard random variables.

Keep in mind the notations and conventions specified in section 4.1. In particular, use $\boldsymbol{\vartheta} = (\boldsymbol{\varphi}, \boldsymbol{\gamma})$ for $\boldsymbol{\theta} = (\boldsymbol{\beta}^* + n^{-1/2}\, \boldsymbol{\varphi}, \boldsymbol{\gamma}) \in \Theta_{n,c}(\boldsymbol{\beta}^*)$ and denote the processes $X$ and $\sigma^2$ under $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^*$ by adding a subscript "0" to the processes.

Denote by $E_{W_{23}}$ the expectation taken with respect to $W_2$ and $W_3$. Conditioning on $\bar{\sigma}^2_{n,t_j}$ we obtain the joint conditional density of $X_{t_1}, \cdots, X_{t_n}$

$$\prod_{j=1}^{n} \left[ (s_n\, \bar{\sigma}^2_{n,t_j})^{-1/2}\, \phi\left( \{X_{t_j} - X_{t_{j-1}} - s_n(\gamma_0 + \gamma_1\, \bar{\sigma}^2_{n,t_j})\}\, (s_n^{1/2}\, \bar{\sigma}_{n,t_j})^{-1} \right) \right],$$

where $\phi$ is the density of the standard normal distribution. Averaging out $\bar{\sigma}^2_{n,t_k}$, $k = 1, \cdots, n$, in the conditional density, we get the joint density function and have the likelihood

$$\begin{aligned} L_{n,2}(\boldsymbol{\vartheta}) &= E_{W_{23}} exp \left( - \sum_{j=1}^{n} (2\, s_n\, \bar{\sigma}^2_{n,t_j})^{-1} \left\{ X_{t_j} - X_{t_{j-1}} - s_n(\gamma_0 + \gamma_1\, \bar{\sigma}^2_{n,t_j}) \right\}^2 \right. \\ &\qquad \left. - \sum_{j=1}^{n} log\, \bar{\sigma}_{n,t_j} - (n/2)\, log(2\, \pi\, s_n) \right), \end{aligned}$$

and thus the likelihood process under $Q_{n,\boldsymbol{\vartheta}^*}$

$$\begin{aligned} &\Lambda_{n,2}(\boldsymbol{\vartheta}) = L_{n,2}(\boldsymbol{\vartheta}) / L_{n,2}(\boldsymbol{\vartheta}^*) \\ = \; & E_{W_{23}} exp \left( - \sum_{j=1}^{n} (2\, s_n\, \bar{\sigma}^2_{n,t_j})^{-1} \{X_{t_j,0} - X_{t_{j-1},0} - s_n(\gamma_0 + \gamma_1\, \bar{\sigma}^2_{n,t_j})\}^2 - \sum_{j=1}^{n} log\, \bar{\sigma}_{n,t_j} \right) \\ & \left\{ E_{W_{23}} exp \left( - \sum_{j=1}^{n} (2\, s_n\, \bar{\sigma}^2_{n,t_j,0})^{-1} (X_{t_j,0} - X_{t_{j-1},0})^2 - \sum_{j=1}^{n} log\, \bar{\sigma}_{n,t_j,0} \right) \right\}^{-1} \\ = \; & E_{W_{23}} \left[ B_n\, exp \left\{ \frac{1}{2} \sum_{j=1}^{n} \left( 1 - \frac{\bar{\sigma}^2_{n,t_j,0}}{\bar{\sigma}^2_{n,t_j}} \right) (z_j^2 - 1) + \frac{1}{2} \sum_{j=1}^{n} \left( 1 + log\, \frac{\bar{\sigma}^2_{n,t_j,0}}{\bar{\sigma}^2_{n,t_j}} - \frac{\bar{\sigma}^2_{n,t_j,0}}{\bar{\sigma}^2_{n,t_j}} \right) \right. \right. \\ & \left. \left. + s_n^{1/2} \sum_{j=1}^{n} \left( \gamma_0 + \gamma_1\, \bar{\sigma}^2_{n,t_j} \right) \bar{\sigma}_{n,t_j,0}\, \bar{\sigma}^{-2}_{n,t_j}\, z_j - \frac{s_n}{2} \sum_{j=1}^{n} \left( \gamma_0 + \gamma_1\, \bar{\sigma}^2_{n,t_j} \right)^2 \bar{\sigma}^{-2}_{n,t_j} \right\} \right] \\ = \; & E_{W_{23}} \left[ B_n\, exp \left\{ \frac{1}{\sqrt{2}} \int_0^1 \bar{V}_{n,t}\, dW^{(n)}_{4,t} + \frac{1}{2} \sum_{j=1}^{n} \bar{H}_{n,t} / n \right. \right. \\ & \left. \left. + \int_0^1 \left( \gamma_0 + \gamma_1\, \bar{\sigma}^2_{n,t} \right) \bar{\sigma}_{n,t,0}\, \bar{\sigma}^{-2}_{n,t}\, dW^{(n)}_{1,t} - \frac{s_n}{2} \sum_{j=1}^{n} \left( \gamma_0 + \gamma_1\, \bar{\sigma}^2_{n,t_j} \right)^2 \bar{\sigma}^{-2}_{n,t_j} \right\} \right], \end{aligned} \tag{34}$$

where the third equation is due to (32),

$$B_n = \left\{ E_{W_{23}} exp \left( - \sum_{j=1}^{n} log\, \bar{\sigma}_{n,t_j,0} \right) \right\}^{-1} exp \left( - \sum_{j=1}^{n} log\, \bar{\sigma}_{n,t_j,0} \right), \tag{35}$$

$$\bar{V}_{n,t} = n^{1/2} \left( 1 - \frac{\bar{\sigma}^2_{n,t,0}}{\bar{\sigma}^2_{n,t}} \right), \qquad \bar{H}_{n,t} = n \left( 1 + log\, \frac{\bar{\sigma}^2_{n,t_j,0}}{\bar{\sigma}^2_{n,t_j}} - \frac{\bar{\sigma}^2_{n,t_j,0}}{\bar{\sigma}^2_{n,t_j}} \right), \tag{36}$$

and

$$W^{(n)}_{1,t} = n^{-1/2} \sum_{k=1}^{[n\, t]} z_k, \qquad W^{(n)}_{4,t} = (2\, n)^{-1/2} \sum_{k=1}^{[n\, t]} (z_k^2 - 1). \tag{37}$$

Note that by Convention 1 in section 4.1 we use the same notation $W_{1,t}^{(n)}$ here for normalized partial sum processes $z_k$ (in the diffusion model), as that for $\varepsilon_k$ (in the GARCH model, see (28)), because they have the same distribution, are uncorrelated with other processes, and play an identical role in the corresponding likelihood processes. However, a notation $W_{4,t}^{(n)}$ different from $W_{3,t}^{(n)}$ is introduced to denote the normalized partial sum process for $z_k^2$. This is due to the fact that unlike the GARCH case that $\varepsilon_k^2$'s are correlated with the conditional variances $\sigma_{n,t}^2$, in the diffusion model $z_k^2$'s are independent of the conditional variances $\bar{\sigma}_{n,t}^2$, and thus jointly $(\varepsilon_k^2, \sigma_{n,t}^2)$ and $(z_k^2, \bar{\sigma}_{n,t}^2)$ are not identically distributed, although marginally $\varepsilon_k^2$'s and $z_k^2$'s are. In fact, this is a key point for the difference between the two likelihoods.

By strong approximation (Komlós, Major and Tusnády, 1975) there exists a standard Brownian motion $W_{4,t}$ independent of $W_{1,t}$, $W_{2,t}$, $W_{3,t}$ such that

$$\sup_{0 \le t \le T} \left\{ |W_{1,t}^{(n)} - W_{1,t}| + |W_{4,t}^{(n)} - W_{4,t}| \right\} = O_p\left( n^{-1/2} log\, n \right). \tag{38}$$

**Lemma 7** *For the average volatility process $\bar{\sigma}_{n,t}^2$ defined in (33), we have*

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

*for the multiplicative GARCH case, and*

$$\bar{\sigma}_{n,t}^2 = \sigma_t^2 \int_0^1 exp\left( -\beta_2\, s_n^{1/2}\, \tilde{W}_{3,u} \right)\, du + O_p(n^{-1})$$

*and for the linear GARCH case, where $\tilde{W}_{i,u} = s_n^{-1/2}\, (W_{i,t} - W_{i,t-s_n u})$ are rescaled Brownian motions.*

**Remark 5.1**. Because of stationarity and the rescaling property of Brownian motions $W_i$, $\tilde{W}_i$ are Brownian motions whose distributions are independent of $t$. Because of Convention 1 in section 4.1, we don't keep track $t$. Also both $\int_0^1 exp\left( -\beta_2\, s_n^{1/2} \int_0^u e^{\beta_1 v}\, d\tilde{W}_{2,v} \right) du$ and $\int_0^1 exp\left( -s_n^{1/2}\, \beta_2\, \tilde{W}_{3,u} \right) du$ are of order $1 + O_p(n^{-1/2})$. For the reason discussed in Remark 4.2, we can't replace them by 1 here.

Proof of Lemma 7. For the multiplicative GARCH case, from the definition of $\bar{\sigma}^2$ in (33) and Lemma 1 we have

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

Similarly, for the linear GARCH case, by Lemma 4 we get

$$
\begin{aligned}
\bar{\sigma}^2_{n,t} &= \sigma^2_t \int_0^1 exp\left\{-s_n\,(\beta_1-\beta_2^2/2)\,u - s_n^{1/2}\,\beta_2\,\tilde{W}_{3,u}\right\}\,du \\
&\quad -\beta_0 \int_0^1 du \int_{t-s_n\,u}^t exp\left\{(\beta_1-\beta_2^2/2)\,(t-s_n\,u-h)+\beta_2\,(W_{3,t-s_n\,u}-W_{3,h})\right\}\,dh \\
&= \sigma^2_t \int_0^1 exp\left(-s_n^{1/2}\,\beta_2\,\tilde{W}_{3,u}\right)\,du + O_p(s_n). \qquad \square
\end{aligned}
$$

**Lemma 8**

$$\bar{V}_{n,t} = V_t + O_p(n^{-1/2}).$$

*where $V_t$ and $\bar{V}_{n,t}$ are defined in (27) and (36), respectively.*

Proof. Thank to Lemma 7, we can now easily show that for both multiplicative and linear GARCH cases,

$$\frac{\bar{\sigma}^2_{n,t,0}}{\bar{\sigma}^2_{n,t}} = \frac{\sigma^2_{t,0}}{\sigma^2_t} + O_p(n^{-1}). \tag{39}$$

Because of similarity, we prove (39) only for the multiplicative GARCH case. Note from Convention 4 in section 4.1 that for $\sigma^2_t$, $\boldsymbol{\beta} = \boldsymbol{\beta}^* + n^{-1/2}\,\boldsymbol{\varphi}$, $\boldsymbol{\beta}^* = (\beta^*_0, \beta^*_1, \beta^*_2, \beta^*_3)$ and $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \varphi_2, \varphi_3)$; and for $\sigma^2_{t,0}$, $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. By Lemma 7 we can derive

$$
\begin{aligned}
\bar{\sigma}^2_{n,t} &= \sigma^2_t \int_0^1 exp\left(-\beta^*_2\,s_n^{1/2}\int_0^u e^{\beta^*_1\,v}\,d\tilde{W}_v\right)\,du + O_p(n^{-1}) \\
&= \bar{\sigma}^2_{n,t,0}\,(\sigma_t/\sigma_{t,0})^2\; + O_p(n^{-1}),
\end{aligned}
$$

which proves (39).

From Lemmas 1 and 4 we get that $\sigma^2_t$ depends on $\boldsymbol{\beta}$ smoothly. Expanding $\sigma^2_t$ at $\boldsymbol{\beta}^*$ and using the definition of $V_t$ we get

$$log\,\sigma^2_t \;=\; log\,\sigma^2_{t,0} + n^{-1/2}\,V_t + O_p(n^{-1}).$$

Finally, combining above equation with (39) we arrive at

$$\bar{V}_{n,t} \;=\; n^{1/2}\left\{1 - exp\left(-n^{-1/2}\,V_t\right) + O_p(n^{-1})\right\} = V_t + O_p(n^{-1/2}). \qquad \square$$

**Lemma 9**

$$\bar{H}_{n,t} = -V_t^2/2 + O_p(n^{-1/2}),$$

*where $V_t$ and $\bar{H}_{n,t}$ are defined in (27) and (36), respectively.*

Proof. The lemma is easily proved by Lemma 8 and the same argument for proving Proposition 2. □

There is a great difficulty in handling $\Lambda_{n,2}(\boldsymbol{\vartheta})$. First, it is generally impossible to derive an explicit and tractable form for $\Lambda_{n,2}(\boldsymbol{\vartheta})$. Second, the factor $B_n$ in $\Lambda_{n,2}(\boldsymbol{\vartheta})$ converges in probability to zero but it is unbounded, because $E\,B_n = 1$, and its first factor [see (35)] has an exponent of order $n^2\, var(\int_0^1 log\, \sigma_t^2\, dt)/8$. Thus we can not do usual asymptotics by exchanging the limit "$n \to \infty$" and the expectation $E_{W_{23}}$. To get around this difficulty and prove the theorems we consider the likelihood process in a local neighborhood around $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*)$ with $\beta_2^* = 0$. In this case $\bar{\sigma}_{n,t,0}^2$ is deterministic but both $\bar{\sigma}_{n,t}^2$ and $V_t$ are stochastic, since $\boldsymbol{\beta} = \boldsymbol{\beta}^* + n^{-1/2}\,\boldsymbol{\varphi}$. Also for simplicity, we take $\boldsymbol{\gamma} = 0$. Thus, $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^*, \mathbf{0})$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{0})$, or equivalently by Convention 3 in section 4.1, $\boldsymbol{\vartheta}^* = (\mathbf{0}, \mathbf{0})$ and $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^+ = (\boldsymbol{\varphi}, \mathbf{0})$.

**Proposition 5** *Suppose $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*, 0, \beta_3^*)$ and define*

$$\Lambda_2(\boldsymbol{\vartheta}^+) = E_{W_{23}} exp\left\{ \frac{1}{\sqrt{2}} \int_0^1 V_t\, dW_{4,t} - \frac{1}{4}\int_0^1 V_t^2\, dt \right\}, \tag{40}$$

*where $V_t$ and $W_{4,t}$ are given by (27) and (38), respectively. Then as $n \to \infty$, $E|\Lambda_{n,2}(\boldsymbol{\vartheta}^+) - \Lambda_2(\boldsymbol{\vartheta}^+)| \to 0$.*

Proof. Since $\beta_2^* = 0$, $\sigma_{t,0}^2$ is deterministic, and then (35) implies $B_n = 1$. Taking $B_n = 1$ and $\boldsymbol{\gamma} = 0$ in (34) we get

$$\Lambda_{n,2}(\boldsymbol{\vartheta}^+) \;=\; E_{W_{23}} exp\left\{ \frac{1}{2}\int_0^1 \bar{V}_{n,t}\, dW_{4,t}^{(n)} + \frac{1}{2}\sum_{j=1}^n \bar{H}_{n,t}/n \right\} \tag{41}$$

It is easy to check that as a likelihood process, $E\Lambda_{n,2}(\boldsymbol{\vartheta}^+) = 1$.

The process $V_t$ is a continuous functional of $W_2$ or $W_3$ and has a.s. continuous sample paths, so conditional on $W_2$ and $W_3$, $\int_0^1 V_t^2\, dt$ is a.s. finite. As $W_{1,t}$, $\{W_{2,t}, W_{3,t}\}$, $W_{4,t}$ are independent but $W_{2,t}$ and $W_{3,t}$ are correlated, we have that conditioning on $W_2$ and $W_3$, Novikov's condition (i.e. $E[exp\{\int_0^1 V_t^2\, dt/4\}|W_2, W_3] < \infty$) holds, and thus the exponential martingale $exp\left\{\frac{1}{\sqrt{2}}\int_0^t V_s\, dW_{4,s} - \frac{1}{4}\int_0^t V_s^2\, ds\right\}$ is indeed a martingale, and $E\left[ exp\left\{\frac{1}{\sqrt{2}}\int_0^1 V_t\, dW_{4,t} - \frac{1}{4}\int_0^1 V_t^2\, dt\right\}\Big| W_2, W_3\right] = 1$. (See Ikeda and Watanabe 1989, Karatzas and Shreve 1997, section 3.5.) Therefore,

$$\begin{aligned} E\Lambda_2(\boldsymbol{\vartheta}^+) &= E_{W_4} E_{W_{23}} exp\left\{ \frac{1}{\sqrt{2}}\int_0^1 V_t\, dW_{4,t} - \frac{1}{4}\int_0^1 V_t^2\, dt \right\} \\ &= E_{W_{23}} E\left[ exp\left\{ \frac{1}{\sqrt{2}}\int_0^1 V_t\, dW_{4,t} - \frac{1}{4}\int_0^1 V_t^2\, dt \right\}\Bigg| W_2, W_3 \right] = 1. \end{aligned}$$

As in section 4.2, by strong approximation (38) and Lemmas 8 and 9 we obtain that the process inside $E_{W_{23}}$ in the expression (41) of $\Lambda_{n,2}(\boldsymbol{\vartheta}^+)$

$$exp\left\{\frac{1}{\sqrt{2}}\int_0^1 \bar{V}_{n,t}\, dW_{4,t}^{(n)} + \frac{1}{2}\sum_{j=1}^{n} \bar{H}_{n,t}/n\right\}, \tag{42}$$

converges in probability to

$$exp\left\{\frac{1}{\sqrt{2}}\int_0^1 V_t\, d\,W_{4,t} - \frac{1}{4}\int_0^1 V_t^2\, d\,t\right\}, \tag{43}$$

which is the process inside $E_{W_{23}}$ in the expression (40) of $\Lambda_2(\boldsymbol{\vartheta}^+)$. Both processes are nonnegative and have expectation one, so applying the Scheffé theorem, we have that the expectation of the absolute difference between (42) and (43) converges to zero. That in turn implies $E|\Lambda_{n,2}(\boldsymbol{\vartheta}^+) - \Lambda_2(\boldsymbol{\vartheta}^+)| \to 0$. $\square$

**Remark 5.2**. As we mentioned in section 2.2 and discussed before the proposition, the likelihood process for a diffusion model is generally not available. Even for the limit of the multiplicative GARCH model (40) with $\beta_0^* = \beta_1^* = 0$, we will show that its explicit form is too complicated to be useful. In this case, as shown in Remark 4.1, we have $V_t = \varphi_2\, W_{2,t}$, and the Karhunen-Loève expansion of $W_{2,t}$,

$$W_{2,t} = \sum_{j=0}^{\infty} 2^{1/2}\, \pi^{-1}\, (j+1/2)^{-1}\, sin\left\{(j+1/2)\, \pi\, t\right\}\, z_j,$$

where $z_j$ are i.i.d. standard normal random variables. Then

$$\int_0^1 W_{2,t}\, dW_{4,t} = \sum_{j=0}^{\infty} \pi^{-1}\, (j+1/2)^{-1}\, w_j\, z_j, \qquad \int_0^1 W_{2,t}^2\, dt = \sum_{j=0}^{\infty} \pi^{-2}\, (j+1/2)^{-2}\, z_j^2,$$

where the Fourier coefficients $w_j = \int_0^1 \sqrt{2}\, sin\left\{(j+1/2)\, \pi\, t\right\}\, dW_{4,t}$ are i.i.d. standard normal random variables and independent of $z_j$. Finally,

$$\begin{aligned}
\Lambda_2(\boldsymbol{\vartheta}^+) &= \prod_{j=0}^{\infty} E_{z_j}\, exp\left\{\frac{\varphi_2\, w_j\, z_j}{\sqrt{2}\,\pi\,(j+1/2)} - \frac{\varphi_2^2\, z_j^2}{4\,\pi^2\,(j+1/2)^2}\right\} \\
&= exp\left\{\sum_{j=0}^{\infty} \frac{\varphi_2^2\, w_j^2}{2\,\pi^2\,(j+1/2)^2 + \varphi_2^2}\right\} \prod_{j=0}^{\infty}\left\{\frac{\varphi_2^2}{2\,\pi^2\,(j+1/2)^2} + 1\right\}^{-1/2}.
\end{aligned}$$

The non-random product series converges, and the exponent of the random part is a positive linear combination of the squared Fourier coefficients of $W_{4,t}$.

# 6 Proof of Theorems

## 6.1 Proof of Theorem 1: Nonequivalence under stochastic volatility

If $\Delta(\mathbb{E}_{n,1}, \mathbb{E}_{n,2}) \to 0$, as $n \to \infty$, Propositions 4 and 5 imply that with $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*, 0, \beta_3^*)$, for all $\boldsymbol{\vartheta}^+ = (\boldsymbol{\varphi}, \mathbf{0})$, $\Lambda_1(\boldsymbol{\vartheta}^+)$ and $\Lambda_2(\boldsymbol{\vartheta}^+)$ must have the same distribution (Le Cam 1986, Proposition 8 and its remark in section 4 of chapter 6, pp. 93-95, Le Cam and Yang 1990, section 2.2 of chapter 2). However, $W_{1,t}$, $\{W_{2,t}, W_{3,t}\}$, $W_{4,t}$ are independent but $W_{2,t}$ and $W_{3,t}$ are correlated, and from (27), Lemmas 1 and 4, we can see that $V_t$ strongly depends on $W_{2,t}$ or $W_{3,t}$, so $V_t$ is correlated with $W_{3,t}$ and independent of $W_{4,t}$. Thus, $\int_0^1 V_t\, dW_{3,t}$ and $\int_0^1 V_t\, dW_{4,t}$ can't have the same distribution. With $\boldsymbol{\beta}^*$ and $\boldsymbol{\vartheta}^+$ given above, from (30) and (40), we get the following expressions for $\Lambda_i(\boldsymbol{\vartheta}^+)$,

$$\Lambda_1(\boldsymbol{\vartheta}^+) = exp\left\{ \frac{1}{\sqrt{2}} \int_0^1 V_t\, dW_{3,t} - \frac{1}{4} \int_0^1 V_t^2\, dt \right\},$$

$$\Lambda_2(\boldsymbol{\vartheta}^+) = E_{W_{23}} exp\left\{ \frac{1}{\sqrt{2}} \int_0^1 V_t\, dW_{4,t} - \frac{1}{4} \int_0^1 V_t^2\, dt \right\},$$

we now easily conclude the contradiction that $\Lambda_1(\boldsymbol{\vartheta}^+)$ and $\Lambda_2(\boldsymbol{\vartheta}^+)$ can't be identically distributed. Indeed, for example, for the multiplicative GARCH case, from (27), (30) and Lemma 1 we obtain

$$V_t\big|_{\boldsymbol{\varphi}=\mathbf{0}} = 0, \qquad \Lambda_1(\boldsymbol{\vartheta}^+)\Big|_{\boldsymbol{\varphi}=\mathbf{0}} = 1,$$

and

$$\frac{\partial V_t}{\partial \varphi_2}\bigg|_{\boldsymbol{\varphi}=\mathbf{0}} = \frac{\partial log\, \sigma_{t,0}^2}{\partial \beta_2^*} = exp(\beta_1^*\, t) \int_0^t exp(-\beta_1^*\, s)\, dW_{2,s}.$$

Thus,

$$\frac{\partial \Lambda_1(\boldsymbol{\vartheta}^+)}{\partial \varphi_2}\bigg|_{\boldsymbol{\varphi}=\mathbf{0}} = \frac{1}{\sqrt{2}} \int_0^1 \frac{\partial V_t}{\partial \varphi_2}\bigg|_{\boldsymbol{\varphi}=\mathbf{0}} dW_{3,t} = \frac{1}{\sqrt{2}} \int_0^1 exp(\beta_1^*\, t) \int_0^t exp(-\beta_1^*\, s)\, dW_{2,s}\, dW_{3,t},$$

which is a double Itô integral and has mean zero and variance

$$\frac{1}{2} \int_0^1 exp(2\, \beta_1^*\, t) \int_0^t exp(-2\, \beta_1^*\, s)\, ds\, dt = \{exp(2\, \beta_1^*) - 2\, \beta_1^* - 1\}\, \{8\, (\beta_1^*)^2\}^{-1} \;\; \geq \;\; 1/4.$$

But due to the independence of $W_2$ and $W_4$, a similar calculation leads to

$$\begin{aligned}\left.\frac{\partial \Lambda_2(\boldsymbol{\vartheta}^+)}{\partial \varphi_2}\right|_{\boldsymbol{\varphi}=\mathbf{0}} &= E_{W_{23}} \int_0^1 \left.\frac{\partial V_t}{\partial \varphi_2}\right|_{\boldsymbol{\varphi}=\mathbf{0}} dW_{4,t} \\ &= \frac{1}{\sqrt{2}} \int_0^1 exp(\beta_1^* \, t)\, E_{W_{23}} \left\{ \int_0^t exp(-\beta_1^* \, s)\, dW_{2,s} \right\} dW_{4,t} = 0.\end{aligned}$$

Therefore, $\Lambda_1(\boldsymbol{\vartheta}^+)$ and $\Lambda_2(\boldsymbol{\vartheta}^+)$ can't be identically distributed.

## 6.2 Proof of Theorem 2: Equivalence under nonstochastic volatility

Because of similarity, we give the arguments only for the multiplicative GARCH case. With deterministic volatility, the two models become a regression model with independent normal errors and the white noise model. Both the GARCH observations $X_{n,t_k}$ [defined in (11)] and the discrete diffusion observations $X_{t_k}$ [given by (32)] have independent increments, and both increments $X_{n,t_k} - X_{n,t_{k-1}}$ and $X_{t_k} - X_{t_{k-1}}$ follow normal distributions with means $\nu_{1,n,k} = (\gamma_0 + \gamma_1 \, \sigma^2_{n,t_k})\, s_n$ and $\nu_{2,n,k} = (\gamma_0 + \gamma_1 \, \bar\sigma^2_{n,t_k})\, s_n$, and variances $\tau^2_{1,n,k} = \sigma^2_{n,t_k}\, s_n$ and $\tau^2_{2,n,k} = \bar\sigma^2_{n,t_k}\, s_n$, respectively. Therefore, we calculate the Hellinger distance, $H(P_{n,\boldsymbol{\theta}}, Q_{n,\boldsymbol{\theta}})$, between the distribution $P_{n,\boldsymbol{\theta}}$ of $X_{n,t_k}$, $k = 0, \cdots, n$, and the distribution $Q_{n,\boldsymbol{\theta}}$ of $X_{t_k}$ , $k = 0, \cdots, n$, as follows.

$$\begin{aligned} & H^2(P_{n,\boldsymbol{\theta}}, Q_{n,\boldsymbol{\theta}}) = H^2\left( \prod_{k=1}^n N(\nu_{1,n,k}, \tau^2_{1,n,k}), \prod_{k=1}^n N(\nu_{2,n,k}, \tau^2_{2,n,k}) \right) \\ =&\ 2 - 2 \prod_{k=1}^n \left\{ 1 - H^2(N(\nu_{1,n,k}, \tau^2_{1,n,k}), N(\nu_{2,n,k}, \tau^2_{2,n,k}))/2 \right\} \\ =&\ 2 - 2 \prod_{k=1}^n \left( \left\{ \frac{2\, \tau_{1,n,k}\, \tau_{2,n,k}}{\tau^2_{1,n,k} + \tau^2_{2,n,k}} \right\}^{1/2} exp\left\{ -\frac{(\nu_{1,n,k} - \nu_{2,n,k})^2}{\tau^2_{1,n,k} + \tau^2_{2,n,k}} \right\} \right) \\ =&\ 2 - 2 \prod_{k=1}^n \left\{ \frac{2\, \sigma_{n,t_k}\, \bar\sigma_{n,t_k}}{\sigma^2_{n,t_k} + \bar\sigma^2_{n,t_k}} \right\}^{1/2} exp\left\{ -\sum_{k=1}^n \frac{\gamma_1^2\, s_n\, (\sigma^2_{n,t_k} - \bar\sigma^2_{n,t_k})^2}{\sigma^2_{n,t_k} + \bar\sigma^2_{n,t_k}} \right\} \\ =&\ 2 - 2 \prod_{k=1}^n \left\{ 1 - \frac{(\sigma_{n,t_k} - \bar\sigma_{n,t_k})^2}{\sigma^2_{n,t_k} + \bar\sigma^2_{n,t_k}} \right\}^{1/2} exp\left\{ -\sum_{k=1}^n \frac{\gamma_1^2\, s_n\, (\sigma^2_{n,t_k} - \bar\sigma^2_{n,t_k})^2}{\sigma^2_{n,t_k} + \bar\sigma^2_{n,t_k}} \right\}, \qquad (44) \end{aligned}$$

where the second equality is by the Hellinger distance property for independent distributions (Le Cam 1986, chapter 4, Le Cam and Yang 1990, section 3.2), the third equality is from the expression for Hellinger distance between two univariate normal distributions obtained by direct calculations (see also Brown and Low 1996, equation 3.7), and the fourth equality is because of substitutions of $\nu$'s and $\tau^2$'s by $\sigma^2$'s.

Since $\beta_2 = 0$, Lemma 1 shows that for the deterministic diffusion volatility process $\sigma_t^2$,

$$log\, \sigma_t^2 = e^{\beta_1 t} \left\{ \beta_3 + \beta_0 \int_0^t e^{-\beta_1 u}\, du \right\}. \tag{45}$$

Taking $\beta_2 = 0$ and dropping all random terms in Lemmas 2 and 3, we obtain that for the deterministic GARCH volatility process $\sigma_{n,t}^2$,

$$\begin{aligned} log\, \sigma_{n,t}^2 &= (1 + s_n\, \beta_1)^{[n\, t]-1}\, \beta_3 + (s_n\, \beta_0)/(1 + s_n\, \beta_1) \sum_{j=1}^{[n\, t]-1} (1 + s_n\, \beta_1)^{[n\, t]-j} \\ &= e^{\beta_1 t} \left\{ \beta_3 + \beta_0 \int_0^t e^{-\beta_1 s}\, ds \right\} + O(n^{-1}) = log\, \sigma_t^2 + O(n^{-1}). \end{aligned} \tag{46}$$

As deterministic $\sigma_t^2$ given by (45) is smooth in $t$ and $\bar\sigma_{n,t}^2 = s_n^{-1} \int_{t-s_n}^t \sigma_u^2\, du$ is the average of $\sigma_t^2$ over an interval of length $s_n$, simple calculations show $\bar\sigma_{n,t}^2 = \sigma_t^2 + O(n^{-1})$. This result together with (46) imply that uniformly over all $\beta_0, \beta_1, \beta_3$ and $t \in [0, 1]$,

$$\bar\sigma_{n,t}^2 = \sigma_{n,t}^2 + O(n^{-1}). \tag{47}$$

Plugging (47) into (44) we conclude that uniformly over all $\boldsymbol{\theta} = (\beta_0, \beta_1, 0, \beta_3, \gamma_0, \gamma_1) \in \Theta'$, $H^2(P_{n,\boldsymbol{\theta}}, Q_{n,\boldsymbol{\theta}})$ is of order $n^{-1}$. Finally we complete the proof by applying (23).

# REFERENCES


Aït-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. Econometrica **64**, 527-560.

Basawa, I. and Prakasa Rao, B. L. S. (1980). Statistical Inference for Stochastic Processes. Academic Press.

Black, F. and Sholes, M. (1973). The pricing of options and corporate liabilities. Journal of Plitical 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. and Low, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. Annals of Statistics **24**, 2384-2398.

Brown, L. and Wang, Y. (2000). Statistical Equivalence at Suitable Frequencies of GARCH and Stochastic Volatility Models with the Corresponding Diffusion Model.

Danielsson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. Journal of Econometrics **64**, 375-400.

Dothan, M. U. (1990). Prices in Financial Markets. Oxford University Press.

Drost, F. C. and Klaassen, C. A. J. (1997). Efficient estimation in semiparametric GARCH models. Journal of Econometrics **81**, 193-221.

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 limit. 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.

Florens-Zmirou, D. (1989). Approximate discrete time schemes for statistics of diffusion processes. Statistics **20**, 547-557.

Gallant, A. R., Hsieh, D., and Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. Journal of Econometrics **81**, 159-192.

Gallant, A. R. and Long, J. R. (1997). Estimating stochastic differential equations efficiently by minimum chi-squared. Biometrika **84**, 125-141.

Gallant, A. R. and Tauchen, G. (1998). Reprojecting partially observed systems with application to interest rate diffusions. Journal of American Statistical Association **93**, 10-24.

Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. Statistics **21**, 99-116.

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.

Harrison, M. and Kreps, D. (1979). Martingales and arbitrage in multiperiod securities markets. Journal of Economic Theory **20**, 381-408.

Harrison, M. and Pliska, S. (1981). Martingales and stochastic integrals in the theory of continuous trading. Stochastic Processes and its Application **11**, 215-260.

Hsieh, D. (1991). Chaos and nonlinear dynamics: applications to financial markets. Journal of Finance **46**, 1839-1877.

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.

Ikeda, N. and Watanabe, S. (1989). Stochastic differential equations and diffusion processes. North-Holland. Second edition.

Ingersoll, J. E. (1987). Theory of Financial Decision Making. Rowman & Littlefield.

Jacod, J. and Shiryaev, A. N. (1987). Limit Theorems for Stochastic Processes. Springer.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion). Journal of Business & Economic Statistics **12**, 371-417.

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

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. Scand. J. Statist. **24**, 211-229.

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. (1990). Asymptotics in Statistics – Some Basic Concepts. Springer.

Merton, R. (1973). The theory of rational option pricing. Bell Journal of Economics and Management Science **4**, 141-183.

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

Mykland, P. A. (1996). Options pricing in incomplete markets: an asymptotic approach. Technical Report no. 430, Department of Statistics, University of Chicago.

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

Nelson, D. B. (1992). Filtering and forecasting with misspecified ARCH models. I. Getting the right variance with the wrong model. Journal of Econometrics **52**, 61-90.

Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. Annals of Statistics **24**, 2399-2430.

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

Prakasa-Rao, B. L. S. (1988). Statistical inference from sampled data for stochastic processes. Contemp. Math. **80**, 249-284.

Rossi, P. E. (1996). Modeling Stock Market Volatility. Academic Press.

Strasser, H. (1985). Mathematical Theory of Statistics. Walter de Gruyter, Berlin.

Stroock, D. W. and Varadhan, S. R. S. (1979). Multidimensional Diffusion Processes. Springer.

Tusnády, G. (1977). A remark on the approximation of the sample distribution function in the multidimensional case. Period. Math. Hungar. **8**, 53-55.

Yoshida, N. (1992). Estimation for diffusion processes from discrete observations. J. Multivar. Anal. **41**, 220-242.

Department of Statistics, University of Connecticut, Storrs, CT 06269