



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Zu, Y. (2015). Nonparametric specification tests for stochastic volatility models based on volatility density. *Journal of Econometrics*, 187(1), pp. 323-344. doi: 10.1016/j.jeconom.2015.02.045

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/8090/>

**Link to published version:** <https://doi.org/10.1016/j.jeconom.2015.02.045>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

---

---

City Research Online:

<http://openaccess.city.ac.uk/>

[publications@city.ac.uk](mailto:publications@city.ac.uk)

---

# Nonparametric specification tests for stochastic volatility models based on volatility density\*

Yang Zu<sup>†</sup>

City University London

March 10, 2015

## Abstract

This paper develops a specification test for stochastic volatility models by comparing the nonparametric kernel deconvolution density estimator of an integrated volatility density with its parametric counterpart.  $L^2$  distance is used to measure the discrepancy. The asymptotic null distributions of the test statistics are established and the asymptotic power functions are computed. Through Monte Carlo simulations, the size and power properties of the test statistics are studied. The tests are applied to an empirical example.

JEL Classification: C58, C12, C14.

Keywords: nonparametric tests, kernel deconvolution estimator, stochastic volatility model.

## 1 Introduction

Consider the following continuous-time stochastic volatility model:

$$\begin{aligned}dX_t &= \sigma_t dB_t, \\d\sigma_t^2 &= b(\sigma_t^2)dt + a(\sigma_t^2)dW_t,\end{aligned}\tag{1}$$

---

\*I wish to thank the co-editor Yacine Aït-Sahalia, the associate editor and three anonymous referees for careful reading the manuscript and providing helpful comments and suggestions, which greatly improve the paper. This paper is revised from a chapter of my PhD thesis at the University of Amsterdam, I wish to thank my supervisor Peter Boswijk for helpful advice and encouragement. I also wish to thank Aurore Delaigle for the helpful email correspondence, Bert van Es and Dennis Kristensen for fruitful discussions and suggestions. Helpful comments from Peter Spreij, Cees Diks, Jan de Gooijer, Philippos Papadopoulos, and the session participants of the 10th World Congress of the Econometric Society are also gratefully acknowledged. All remaining errors are my own.

<sup>†</sup>Email: yang.zu@city.ac.uk. Department of Economics, City University London, Northampton Square, EC1V 0HB London, United Kingdom. Telephone: +44 (0)20 7040 8619. Fax: +44 (0)20 7040 8580.

where  $B$  and  $W$  are two independent standard Brownian motion processes.  $X$  is assumed to be observed discretely at  $t_i = i\Delta$ , for  $i = 0, 1, \dots, n$ ,  $\Delta$  is assumed to be fixed, while  $\sigma^2$  is assumed to be unobservable.

The model (1) is used to describe the evolution of asset prices in financial markets. Nonparametric estimation of continuous-time stochastic volatility models has been considered in Franke et al. (2003), Reno (2006), Reno (2008), Kanaya and Kristensen (2015) and Comte et al. (2010), among others.

Assume that in the stochastic volatility model (1), the true model is characterized by the functions  $\{b_0(\cdot), a_0(\cdot)\}$ , and let a parametrization of the model be:

$$\{b(\cdot; \theta), a(\cdot; \theta), \theta \in \Theta \subseteq \mathbb{R}^k\}.$$

This paper aims to study the problem of testing the null hypothesis

$$\begin{aligned} \mathcal{H}_0 : \{ \exists \theta_0 \in \Theta, b(\cdot; \theta_0) = b_0(\cdot), a(\cdot; \theta_0) = a_0(\cdot) \}, \\ \text{vs.} \\ \mathcal{H}_1 : \{ b(\cdot; \theta) \neq b_0(\cdot), a(\cdot; \theta) \neq a_0(\cdot), \forall \theta \in \Theta \}. \end{aligned} \tag{2}$$

If  $\sigma^2$  was observable (in discrete time), this problem would be reduced to a specification test problem for diffusion processes. For example, Aït-Sahalia (1996) proposed a specification test for diffusion processes by comparing the nonparametric kernel density estimate of the stationary density of the process with its parametric counterpart. His test statistic for the present model would be as follows:

$$T_n = \sum_{i=1}^n \left( \hat{\pi}(\sigma_i^2) - \pi(\sigma_i^2; \hat{\theta}_n) \right)^2, \tag{3}$$

where  $\hat{\pi}(x) = \sum_{i=1}^n K((x - \sigma_i^2)/h)/(nh)$  is the nonparametric kernel density estimator of the stationary volatility density,  $\pi(x; \theta)$  is the corresponding parametric density under the null hypothesis, and  $\hat{\theta}_n$  is an estimator of the parameters. The sum in (3) is over the grid of observations. However, the unobservability of  $\sigma^2$  in the stochastic volatility model (1) makes this test not applicable to the problem in this paper.

Although the volatility process is not observable in model (1), Van Es et al. (2003) notice that the volatility density of the model can still be estimated nonparametrically by a deconvolution kernel density estimator using the observed log return data because the discretized stochastic volatility model can be rewritten into a convolution model: let the log price  $X$  be observed at equally  $\Delta$ -spaced times  $t_0, \dots, t_n$ , and define the return

sequence:

$$\begin{aligned} y_i &:= \frac{1}{\Delta^{1/2}} (X_{t_i} - X_{t_{i-1}}) \\ &= \frac{1}{\Delta^{1/2}} \int_{t_{i-1}}^{t_i} \sigma_s dB_s \\ &\sim N\left(0, \frac{1}{\Delta} \int_{t_{i-1}}^{t_i} \sigma_s^2 ds\right) \end{aligned}$$

for  $i = 1, \dots, n$ , where  $\sim$  means “distributed as”. Define  $\eta_i := \int_{t_{i-1}}^{t_i} \sigma_s^2 ds / \Delta$  and use  $\varepsilon_i$ ,  $i = 1, \dots, n$  to denote independent and identically distributed standard normal variables. The above equation can be written as:

$$y_i = \eta_i^{1/2} \varepsilon_i, i = 1, \dots, n.$$

Squaring both sides and taking the logarithms,

$$\log y_i^2 = \log \eta_i + \log \varepsilon_i^2, i = 1, \dots, n, \quad (4)$$

such that the variable  $\log y_i^2$  is the convolution of  $\log \eta_i$  with  $\log \varepsilon_i^2$ , which has a completely known log chi-square distribution.

In statistics, model (4) is known as a measurement error model — the signal  $\log \eta_i$  is measured with a noise  $\log \varepsilon_i^2$ , and only  $\log y_i^2$  is observed. Recovering the density of the signal  $\log \eta_i$  from the observed  $\log y_i^2$ 's is called density deconvolution. Density deconvolution can be done in several ways, see Meister (2009) for a review. Denoting the density function of  $\log \eta_i$  as  $g(x)$ , it can be estimated by the following kernel deconvolution estimator:

$$\hat{g}(x) = \frac{1}{2\pi} \frac{1}{n} \sum_{j=1}^n \int_{-\infty}^{+\infty} \frac{\phi_K(th)}{\phi_k(t)} e^{-it(x - \log y_j^2)} dt,$$

where the definitions of  $\phi_K$  and  $\phi_k$  are left in Section 2.

Motivated by these studies, this paper proposes to test the hypothesis  $\mathcal{H}_0$  against  $\mathcal{H}_1$  by comparing the parametric estimate and the nonparametric kernel deconvolution estimate of the stationary density function of  $\log \eta_i$ . Notice that the asymptotic scheme considered in this paper is different from that of Van Es et al. (2003), where both the in-fill and long span asymptotic schemes are used. In this paper the sampling interval  $\Delta$  is assumed to be fixed and only a long span asymptotic scheme is used. This is consistent with the usual practice in financial econometrics that a stochastic volatility model is usually used to model returns sampled at daily or lower frequency. Under our assumption, the object of comparison is not exactly the stationary volatility density but the density function of  $\log \eta_i := \log \left( \int_{t_{i-1}}^{t_i} \sigma_s^2 ds / \Delta \right)$  – the *log integrated volatility density*. Denote  $g_0(x)$  as the true log integrated volatility density function and denote  $g(x; \theta)$

as the corresponding parametric density function under the null hypothesis. The test statistic could be formulated by calculating the  $L^2$  distance between  $\hat{g}(x)$  and  $g(x; \hat{\theta})$ ,

$$T_0 = \int_{\mathbb{R}} \left( \hat{g}(x) - g(x; \hat{\theta}) \right)^2 dx.$$

The actual testing problem studied in this paper can be expressed as follows:

$$\mathcal{H}'_0 : \{ \exists \theta_0 \in \Theta, g(\cdot; \theta_0) = g_0(\cdot) \} \quad \text{vs.} \quad \mathcal{H}'_1 : \{ g(\cdot; \theta) \neq g_0(\cdot), \forall \theta \in \Theta \}. \quad (5)$$

The reformulated testing problem (5) is *not equivalent* to the original problem in (2), which tests  $\mathcal{H}_0$  against  $\mathcal{H}_1$ . There are certain deviations in  $\mathcal{H}_1$  can not be detected by the reformulated testing problem. For example, this test will not be able to detect a volatility model with misspecified transitional density, but with a correctly specified marginal density. To address this problem, a possible extension to base the test on the bivariate volatility density is discussed in Section 5.4.

The idea of comparing parametric and nonparametric estimates for specification tests is not new. Bickel and Rosenblatt (1973) and Härdle and Mammen (1993), among others, made early contributions. Nonparametric tests with weakly dependent data were studied by Fan (1994), Fan and Ullah (1999), Fan and Li (1999), and Aït-Sahalia et al. (2001), among others. Nonparametric specification tests for diffusion processes were investigated by Aït-Sahalia (1996), Hong and Li (2005), Corradi and Swanson (2005), Chen et al. (2008), Aït-Sahalia et al. (2010) and Aït-Sahalia and Park (2012), among others. Specification tests involving a deconvolution kernel estimator is relatively new, see Butucea (2007) and Holzmann et al. (2007).

There are several existing tests for stochastic volatility models. Early contributions include the moment restrictions based tests (such as (Gallant et al., 1997)). Recent contributions include Corradi and Distaso (2006), who propose testing the specification of stochastic volatility models based on the moment information of realized volatility measures, and Corradi and Swanson (2011), who propose basing the test on the one-step predictive density of observed series. Zu and Boswijk (2009) propose tests based on the density function and distribution function of the observed returns.

It may seem natural to perform the test based on the observable return distributions and/or densities. However, the volatility density based test considered in this paper is useful in detecting certain types of local deviations to the null model. This is because the return density is a convolution of the volatility density with the log chi-square density. It can happen that two volatility densities at a certain  $L^2$  distance will become very close to each other after the convolution, and become very hard to distinguish by looking at the densities of the return densities. Example 2 of Holzmann et al. (2007) gives a mathematical example of such phenomenon: a sequence of nonlinear local alternatives is

specified to converge to the null model at a certain rate, but after taking a convolution operation with a Laplace density, the convergence rate could become arbitrarily faster. In this case, the testing problem becomes harder after the convolution and it would be more efficient to formulate the test in terms of the volatility density. In practice, since it is not known which kind of deviation the data generating process might have, the test proposed in this paper could be served as a useful complement to the return density/distribution based tests.

The methodology proposed in this paper only test statistical goodness-of-fit of a stochastic volatility model, it does not consider the model tractability and the related issues, which are important in determining a model specification in practical financial applications, say option pricing. Thus it is important that the test developed in this paper be interpreted appropriately and used with caution when invalidating a stochastic volatility model.

Model (1) does not allow for a drift term and a jump term, which is not consistent with empirical facts observed in realistic financial returns. Nevertheless, the model is still useful as it can be applied to the series which has been demeaned and has the jumps been filtered out. This issue will be discussed further in detail in Section 5.2. By considering model (1) the attention is focused on the (mis)specification of the volatility process.

This paper is organized as follows. Deconvolution kernel density estimation for the log integrated volatility density is discussed in Section 2. The parametric estimate of the log integrated volatility density is discussed in Section 3. The log integrated volatility density usually does not have an explicit formula. This section also discusses methods of approximating the log integrated volatility density. The test statistic, its asymptotic null distribution, and the asymptotic power of the test under fixed and local alternatives are studied in Section 4. A parametric bootstrap procedure is also proposed in this section to approximate the finite sample null distribution. Section 5 discusses possible extensions along different directions. Section 6 performs Monte Carlo simulations. The size and power properties of the tests are studied under various realistic scenarios. The test is applied to a real example in Section 7. Section 8 concludes the paper. Technical assumptions are stated in Appendix A. The proofs are collected in Appendix B. Technical lemmas used in the proofs are collected in Appendix C. An alternative method of approximating the parametric volatility density is discussed in Appendix D.

In this paper,  $\phi_g(t) = \int_{-\infty}^{\infty} e^{itx} g(x) dx$  is defined as the Fourier transform of function  $g(x)$ ; the inverse Fourier transform is defined as  $g(x) = (2\pi)^{-1} \int_{-\infty}^{+\infty} e^{-itx} \phi_g(x) dt$ . An integral  $\int$  with no upper and lower limit implicitly means it is an integral over the real line.  $\xrightarrow{d}$  is used to denote convergence in distribution and  $\sim$  to denote “distributed as”.  $\|f\|_p := (\int |f(x)|^p dx)^{1/p}$ ,  $p > 0$  denotes the  $p$ th norm of function  $f$ , provided that it exists. By definition  $\|f\|_p$  is always nonnegative, and  $\|f\|_p = 0$  only when  $f(x) = 0$  almost everywhere on the real line. Assume further that all the density functions considered

in this paper have strictly positive norms.  $g * k(x) = \int g(x - y)k(y)dy$  denotes the convolution of  $g(\cdot)$  and  $k(\cdot)$ . The *volatility process* in this paper is always used to refer to the  $\sigma^2$  process, not the  $\sigma$  process.

## 2 Nonparametric log integrated volatility density

Assume that the density functions of  $\log y_i^2$ ,  $\log \eta_i$  and  $\log \varepsilon_i^2$  exist, denoted as  $f(\cdot)$ ,  $g(\cdot)$ , and  $k(\cdot)$ , respectively. Equation (4) implies the following convolution relationship:

$$f(x) = g * k(x).$$

When  $Y_i := \log y_i^2$ ,  $i = 1, \dots, n$  are observable and  $k(x)$  is fully known, the density function  $g(x)$  can be estimated by the deconvolution kernel density estimator:

$$\hat{g}(x) = \frac{1}{2\pi} \frac{1}{n} \sum_{j=1}^n \int_{-\infty}^{+\infty} \frac{\phi_K(th)}{\phi_k(t)} e^{-it(x-Y_j)} dt,$$

where  $\phi_K$  is the Fourier transform of a kernel function  $K$  and  $\phi_k(t)$  is the characteristic function corresponding to  $k(x)$ . The kernel deconvolution estimator was first proposed for the measurement error model by Carroll and Hall (1988) and Stefanski and Carroll (1990).

If

$$\nu_h(x) := \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\phi_K(t)}{\phi_k(t/h)} e^{-itx} dt,$$

is defined to be the deconvolution kernel function, the estimator can be written in a kernel form:

$$\hat{g}(x) = \frac{1}{nh} \sum_{j=1}^n \nu_h \left( \frac{x - Y_j}{h} \right). \quad (6)$$

It is known that  $\nu_h(x)$  is a real-valued function when  $\phi_K(t)$  is even and real (see, e.g., Stefanski and Carroll (1990)).

The error term  $\log \varepsilon_i^2$  follows a log chi-square distribution. Its probabilistic properties can be found in e.g. Van Es et al. (2005) and Comte (2004). In particular, the characteristic function of the log chi-square distribution is  $\phi_k(t) = 2^{it} \Gamma(1/2 + it) / \sqrt{\pi}$ ; the tail decay rate of its modulus function is described by  $|\phi_k(t)| = \sqrt{2} \exp(-\pi|t|/2)(1 + O(1/|t|))$  as  $|t| \rightarrow \infty$ . According to Fan (1991), this error belongs to the so-called super-smooth errors because the tail of the modulus function decays exponentially fast.

The kernel function considered first in this paper is the so-called sinc kernel:

$$K(x) = \begin{cases} \sin(x) / (\pi x) & \text{if } x \neq 0, \\ 1 & \text{if } x = 0, \end{cases} \quad (7)$$

The sinc kernel is also called Fourier Integral kernel. The usage of this kernel in density deconvolution dates back to Stefanski and Carroll (1990). Its Fourier transform is  $\phi_K(t) = I\{|t| \leq 1\}$ , where  $I\{\cdot\}$  is an indicator function. The simplicity of  $\phi_K(t)$  aids the computation and thus the sinc kernel is favoured in theoretical literature (e.g. Butucea (2007), among others). It is a so-called “infinite order” kernel, see Delaigle and Hall (2010) Section 3.1 for a detailed discussion about this property in general and in the context of kernel deconvolution.

Although  $\int K(x) = 1$ , it is not a proper probability density because it takes negative values. A main drawback of the sinc kernel is its numerical stability; it often causes unwanted oscillations in the estimator (see Delaigle and Hall (2010) and Meister (2009) for the discussions). The extensions to a more numerically stable kernel is discussed in Section 5.1.

## 3 Parametric log integrated volatility density

### 3.1 Parametric estimation and approximating the parametric volatility density

Estimating stochastic volatility models was an active research area in the past decades. This paper does not rely on any specific parametric estimation method, it is only assumed that the parametric estimator is  $\sqrt{n}$ -consistent under the null hypothesis and that the parametrization is smooth. These requirements are easily satisfied by popular methods, such as the Efficient Method of Moments (EMM) by Gallant and Tauchen (1996) and the Generalized Method of Moments (GMM) by Meddahi (2002).

The density function  $g(x; \theta)$  usually does not have a closed-form expression in terms of the functions  $a(\cdot; \theta)$  and  $b(\cdot; \theta)$  because of the fixed interval sampling scheme used in this paper. Two methods for approximating the function  $g(x; \hat{\theta})$  are proposed. The first is based on numerical simulation of the volatility process and is discussed in this section. Its precision can be made arbitrarily high by simulating the path sufficiently fine and long. In Appendix D a second heuristic method is discussed: the stationary density, which is known in closed form, is used as an approximation of the integrated volatility density; however, the magnitude of the approximation error is unknown.

#### 3.1.1 Approximation by simulation

Given the estimated parameter values  $\hat{\theta}$ , one can simulate the estimated model using standard numerical simulation methods for stochastic differential equations, such as the Euler Scheme and the Milestein Scheme, both methods are convergent in either the weak or strong criteria, see e.g. Chapter 6 of Glasserman (2004).

To be specific, given an estimate  $\hat{\theta}$ , the parameterization  $b(\cdot; \theta)$  and  $a(\cdot; \theta)$  and the step size  $\Delta$ , taking  $\delta = \Delta/N$  as a finer step, one can simulate  $M$  consecutive blocks of  $N$  observations with step length  $\delta$ , making  $M \times N$  observations of  $\sigma_1^2, \dots, \sigma_{MN}^2$  over *one path* of the model. Then, calculating the average in each of the  $M$  blocks and taking the logarithm, one can produce a sequence of approximated realizations:  $\log \left( \int_{t_{i-1}}^{t_i} \sigma_s^2 ds / \Delta \right)$ ,  $i = 1, \dots, M$ . In this simulated random world, the standard nonparametric kernel density estimation methodology can be applied to “estimate” the density function of  $\log \left( \int_{t_{i-1}}^{t_i} \sigma_s^2 ds / \Delta \right)$ . Denote the resulting simulated sample by  $Y_1^*, \dots, Y_M^*$ ; using a classical kernel density estimator for the simulated volatility observations (not the deconvolution kernel density estimator),  $g(x; \hat{\theta})$  can be approximated by  $g^s(x; \hat{\theta})$  as

$$g^s(x; \hat{\theta}) = \frac{1}{M h_M} \sum_{i=1}^M K^* \left( \frac{x - Y_i^*}{h_M} \right),$$

where  $K^*(\cdot)$  is a kernel function and  $h_M$  is the bandwidth parameter. Notice that here the kernel function and the bandwidth parameter are different with those used for kernel deconvolution estimator in (6). The choices of the kernel function and the bandwidth are also less important in this simulation world than in real world, because the number of simulated observations  $M$  can be chosen to be large.

The standard consistency results for the classical kernel density estimators and convergence theorems for the Euler scheme (or Milestein scheme) simulation imply that when  $M \rightarrow \infty$ ,  $h_M \rightarrow 0$  and  $N \rightarrow \infty$ ,  $g^s(x; \hat{\theta}) \xrightarrow{p} g(x; \hat{\theta})$  pointwisely for all  $x \in \mathbb{R}$ . The convergence in probability is understood as in the probability space of numerical simulations. For technical conditions on the kernel function  $K^*$ , bandwidth  $h_M$  and the consistency results of the classical kernel density estimator, see e.g., Section 2.6.2 of Pagan and Ullah (1999). For the convergence results of the Euler scheme (or Milestein scheme) simulation, refer to Kloeden and Platen (1992). The accuracy of this approximation is determined by the number of blocks  $M$  and the number  $N$  in each block. Because here  $M$  and  $N$  need not depend on the sample size  $n$ , they can be chosen as very large to make the approximation error arbitrarily small.

## 4 Test statistics and asymptotic properties

### 4.1 Test statistic and asymptotic null distribution

The test statistic

$$T_0 = \int_{\mathbb{R}} \left( \hat{g}(x) - g(x; \hat{\theta}) \right)^2 dx,$$

calculates the  $L^2$  distance between the parametric and the nonparametric density estimates of the volatility process. Its asymptotic distribution is given in the next Theorem.

**Theorem 1** *Assume Conditions (SV0)–(SV5) and Assumptions B1, B2, and C1 in Appendix A, such that  $\{Y_i\}$  is a stationary,  $\beta$ -mixing sequence with coefficients  $\beta(k) = e^{-\lambda k}$  for some  $\lambda > 0$ ; when the kernel function considered is the sinc kernel defined in (7); then under  $\mathcal{H}_0$ , when  $n \rightarrow \infty$ ,  $h \rightarrow 0$ ,  $\exp(\pi/h)/n \rightarrow 0$  and  $\exp(\pi/h)/(nh^2) \rightarrow \infty$ ,*

$$\frac{1}{\sigma_{n,1}} (T_0 - \mu_{n,1}) \xrightarrow{d} N(0, 1),$$

where  $\mu_{n,1} = \exp(\pi/h)/(2\pi^2n)$ ,  $\sigma_{n,1} = \exp(\pi/h)\|f\|_2/(2\pi n)$ .

Notice that  $\|f\|_2$  is the  $L^2$  norm of the density of the observed return  $Y_i$ 's. Let  $\widehat{\|f\|_2}$  be a consistent estimator for  $\|f\|_2$ , and define  $\hat{\sigma}_{n,1} = \exp(\pi/h)\widehat{\|f\|_2}/(2\pi n)$ , it follows that

$$\frac{T_0 - \mu_{n,1}}{\hat{\sigma}_{n,1}} \xrightarrow{d} N(0, 1).$$

Theorem 1 does not assume a specific consistent estimator for  $\|f\|_2$ . Estimating the  $L^2$  norm of a density using kernel method with direct observations is classical in the nonparametric literature, several consistent estimators have been exist. For example, let  $\hat{f}(x)$  be the classical kernel density estimator for  $f(x)$ , it is known that  $(1/n)\sum_{i=1}^n \hat{f}(Y_i)$  can estimate  $\|f\|_2$  consistently, see the result in, e.g. Fan and Ullah (1999) Theorem 4.1 for the stationary  $\beta$ -mixing data case. See Hall and Marron (1987) for other possible estimators.

The assumptions  $\exp(\pi/h)/n \rightarrow 0$  and  $\exp(\pi/h)/(nh^2) \rightarrow \infty$  essentially restrict the bandwidth  $h$  in a narrow band. To see this, notice that  $h$  should converge to 0 slower than  $\pi/\log n$  because of the assumption  $\exp(\pi/h)/n \rightarrow 0$ , but  $\pi/(\log n^\lambda)$  for any  $\lambda \in (0, 1)$  will be too slow as this will fail the assumption  $\exp(\pi/h)/(nh^2) \rightarrow \infty$ . To give a specific example, the bandwidth  $h = \pi/\log(n/\log n)$  satisfies both assumptions, and it will deliver a very slow convergence rate  $(\log n)^{-1}$  for the test statistic.

The convergence rate of the test statistic  $T_0$  is faster when  $h \rightarrow 0$  at a slower rate. However, this is restricted by the assumption  $\exp(\pi/h)/(nh^2) \rightarrow \infty$ , which is necessary for the integrated squared bias to be dominated as in the proof of Theorem 1. From that proof, it is clear that the following bias-corrected test statistic

$$T_1 = \int_{\mathbb{R}} \left( \hat{g}(x) - K_h * g(x; \hat{\theta}) \right)^2 dx,$$

where  $K_h(x) = K(x/h)/h$ , does not have the integrated squared bias term by definition, thus does not need the assumption  $\exp(\pi/h)/(nh^2) \rightarrow \infty$  any more. In the literature of nonparametric specification test, using bias-corrections in the definition of test statistics has already been seen in Härdle and Mammen (1993) and Fan (1994), among others. The results for the bias-corrected statistic is summarized in the next Theorem.

**Theorem 2** *Assume Conditions (SV0)–(SV5) and Assumptions B1, B2, and C1 in Appendix A, such that  $\{Y_i\}$  is a stationary,  $\beta$ -mixing sequence with coefficients  $\beta(k) = e^{-\lambda k}$  for some  $\lambda > 0$ ; when the kernel function considered is the sinc kernel defined in (7); then under  $\mathcal{H}_0$ , when  $n \rightarrow \infty$ ,  $h \rightarrow 0$  and  $\exp(\pi/h)/n \rightarrow 0$ ,*

$$\frac{1}{\sigma_{n,1}} (T_1 - \mu_{n,1}) \xrightarrow{d} N(0, 1),$$

where  $\mu_{n,1} = \exp(\pi/h)/(2\pi^2n)$ ,  $\sigma_{n,1} = \exp(\pi/h)\|f\|_2/(2\pi n)$ .

Use the same notation  $\hat{\sigma}_{n,1}$  as in Theorem 1. Define  $T_2 := (T_1 - \mu_{n,1})/(\hat{\sigma}_{n,1})$ , we have that  $T_2 \xrightarrow{d} N(0, 1)$ .

The bandwidth condition imposed in Theorem 2 is actually equivalent to that needed for the deconvolution estimator to be pointwise consistent (see e.g. Stefanski and Carroll (1990) Theorem 2.1). The convergence rates of the test statistics  $T_1$  and  $T_2$  are both  $n^{-1} \exp(\pi h^{-1})$ . Since the lower bound for  $h \rightarrow 0$  has been removed, one could choose  $h$  to converge to 0 slower to achieve better rate of convergence. For example, when  $h = \pi/(\log n^\lambda)$  for any  $\lambda \in (0, 1)$ , the rate becomes  $n^{-1+\lambda}$ , which could be any rate slower than  $n^{-1}$ . This is similar to the error-free nonparametric density-based tests (e.g. Fan (1994) and Gao and King (2004)), where the rate is  $n^{-1}h^{-1/2}$ , which could also be any rate slower than  $n^{-1}$ .

Next, only the biased-corrected tests  $T_1$  and  $T_2$  are considered for their power properties, because they require less restrictive bandwidth assumptions.

## 4.2 Asymptotic power properties

For the asymptotic power of the test statistics, a fixed alternative is first considered:

$$\mathcal{H}'_1 : \{g(x) = g_1(x) \neq g(x; \theta), \forall \theta \in \Theta\}.$$

**Theorem 3** *Assume Conditions (SV0)–(SV5) and Assumptions B1a, B2, and C1 in Appendix A; when the kernel function considered is the sinc kernel defined in (7); let  $\alpha \in (0, 1)$  be a level of significance, and  $Z_{1-\alpha}$  be the  $1 - \alpha$  quantile of the standard normal distribution. Then under  $\mathcal{H}'_1$ ,*

$$P\left(\frac{1}{\sigma_{n,1}} (T_1 - \mu_{n,1}) > Z_{1-\alpha}\right) \rightarrow 1,$$

and

$$P(T_2 > Z_{1-\alpha}) \rightarrow 1,$$

as  $n \rightarrow \infty$ ,  $h \rightarrow 0$  and  $\exp(\pi/h)/n \rightarrow 0$ .

As with most nonparametric tests, both tests are consistent. That is, they can detect any fixed deviation to the true model as long as the sample size is sufficiently large.

The asymptotic power properties of the tests under linear local alternatives are considered next. Let

$$\mathcal{H}'_{1n} : \{g_n(x) = g(x, \theta_0) + \gamma_n d(x)\},$$

be a sequence of local alternatives, where  $\gamma_n$  and  $d(x)$  have to be chosen such that  $g_n$  is a density, with  $\gamma_n \rightarrow 0$  as  $n \rightarrow \infty$ ,  $\int d(x)dx = 0$ ,  $\int d^2(x)dx < \infty$  and  $d(x)$  is bounded. This sequence of linear local alternative models is also called regular or Pitman alternative, it converges to the null density at a rate of  $\gamma_n$ .

**Theorem 4** *Assume Conditions (SV0)–(SV5) and Assumptions B1a, B2, and C1 in Appendix A; when the kernel function considered is the sinc kernel defined in (7); let  $\alpha \in (0, 1)$  be a level of significance, and  $Z_{1-\alpha}$  be the  $1 - \alpha$  quantile of the standard normal distribution. Then, under  $\mathcal{H}'_{1n}$ , for  $\gamma_n^2 = \exp(\pi/h)/n$ ,*

$$\frac{1}{\sigma_{n,1}} (T_1 - \mu_{n,1}) \xrightarrow{d} N\left(\frac{2\pi}{\|f\|_2} \int d^2(x)dx, 1\right),$$

and

$$T_2 \xrightarrow{d} N\left(\frac{2\pi}{\|f\|_2} \int d^2(x)dx, 1\right).$$

It follows that

$$P\left(\frac{1}{\sigma_{n,1}} (T_1 - \mu_{n,1}) > Z_{1-\alpha}\right) \rightarrow \Phi\left(Z_{1-\alpha} - \frac{2\pi}{\|f\|_2} \int d^2(x)dx\right),$$

and

$$P(T_2 > Z_{1-\alpha}) \rightarrow \Phi\left(Z_{1-\alpha} - \frac{2\pi}{\|f\|_2} \int d^2(x)dx\right),$$

as  $n \rightarrow \infty$ ,  $h \rightarrow 0$  and  $\exp(\pi/h)/n \rightarrow 0$ .

The tests  $T_1$  and  $T_2$  have power against local alternatives that converge at a rate  $n^{-1/2} \exp(\pi/(2h))$ . If  $h$  is chosen to converge to 0 as slow as possible, the rate of local alternatives can be detected can be any rate slower than  $n^{-1/2}$ . For example, when  $h = \pi/(\log n^\lambda)$  for any  $\lambda \in (0, 1)$ , the rate becomes  $n^{-1/2+\lambda/2}$ , which could be any rate slower than  $n^{-1/2}$ . The tests are less powerful than the for the local alternatives  $\mathcal{H}'_{1n}$  than the Kolmogorov-Smirnov test, which has the rate  $n^{-1/2}$ . For the error-free nonparametric density-based tests, the rate of local alternatives detectable is  $n^{-1/2}h^{-1/4}$  (see e.g. Fan (1994) page 329 for a discussion), which also could be any rate slower than  $n^{-1/2}$ .

### 4.3 Bootstrap null distribution

$T_1$  is not admissible as an asymptotic test because there are unknown quantities in the asymptotic variance of the limiting distribution. As for the test  $T_2$ , in finite sample its asymptotic approximation is usually poor. This happens in most nonparametric test statistics (see e.g. Fan (1995) for a discussion), and a bootstrap method is usually used to approximate the null distribution.

The null distribution of the test statistic need be approximated under both the null and the alternative hypothesis of the test. In particular, under the alternative hypothesis, the null distribution of the statistic is actually its distribution under the pseudo-true model. In this paper a parametric bootstrap procedure is considered. Roughly, parametric bootstrap refers to the method of resampling from a parametrically estimated model (see Section 6.5 of Efron and Tibshirani (1994)). In parametric bootstrap, bootstrap samples are generated from the estimated null model, which mimics the null model under the null hypothesis and mimics the pseudo-true model under the alternative hypothesis. These ensure that the null distribution is approximated both under the null and under the alternative hypothesis. Stationarity and the dependence structure of the bootstrap sample are ensured by checking the assumptions (SV0) to (SV5) in the Appendix A of the estimated null model.

Parametric bootstrap has been used in various nonparametric testing problem, though sometime it is called differently. For the i.i.d. data case, see Fan (1995), Andrews (1997). For dependent data case, see Aït-Sahalia et al. (2009), the *autoregression bootstrap* in Franke et al. (2002) and the discussion for the *recursive simulation scheme* in Gao and Gijbels (2008) after their description of how to simulate the null distribution.

The parametric bootstrap approximation procedure for the null distribution of  $T_i, i = 1, 2$  is as follows:

**Step 1** Given a parametric estimate  $\hat{\theta}$  (which has to satisfy the stationarity condition of the null model), step size  $\Delta$ , simulate  $n$  (original sample size) discretely observed  $\Delta$ -returns, which is called one bootstrap sample. This step has to be done over a fine grid, as in Section 3.1.1.

**Step 2** With this bootstrap sample, compute the test statistic  $T_i$ , and call it  $T_i^*$ .

**Step 3** Repeat steps 1 and 2  $B$  times to obtain  $B$  realizations of the bootstrapped test statistic  $T_i^{*1}, \dots, T_i^{*B}$  for the statistic  $T_i$ .

When  $B$  is big, the empirical distribution of  $T_i^{*1}, \dots, T_i^{*B}$  approximates the finite sample null distribution.

The asymptotic validity of parametric bootstrap has been shown for various non-parametric test statistics with direct observations (e.g. Fan (1995), Andrews (1997) and

Neumann and Paparoditis (2000)), for both density based and distribution based tests, and for both the  $L^2$  type tests and the corresponding model-free tests. Asymptotic theory is lacking for the parametric bootstrap procedure proposed in this section. It is conjectured that it could be developed with the strategy used in the aforementioned literature. In the absence of such theory, I examine the bootstrap approach by extensive Monte Carlo experiment.

**Remark** Another type of bootstrap, the so-called block bootstrap, could also be considered in the current model. Corradi and Swanson (2005) have used the block bootstrap to approximate the null distribution in testing the specification of diffusion processes. The mechanism of block bootstrap is different with the parametric bootstrap, because the block bootstrap mimics the data generating process always: under the null hypothesis, the block bootstrap mimics the null model; and in particular under the alternative hypothesis, it mimics the alternative model. For this reason, the block bootstrap distribution of the test statistic under the alternative hypothesis is no longer the null distribution one needs to approximate. This is why Corradi and Swanson (2005) has to use a re-centered statistic (equation (11) and (12) in their paper), instead of the original test statistic (equation (5) in their paper), with the block bootstrap sample. It is interesting to investigate the possibility of using a block bootstrap procedure for the current statistic in future research and it is conjectured that a similar re-centering strategy would be necessary.

## 5 Extensions and related issues

### 5.1 Other kernel function

The sinc kernel is used in presenting the main theorems because it simplifies the expressions for the asymptotic mean and asymptotic variance. For numerical implementation, it is known (e.g. Delaigle and Gijbels (2007)) that the following kernel:

$$K_1(x) = \frac{48x(x^2 - 15) \cos x - 144(2x^2 - 5) \sin x}{\pi x^7} \quad (8)$$

is more stable. The Fourier transform of the kernel is:

$$\phi_{K_1}(u) = (1 - u^2)^3 I(|u| \leq 1)$$

With this kernel, the central limit theorem can be derived analogously. The derivation does not differ so much from that of Theorem 1 and Theorem 2, other than in most of the places the integrals becomes tedious to evaluate due to the extra  $(1 - u^2)^3$  factor in the integrand. However, with the help of a computer algebra system, e.g. the Wolfram

Mathematica as being used in this paper, the analytical evaluation of integrals can be done in a fast and accurate way. The results are summarized in the next proposition.

**Proposition 1** *Under the same conditions as in Theorem 2, other than using the kernel function  $K_1$  as defined in (8), and the bandwidth satisfies  $\exp(\pi/h)h^6/n \rightarrow 0$  as  $n \rightarrow \infty$  and  $h \rightarrow 0$ , it holds that*

$$\frac{1}{\sigma_{n,2}}(T_1 - \mu_{n,2}) \xrightarrow{d} N(0, 1),$$

where  $\mu_{n,2} = C_1 \exp(\pi/h)h^6/n$ ,  $\sigma_{n,2} = C_2 \exp(\pi/h)h^6\|f\|_2/n$ , and  $C_1 = 23040/\pi^8 \approx 2.42819$ ,  $C_2 = 720\sqrt{310}/\pi^7 \approx 3.62318$ .

Still, let  $\widehat{\|f\|_2}$  be a consistent estimator for  $\|f\|_2$ , define  $\hat{\sigma}_{n,2} := C_2 \exp(\pi/h)h^6\widehat{\|f\|_2}/n$ , then

$$T_2 := \frac{1}{\hat{\sigma}_{n,2}}(T_1 - \mu_{n,2}) \xrightarrow{d} N(0, 1).$$

It is interesting to notice that the convergence rate of the test statistic has improved to  $n^{-1} \exp(\pi/h)h^6$  with a factor of  $h^6$  as compared to the result of Theorem 2. The kernel  $K_1$  is used throughout in the Monte Carlo experiment and the empirical application.

## 5.2 More general specifications

The approach developed in this paper can be adapted to the following model:

$$\begin{aligned} dX_t &= \sigma_t dB_t, \\ \sigma_t &= F(V_t), \\ dV_t &= b(V_t)dt + a(V_t)dW_t, \end{aligned}$$

where the volatility process is specified as a function  $F(\cdot)$  of a latent factor  $V_t$ . The  $V_t$  process does not need to be positive always. For example, both positive and negative values can be taken for  $V_t = \ln(\sigma_t^2)$  in the continuous-time Log SARV model, where the positivity of  $\sigma_t$  is ensured by using a  $F(x) = \exp(x)$ .

Tests could also be formulated based on the log integrated volatility density, but the approximation for the parametric part now involves an extra function. On the technical side, the domain  $(0, +\infty)$  for the volatility process in assumptions (SV0)-(SV5) of Appendix A need be adapted accordingly to the domain of  $V_t$ . For example, the domain of  $V_t$  is  $(-\infty, +\infty)$  for the log SARV model.

Another direction to generalize the model (1) is to consider a more general log price model:

$$dX_t = \mu_t dt + \sigma_t dB_t + dJ_t,$$

where  $\mu_t$  is a drift process, and  $J_t$  is a jump process. It would be challenging to adapt the approach in this paper to study the misspecification in  $\mu_t$  and  $J_t$  jointly with the misspeci-

fication in the volatility process, because of the possible interaction of the misspecification in these components. On the other hand, if one is convinced that the drift part and the jump part are correctly specified, for practical purpose the tests developed in this paper could be applied to the residuals of the estimated model with these components removed, to detect the possible misspecification in the volatility process.

### 5.3 Discrete-time models and more general error distributions

The approach can be adapted to discrete-time stochastic volatility models straightforwardly. Under general assumptions, the kernel deconvolution estimator can be used to estimate the volatility density from the following nonparametric stochastic volatility model:

$$y_t = \sigma_t \varepsilon_t,$$

with  $\varepsilon_t$  to be either a standard normal distribution or standardized  $t$  distribution, as long as the volatility process is assumed to be stationary. One could then apply the strategy developed in this paper to test any parametric structure imposed on the volatility process.

To give an example, consider the classical discrete-time Stochastic AutoRegressive Volatility (SARV) model:

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, \\ \log \sigma_t^2 &= \omega + \gamma \log \sigma_{t-1}^2 + \sigma_\eta \eta_t, \\ (\varepsilon_t, \eta_t) &\sim i.i.d.N(0, I_2), \end{aligned}$$

where  $y_t$  is the log return. When  $\gamma < 1$  and the volatility process  $\log \sigma_t^2$  is initiated from its stationary distribution  $N(\omega/(1-\gamma), \sigma_\eta^2/(1-\gamma^2))$ , the volatility process is strictly stationary. The process is also  $\beta$ -mixing with exponentially decaying coefficients (see Pham and Tran (1985)) and thus ergodic. To allow for fatter tails in the unconditional distribution, the  $\varepsilon_t$  can be specified as distributions with fatter tails, such as the student  $t$  distribution.

Using the same strategy of constructing the test statistic as in Section 4.1, one can compare the kernel deconvolution volatility density estimate with its parametric counterpart implied by the discrete-time SARV model. The log volatility density of the discrete-time SARV model is known to be normal and needs no approximation. However, the stationary densities of general discrete time volatility models are usually difficult to obtain.

More general specification for the distribution of the error term  $\varepsilon_t$  could also be considered. For example, Lambert and Laurent (2001) consider a standardized skew-student- $t$  distribution to account for the skewness observed in certain financial returns. Studying the asymptotic theory of the test under different error distributions would be

an interesting problem. In the general measurement error models, Holzmann et al. (2007) have studied the nonparametric testing with indirect observations when the errors are ordinary smooth distributions. In a closely related work, Butucea (2004) derives the asymptotic theory of the Integrated Squared Error (ISE) of the deconvolution estimator under various specifications of the signal distribution and error distribution. Adapting these results to the context of stochastic volatility models would be an interesting yet challenging task, I thus leave this for future research.

## 5.4 Test based on bivariate volatility density

The test proposed in this paper is based on the marginal volatility density. As discussed in the introduction section, possible misspecification in the transitional density of the volatility process may not be detectable by this method, if the marginal density is correctly specified. This problem could be partially solved by defining the test based on the bivariate density of the volatility process. The nonparametric deconvolution estimator for the bivariate density is a straightforward extension of the univariate case, which was discussed in Van Es and Spreij (2011). The convergence rate of the bivariate estimator will be slower than the univariate case because of the well-known “curse of dimensionality”, but Van Es and Spreij (2011) have given a simulated numerical example which shows the well practical performance of bivariate deconvolution estimator. Then the parametric bivariate density can be obtained analogously with the numerical simulations and the test statistics could be formulated. The bivariate density contains information about both the marginal density and the transitional density, and should be able to detect the misspecification in the dynamics of the volatility process when used in combination with the marginal density based test.

## 6 Monte Carlo simulations

The finite sample size and power properties of the volatility based tests proposed in this paper are studied and compared with the three return based tests  $T_3$ ,  $T_4$  and  $T_5$  from Zu and Boswijk (2009). The test  $T_3$  compares the nonparametric estimate and the parametric estimate of the density function of the return sequence:

$$T_3 = \int \left( \hat{q}(x) - L_b * q(x; \hat{\theta}) \right)^2 dx,$$

where  $\hat{q}(x)$  and  $q(x; \hat{\theta})$  are the classical kernel density estimator and the parametric estimator of the observed return density, respectively;  $L_b(x) = L(x/b)/b$ , and  $L(x)$  is the

kernel function and  $b$  is the bandwidth parameter. Define

$$T_4 := \frac{nh^{1/2} (T_3 - (nh)^{-1} \int L^2(u)du)}{\hat{\sigma}},$$

where

$$\hat{\sigma}^2 = \frac{2}{n} \sum_{i=1}^n \hat{q}(y_i) \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} L(u)L(u+v)du \right]^2 dv.$$

$T_4$  is a studentized version of  $T_3$  and it is asymptotically distributed as  $N(0, 1)$ , thus is model-free.  $T_5$  is the Cramer-von Mises type test for the return distribution function,

$$T_5 = n \int \left( \hat{Q}(x) - Q(x; \hat{\theta}) \right)^2 dQ(x; \hat{\theta}),$$

where  $\hat{Q}(x)$  and  $Q(x; \hat{\theta})$  are the empirical distribution function and the parametrically estimated distribution function of the observed return, respectively.

The parametric bootstrap method described in Section 4.3 is used to determine the null distributions of the test statistics. The bandwidth of the tests are selected using the Cross-Validation (CV) method of Stefanski and Carroll (1990), see also Delaigle and Gijbels (2004) for a discussion. The parametric estimator used in this section is the GMM estimator of Meddahi (2002). A total of 12 unconditional moments were used in defining the GMM objective function. The GMM estimator is less efficient than the simulated likelihood based method as it only uses moment information, but it is also less computational demanding and thus more practical for the Monte Carlo experiment. The Milestein scheme is used to simulate from all the stochastic differential equations. The kernel  $K_1$  is used in all the volatility based test statistics. For the return density based test, the bandwidth is selected using the Cross-Validation method for classical kernel density estimator (see e.g. Wasserman (2004))<sup>1</sup>; the kernel  $L$  used is the Gaussian kernel. Since calculating the CV bandwidth is too computational intensive to apply to each simulated path, I simulate 120 paths of the model, calculate the corresponding CV bandwidths, and use the average of as our CV bandwidth in all the simulations. It should be emphasized that the data-driven CV method used here is just to avoid *ad hoc* selection of bandwidth, it does not have any optimality implications in terms of test performance.

---

<sup>1</sup>Exceptions are in the power simulation of the tests  $T_3$  and  $T_4$ , where the CV bandwidth seems to give bad power so I revert to use the Normal Reference Rule to select the bandwidth.

## 6.1 Size of the test

1000 sample paths of 5 years, 10 years and 15 years of daily observations are simulated from a continuous time log SARV model:

$$\begin{aligned} dY_t &= \sigma_t dW_t, \\ d \ln \sigma_t^2 &= \alpha(\beta - \ln \sigma_t^2)dt + \gamma dB_t, \end{aligned} \tag{9}$$

where the two Brownian motions  $W$  and  $B$  are independent. The unit interval is taken to be 1 year, which is assumed to have 252 trading days, so  $\Delta = 1/252$  is taken for daily observations. The parameter values  $\alpha = 10$ ,  $\beta = -3$ , and  $\gamma = 3$  is considered.<sup>2</sup>

With the given parameters and sample sizes, the test statistics  $T_1$  and  $T_2$  are simulated 1000 times. The distributions of the simulated test statistics can be obtained using the kernel density estimate and are considered the true distribution (except the Monte Carlo errors). For each of the 1000 paths, 5 bootstrap samples are obtained and their resulting test statistics are computed. Aggregating them together across 1000 samples yields 5000 bootstrap statistics. Their sampling distributions, computed via the kernel density estimate, are considered as the distributions of the tests of the bootstrap method.

To get a visual illustration of the performance of the parametric bootstrap procedure, plots of the null distribution and the bootstrap null distribution for the test statistics  $T_1$  and  $T_2$  under different sample sizes are given in Figure 1. It is observed that the bootstrap procedure seems to provide rather good approximation to the null distribution across all the sample sizes. It is also noticed that the null distribution itself seems to be better approximated by a normal distribution when the sample size is large.

[Figure 1 about here.]

The rejection rates of the tests are computed under different nominal levels and across different sample sizes. These are compared with the size properties of the test  $T_3$ ,  $T_4$  and  $T_5$ . The results are given in Table 1. It is observed that the  $L^2$  distance based test  $T_1$  is in general mildly oversized and the accuracy is acceptable. The corresponding model-free test  $T_2$  seems to always has a better size than  $T_1$ . This is in line with the theory as  $T_2$  is a pivotal test statistic and would benefit more from bootstrap. The return based tests seem to have the similar oversizing problem, though the model-free test  $T_4$  seems to suffer less than the original test  $T_3$ , and the the Cramer von Mises test  $T_5$  seem to most seriously oversized.

[Table 1 about here.]

---

<sup>2</sup>These parameter values are similar to those implied from the discrete-time SARV model estimated in Van der Sluis (1997) and Andersen et al. (2002). The sensitivity of the size performance to the values of the mean reversion speed parameter is also studied. The results are similar to those presented here so they are omitted to save space.

## 6.2 Power of the tests

The power performance of the tests is studied still under the three sample sizes of 5 years, 10 years and 15 years of daily data. The continuous-time Log SARV model (9) is still taken as the null model. The power functions of the five test statistics under two families of alternative models are evaluated with Monte Carlo simulation. In the first family of alternative models, the drift function of the volatility process deviates from the Log SARV model. In the second family of models, the diffusion function deviates from the Log SARV model. Power functions of all the tests are calculated at 1%, 5% and 10% nominal level. To save space, only the power functions at 5% level are reported in this section.

### 6.2.1 Misspecification in the drift function

The power functions of the test statistics are evaluated under the following sequence of alternative models,

$$d \ln \sigma_t^2 = \{(1 - \tau)(\alpha(\beta - \ln \sigma_t^2) + \tau\mu(\ln \sigma_t^2))\}dt + \gamma dW_t, \quad (10)$$

for  $\tau = 0, 0.1, \dots, 1$ , where  $\mu(x) = a \exp(-x) - b$  with  $a = 0.120$ ,  $b = 6.645$ . The functional form of  $\mu(x)$  is highly nonlinear and it is motivated by the drift function of Heston model and the GARCH diffusion model after taking the log transformation, which can be done by a simple application of Itô's lemma.

Figure 2 gives the plot for the drift function at different weights and the power curves for all the tests under the three sample sizes considered. It is observed that all the tests exhibit higher power when the sample size increases. For the density based tests, the model-free test and the original  $L^2$  type test seems to have almost identical power; while for the return based tests, the model-free test  $T_4$  seems more powerful than the original test  $T_3$ . When the sample size is small, the return density based tests outperform the volatility density based tests when the deviation is large (when  $\tau > 0.5$ ). When the sample size is large, the volatility density based test seems to be more powerful than the return based tests, except when the deviation to the null model is large (when  $\tau > 0.9$ ).

### 6.2.2 Misspecification in the diffusion function

In this sequence of alternative models, the drift function remains the same, but the diffusion function is deviating away from a constant:

$$d \ln \sigma_t^2 = \alpha(\beta - \ln \sigma_t^2)dt + \{(1 - \tau)\gamma + \tau\rho(\sigma_t^2)\}dW_t \quad (11)$$

for  $\tau = 0, 0.1, \dots, 1$ , where  $\rho(x) = c\sqrt{\exp(x)}$  with  $c = 3.3$ .

Figure 4 gives the plot for the diffusion function at different weights and the power curves for all the tests under the three sample sizes. The volatility density based tests strictly dominate the return based tests in the scenarios considered; the return density based tests seem to be sensitive to the sample size, it starts to get power only when the sample size is large (15 years data), the Cramer von Mises test seem to have no power for this type of alternative models, at least for the sample size used in the simulation.

To better understand the power for different tests under the two sequences of alternative models, the return density and volatility density are plotted under both the null hypothesis and the alternative hypothesis for the two types of alternative models in Figure 3 and 5. In Figure 3, the nonlinear deviation in the drift function cause both visible changes in the volatility density and the return density. In Figure 5 however, it is noticed that the deviation in the diffusion function cause big change in the volatility density, but only a small change in the variance of the return density. This perhaps explains why the volatility density based tests performs much better than the return density based tests for the deviations in the diffusion function.

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

## 7 Empirical example

In this section, the tests developed in this paper are applied to a daily British pound/Canadian dollar exchange rate dataset from January 1971 to August 1996. This dataset was used to estimate a discrete-time SARV model in Van der Sluis (1997). Figure 6 provides a plot of the exchange rate returns and the sample AutoCorrelation Function (ACF) of the squared returns, from which volatility clustering is observed.

The following continuous-time Log SARV model (9) is estimated with this dataset using the GMM estimator of Meddahi (2002):

$$\begin{aligned} dY_t &= \sigma_t dW_t, \\ d \ln \sigma_t^2 &= 26.8630(-4.4991 - \ln \sigma_t^2)dt + 3.0999dB_t, \end{aligned}$$

where the parameter values are annualized. The parameter values are very close to those implied by the discrete-time model estimated by Van der Sluis (1997).

The kernel  $K_1$  defined in (8) is used and the Cross-Validation method of Stefanski and Carroll (1990) is applied to the log returns to select the bandwidth, which gives

a bandwidth 0.28. Based on 1000 bootstrap samples, the  $p$ -values of all the tests are estimated. The  $p$ -values of all these tests are reported in the following table. All the tests show strong evidence of rejection of the null model, other than the Cramer von Mises test  $T_5$ , which can only reject the model if the significance level 0.1 is used.

[Figure 6 about here.]

[Table 2 about here.]

## 8 Conclusion

This paper studies volatility density based nonparametric specification tests for stochastic volatility models. The asymptotic null distributions of the test statistics and their asymptotic power properties are derived. A parametric bootstrap procedure is proposed to obtain the null distributions and the critical values. The finite sample properties of the method are studied using Monte Carlo simulations. The tests are applied to a simple empirical application.

With high-frequency data, alternative test statistics could be proposed. For example, using the methods developed in Kanaya and Kristensen (2015), nonparametric estimates of the drift and diffusion functions of the volatility process could be obtained, and test statistics could be formulated based on these functions. This approach of constructing tests will provide opportunities to test the original null hypothesis  $\mathcal{H}_0$  against the alternative hypothesis  $\mathcal{H}_1$  directly.

## Appendix A: Assumptions and probability properties of the stochastic volatility models

In this Appendix, the basic setup and the assumptions of the model (1) are established. These assumptions are made for the observed return sequence to be stationary, ergodic and  $\beta$ -mixing with exponentially decaying coefficients, which are necessary for the limiting theorems to work. In the nonparametric model, it is sufficient to assume the observed return sequence  $\{y_{t_i}\}_{i=1}^n$  to satisfy the above conditions directly. In the parametric model, assumptions are imposed on the  $b(x; \theta)$  and  $a(x; \theta)$  functions.

In the parametric stochastic volatility model (1), it is assumed that

**(SV0)**  $(B, W)$  is a standard Brownian motion in  $\mathbb{R}^2$ , defined on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , and  $\sigma_0^2$  is random variable defined on the same probability space, independent of  $(B, W)$ .

**(SV1)** The functions  $b(x; \theta)$  and  $a(x; \theta)$  are continuous function on  $\mathbb{R}^+$ , and continuously differentiable functions on  $(0, +\infty)$  such that

$$\exists C > 0, \quad \forall x > 0, \quad b^2(x; \theta) + a^2(x; \theta) \leq C(1 + x^2),$$

and

$$\forall x > 0, \quad a(x; \theta) > 0.$$

Now define, for  $v_0 > 0$ , the scale measure

$$s(x; \theta) = \exp \left( -2 \int_{v_0}^x \frac{b(v; \theta)}{a^2(v; \theta)} du \right),$$

and the speed measure

$$m(x; \theta) = \frac{1}{a^2(x; \theta)s(x; \theta)},$$

and assume

**(SV2)**

$$\int_0^u s(x; \theta) dx = +\infty, \quad \int_v^{+\infty} s(x; \theta) dx = +\infty, \quad \int_0^{+\infty} m(x; \theta) dx = M < +\infty,$$

where  $u > 0$ ,  $v < +\infty$  are arbitrary points in the domain of  $s(x; \theta)$ .

**(SV3)** The initial random variable  $\sigma_0^2$  has distribution  $\pi(dx) = \pi(x; \theta)dx$ .

**(SV4)**

$$\lim_{x \downarrow 0} a(x; \theta)m(x; \theta) = 0, \quad \lim_{x \uparrow +\infty} a(x; \theta)m(x; \theta) = 0.$$

**(SV5)** Define

$$\gamma(x; \theta) = a'(x; \theta) - \frac{2b(x; \theta)}{a(x; \theta)},$$

then as  $x \downarrow 0$  and  $x \uparrow +\infty$ , the limit of  $1/\gamma(x; \theta)$  exist.

(SV0) rules out the possibility of so called leverage effects. (SV1) ensures the existence and uniqueness of a almost surely positive strong solution to the volatility process. (SV2) implies the solution is positive recurrent on  $(0, +\infty)$  (positive recurrent is also called ergodic). The last condition in (SV2) guarantees the existence of a stationary distribution for the volatility process, with density defined as

$$\pi(x; \theta) = \frac{m(x; \theta)}{M} I\{x > 0\}. \quad (12)$$

So if the process is initiated from this stationary distribution as in (SV3), the volatility process is strictly stationary. Also since the process is aperiodic under (SV1) and (SV2),

the volatility process is  $\beta$ -mixing. These conditions are standard for diffusion processes, see Genon-Catalot et al. (1998) and Chen et al. (2010) for a detailed discussion. (SV4) and (SV5) (together with (SV1) and (SV2)) are actually sufficient conditions for the volatility process to be  $\rho$ -mixing. From Theorem 3.6 in Chen et al. (2010), a sufficient condition (together with (SV1) and (SV2)) for exponential decay  $\beta$ -mixing coefficients is the process to be  $\rho$ -mixing — this is not a very strong assumption, because as discussed in the same paper,  $\beta$ -mixing and  $\rho$ -mixing with exponential decay are almost equivalent concepts for a scalar diffusion. It is also a known result that if a diffusion process is  $\rho$ -mixing, its  $\rho$ -mixing coefficients decay at exponential rate (Bradley (2005), theorem 3.3, or Genon-Catalot et al. (2000), proposition 2.5). So with (SV4) and (SV5), the volatility process is  $\beta$ -mixing with exponentially decaying coefficients.

The exponential rate decay of the  $\beta$ -mixing coefficients for the observed returns is a sufficient condition to use the limiting theorem for U-statistics; it could be reduced to a polynomial rate if only for the purpose of applying the limiting theorem. However, establishing sufficient conditions for the volatility process in the parametric model to be  $\beta$ -mixing with a polynomial rate is not straightforward. In this sense, the exponential rate decay of the  $\beta$ -mixing coefficients should not be understood as a strong assumption.

The conditions above only deliver properties for the volatility process, the following lemma shows that the return sequence  $y_i = \int_{t_{i-1}}^{t_i} \sigma_s dB_s / \sqrt{\Delta}$ ,  $i = 1, \dots, n$ , which is a sequence of stochastic integrals of the volatility process with respect to an independent Brownian motion  $B$  over small fixed intervals, inherits the probabilistic properties of the volatility process.

**Lemma 1** *In model (1), if the volatility process is stationary, ergodic and  $\beta$  mixing with a certain decay rate, then the normalized return sequence  $y_i, i = 1, \dots, n$ , is also stationary, ergodic and  $\beta$  mixing with coefficients decaying at least as fast as that of the volatility.*

**Proof** Using Theorem 3.1 in Genon-Catalot et al. (2000).  $(y_i)_{i=1}^n$  satisfy a Hidden Markov model with hidden chain  $U_i := (\int_{t_{i-1}}^{t_i} \sigma_s^2 ds, \sigma_{t_i}^2)$ ,  $i = 1, \dots, n$ , thus also a Generalized Hidden Markov Model as in the definition in Carrasco and Chen (2002). Applying the proposition 4 in the latter paper, the return series  $(y_i)_{i=1}^n$  is ergodic, strictly stationary and  $\beta$ -mixing with at least the rate in the Hidden chain  $U_i, i = 1, \dots, n$ . As noted in the proof of proposition 3.2 in Genon-Catalot et al. (2000),

$$\beta_U(k) \leq \beta_{\sigma^2}((k-1)\Delta),$$

meaning that the decay of the  $\beta$ -mixing coefficient of  $(U_i)_{i=1}^n$  is at least as fast as that of  $(\sigma_{t_i}^2)_{i=1}^n$ , which completes the proof.  $\square$

It is assumed that:

(C1)  $g(x)$  is first differentiable,  $g(x)$  and its derivative are bounded and square integrable.  $f(x)$  is Lipschitz and bounded. Also assume the density of all the finite dimensional distributions of the observed returns to be bounded.

(B1) Under the null hypothesis, the parametric estimator is  $\sqrt{n}$ -consistent for the true parameter value.

(B1a) Under both the alternative hypotheses  $\mathcal{H}'_1$  and  $\mathcal{H}'_{1n}$ , the parametric estimator is  $\sqrt{n}$ -consistent for a pseudo-true value.

(B2)  $g(x, \theta)$  is Lipschitz in the parameter  $\theta$ , with the Lipschitz constant  $L(x)$  to be square integrable.

## Appendix B: Proofs of the theorems

**Proof** (of Theorem 1) The derivation of Theorem 1 is closely related to the derivation of the asymptotic normality of the Integrated Squared Error (ISE) of the kernel deconvolution estimator. For independent and identically distributed observations, the asymptotic distributions of ISE for different classes of signal densities and noise distributions were derived by Butucea (2004). The result in Theorem 1 extends the corresponding results of Theorem 4 in Butucea (2004) to dependent observations.

First the estimator can be written as

$$\hat{g}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \nu_h \left( \frac{x - Y_i}{h} \right) = \frac{1}{n} \sum_{i=1}^n \omega_h(x - Y_i),$$

where  $\omega_h$  is defined implicitly.

Notice that

$$\begin{aligned} T_0 &= \int \left( \frac{1}{n} \sum_{i=1}^n \omega_h(x - Y_i) - g(x; \hat{\theta}) \right)^2 dx \\ &= \int \left( \frac{1}{n} \sum_{i=1}^n \omega_h(x - Y_i) - K_h * g(x) + K_h * g(x) - g(x) + g(x) - g(x; \hat{\theta}) \right)^2 dx \\ &= T_1 + \int (K_h * g(x) - g(x))^2 dx + \int (g(x) - g(x; \hat{\theta}))^2 dx \\ &\quad + 2 \int \left( \frac{1}{n} \sum_{i=1}^n \omega_h(x - Y_i) - K_h * g(x) \right) (K_h * g(x) - g(x)) dx \\ &\quad + 2 \int \left( \frac{1}{n} \sum_{i=1}^n \omega_h(x - Y_i) - K_h * g(x) \right) (g(x) - g(x; \hat{\theta})) dx \\ &\quad + 2 \int (K_h * g(x) - g(x)) (g(x) - g(x; \hat{\theta})) dx, \end{aligned}$$

where  $T_1$  is the bias-corrected statistic defined in Section 4.1. It will be shown later that  $T_1 = O_p(\exp(\pi/h)/n)$ . Notice that the integrated squared bias  $\int (K_h * g(x) - g(x))^2 dx = O_p(h^2)$  because of Assumption C1 and the ‘‘infinite order kernel’’ property of the sinc kernel (see e.g. Theorem 8.1 of Glad et al. (2007));  $\int (g(x) - g(x; \hat{\theta}))^2 dx = O_p(n^{-1})$  follows from Assumption B2; so both terms are dominated by  $T_1$  under the bandwidth assumption. The cross product terms are also dominated by  $T_1$  because of the Cauchy-Schwarz inequality. One then has  $T_0 = T_1(1 + o_p(1))$ . Next the asymptotic distribution of  $T_1$  is derived.

First  $T_1$  is decomposed as follows,

$$\begin{aligned} T_1 &= \int \left( \frac{1}{n} \sum_{i=1}^n \omega_h(x - Y_i) - K_h * g(x) \right)^2 dx \\ &= \frac{1}{n^2} \sum_{i=1}^n \int (\omega_h(x - Y_i) - K_h * g(x))^2 dx \\ &\quad + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \int (\omega_h(x - Y_i) - K_h * g(x)) (\omega_h(x - Y_j) - K_h * g(x)) dx \\ &=: S_1 + S_2, \end{aligned}$$

where  $S_1$  and  $S_2$  are defined implicitly.

It is shown that

1.  $\left( \frac{4\pi^2 n^2}{\exp(2\pi/h) \|f\|_2^2} \right)^{1/2} |S_1 - ES_1| = o_p(1)$ , where  $ES_1 = (1/(2\pi^2 n)) \exp(\pi/h)$ .
2.  $\left( \frac{4\pi^2 n^2}{\exp(2\pi/h) \|f\|_2^2} \right)^{1/2} S_2 \xrightarrow{d} N(0, 1)$ .
3. Then it is proved that

$$\left( \frac{4\pi^2 n^2}{\exp(2\pi/h) \|f\|_2^2} \right)^{1/2} (T_1 - ES_1) \xrightarrow{d} N(0, 1),$$

by combining the above two results. This also derives the asymptotic distribution of  $T_0$ .

**1. Order of  $S_1$**  The order of  $E(S_1)$  is first evaluated. By stationarity,

$$\begin{aligned}
ES_1 &= \frac{1}{n} E \int (\omega_h(x - Y_1) - K_h * g(x))^2 dx \\
&= \frac{1}{n} \int E (\omega_h^2(x - Y_1) - 2\omega_h(x - y_1)K_h * g(x) + (K_h * g(x))^2) dx \\
&= \frac{1}{n} \left( \int E (\omega_h(x - Y_1))^2 dx - \int (K_h * g(x))^2 dx \right) \\
&= \frac{1}{n} \left( \int \int \frac{1}{h^2} \nu_h^2 \left( \frac{x - Y_1}{h} \right) f(y_1) dy_1 dx - \int (K_h * g(x))^2 dx \right) \\
&= \frac{1}{n} \left( \int \int \frac{1}{h} \nu_h^2(z) f(x - zh) dz dx - \int (K_h * g(x))^2 dx \right) \\
&= \frac{1}{n} \left( \int \int \frac{1}{h} \nu_h^2(z) f(x) dz dx (1 + o(1)) - \int (K_h * g(x))^2 dx \right) \\
&= \frac{1}{nh} \|\nu_h\|_2^2 (1 + o(1)) - \frac{1}{n} \int (K_h * g(x))^2 dx,
\end{aligned}$$

where in the third step  $E\omega_h(x - Y_1) = K_h * g(x)$  (see e.g. Stefanski and Carroll (1990)) is used, in the fifth step a change of variable using  $z = (x - y_1)/h$  is applied and in the sixth step the Lipschitz assumption for  $f(x)$  is used.

Use Lemma 3, it can be calculated that the first term is  $(1/(2\pi^2n)) \exp(\pi/h)$ ; the second term is  $O(1/n)$  by noticing that  $\int (K_h * g(x))^2 dx = (1/(2\pi)) \int |\phi_K(th) * \phi_g(t)|^2 dt \leq (1/(2\pi)) \int |\phi_g(t)|^2 dt$  and that  $g(x)$  is square integrable; so the first term is dominating and it is shown that

$$ES_1 = (1/(2\pi^2n)) \exp(\pi/h). \quad (13)$$

Then the order of the variance of  $Var(S_1)$  is evaluated.

$$\begin{aligned}
&Var(S_1) \\
&= \frac{1}{n^4} \sum_{i=1}^n Var \left( \int (\omega_h(x - Y_i) - K_h * g(x))^2 dx \right) + \\
&\quad \frac{2}{n^4} \sum_{i=1}^{n-1} (n - i) Cov \left( \int (\omega_h(x - Y_1) - K_h * g(x))^2 dx, \int (\omega_h(x - Y_{i+1}) - K_h * g(x))^2 dx \right)
\end{aligned} \quad (14)$$

For the diagonal terms, it is noticed that

$$\begin{aligned}
& \text{Var} \left( \int (\omega_h(x - Y_i) - K_h * g(x))^2 dx \right) \\
& \leq E \left( \int (\omega_h(x - Y_i) - K_h * g(x))^2 dx \right)^2 \\
& = E \left( \int (\omega_h(x - Y_i))^2 dx \right)^2 (1 + o(1)) \\
& = \frac{1}{h^2} \|\nu_h\|_2^4 (1 + o(1)),
\end{aligned} \tag{15}$$

where a similar strategy in deriving the order for  $ES_1$  is used.

For the covariance term, notice that

$$\begin{aligned}
& \text{Cov} \left( \int (\omega_h(x - Y_1) - K_h * g(x))^2 dx, \int (\omega_h(x - Y_{i+1}) - K_h * g(x))^2 dx \right) \\
& \leq \beta(i)^{\delta/(1+\delta)} E \left( \left| \int (\omega_h(x - Y_i) - K_h * g(x))^2 dx \right|^{1+\delta} \right)^{1/(1+\delta)} \\
& = \beta(i)^{\delta/(1+\delta)} \frac{1}{h} \|\nu_h\|_2^2 (1 + o(1))
\end{aligned} \tag{16}$$

for  $\delta > 0$ , where Lemma 1 of Yoshihara (1976) is used.

Plug the results (15) and (16) back into (14), one can obtain that

$$\begin{aligned}
& \text{Var}(S_1) \\
& = \frac{1}{n^3 h^2} \|\nu_h\|_2^4 (1 + o(1)) + \frac{2}{n^3 h} \|\nu_h\|_2^2 \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \beta(i)^{(\delta-2)/\delta} (1 + o(1)) \\
& \leq \frac{C}{n^3 h^2} \|\nu_h\|_2^4 (1 + o(1)),
\end{aligned}$$

for a large constant  $C > 0$  because  $\sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \beta(i)^{(\delta-2)/\delta} = O(1)$  by the exponential rate of decay for the  $\beta$ -mixing coefficients. Use Lemma 3, it is obtained that

$$\text{Var}(S_1) = O \left( \frac{1}{n^3} \exp\left(\frac{2\pi}{h}\right) \right). \tag{17}$$

Combine the results in (13) and (17), and use the Markov's inequality one has

$$S_1 = O_p \left( \frac{\exp(\pi/h)}{n} \right)$$

and

$$(S_1 - ES_1) = O_p \left( \frac{\exp(\pi/h)}{n^{3/2}} \right).$$

**Limiting distribution of  $S_2$**  Recall that

$$S_2 = \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \int (\omega_h(x - Y_i) - K_h * g(x)) (\omega_h(x - Y_j) - K_h * g(x)) dx$$

This is a U-statistic with kernel function

$$H_n(x, y) = \frac{2}{n^2} \int (\omega_h(u - x) - K_h * g(u)) (\omega_h(u - y) - K_h * g(u)) du.$$

It is easy to see this is a symmetric kernel, and

$$EH_n(Y_i, x) = 0,$$

$\forall x \in \mathbb{R}$ . Now the central limit theorem from Hjellvik et al. (1998) is applied to derive the limiting distribution of  $S_2$ .<sup>3</sup>

First, the asymptotic variance  $\sigma_n^2$  is calculated. This is done by calculating the asymptotically equivalent quantity  $n^2\sigma_0^2/2$ , where  $\sigma_0^2 = \int H_n^2(Y_1, Y_2)dP(Y_1)dP(Y_2)$ . Define the notation  $\xi_h(x) = \int \nu_h(x+z)\nu_h(z)dz$ , one has:

$$\begin{aligned} & \sigma_0^2 \\ &= \frac{4}{n^4} \int \int \left( \int (\omega_h(u - y_1) - K_h * g(u)) (\omega_h(u - y_2) - K_h * g(u)) du \right)^2 f(y_1)f(y_2)dy_1dy_2 \\ &= \frac{4}{n^4} \int \int \left( \int \omega_h(u - y_1)\omega_h(u - y_2)du \right)^2 f(y_1)f(y_2)dy_1dy_2 (1 + o(1)) \\ &= \frac{4}{n^4h^4} \int \int \left( \int \nu_h\left(\frac{u - y_1}{h}\right)\nu_h\left(\frac{u - y_2}{h}\right)du \right)^2 f(y_1)f(y_2)dy_1dy_2 \\ &= \frac{4}{n^4h^2} \int \int \left[ \xi_h^2\left(\frac{y_1 - y_2}{h}\right) \right] f(y_1)f(y_2)dy_1dy_2 \\ &= \frac{4}{n^4h} \left[ \int \xi_h^2(x) dx \times \|f\|_2^2 + o\left( \int \xi_h^2(x) dx \right) \right]. \end{aligned}$$

The last approximation above is valid because for an arbitrarily small  $\varepsilon > 0$ ,  $\varepsilon \rightarrow 0$  and

---

<sup>3</sup>Notice that the assumption  $E[H_n(Y_i, Y_j)|\mathcal{F}_{j-1}] = 0$ , for any  $i < j$  does not hold apparently, but as discussed in the remark A.2 in Gao and King (2004), this assumption is not essential and can actually be removed in this type of central limit theorems for U-statistics.

$\varepsilon/h \rightarrow \infty$  when  $n \rightarrow \infty$ , the following holds:

$$\begin{aligned}
& \left| \int \int \left[ \frac{1}{h} \xi_h^2 \left( \frac{y_1 - y_2}{h} \right) \right] f(y_1) f(y_2) dy_1 dy_2 - \int \xi_h^2(x) dx \times \|f\|_2^2 \right| \\
&= \left| \int \int \xi_h^2(u) f(uh + y_2) f(y_2) du dy_2 - \int \xi_h^2(x) dx \int f^2(x) dx \right| \\
&= \left| \int \int \xi_h^2(u) (f(uh + y_2) - f(y_2)) f(y_2) du dy_2 \right| \\
&\leq \int \int \xi_h^2(u) |f(uh + y_2) - f(y_2)| f(y_2) du dy_2 \\
&= \int \int_{|uh| \leq \varepsilon} \xi_h^2(u) |f(uh + y_2) - f(y_2)| f(y_2) du dy_2 \\
&\quad + \int \int_{|uh| > \varepsilon} \xi_h^2(u) |f(uh + y_2) - f(y_2)| f(y_2) du dy_2 \\
&\leq \int \int_{|uh| \leq \varepsilon} \xi_h^2(u) f(y_2) du dy_2 \varepsilon L \\
&\quad + \int \int_{|uh| > \varepsilon} \xi_h^2(u) f(y_2) du dy_2 \times 2 \sup_f \|f\|_\infty \\
&\leq o\left(\int \xi_h^2(x) dx\right) + o(1),
\end{aligned}$$

where in the fifth step the Lipschitz and boundedness assumptions for  $f(x)$  are used.

Using Lemma 3, one has

$$\sigma_n^2 \sim \frac{n^2}{2} \sigma_0^2 = \frac{1}{4\pi^2 n^2} \exp\left(\frac{2\pi}{h}\right) \times \|f\|_2^2 \times (1 + o(1)). \quad (18)$$

Next, upper bounds for  $M_{in}$  for  $i = 1, \dots, 6$  are evaluated. Basically, these are terms of the form,

$$\begin{aligned}
H_{ij} H_{kl} &= H_n(Y_i, Y_j) H_n(Y_k, Y_l) \\
&= \frac{4}{n^4} \int (\omega_h(u - Y_i) - K_h * g(u)) (\omega_h(u - Y_j) - K_h * g(u)) du \\
&\quad \times \int (\omega_h(u - Y_k) - K_h * g(u)) (\omega_h(u - Y_l) - K_h * g(u)) du \\
&= \frac{4}{n^4} \int \omega_h(u - Y_i) \omega_h(u - Y_j) du \int \omega_h(u - Y_k) \omega_h(u - Y_l) du (1 + o(1)),
\end{aligned}$$

where only the dominating term is kept.

I start with the quantity  $M_{n2}$ .

$$\begin{aligned}
& E|H_{1j}H_{ij}|^{2(1+\delta)} \\
&= \left(\frac{1}{n^4h^2}\right)^{2(1+\delta)} E \left| \xi_h \left( \frac{Y_1 - Y_j}{h} \right) \xi_h \left( \frac{Y_i - Y_j}{h} \right) \right|^{2(1+\delta)} \\
&= \left(\frac{1}{n^4h^2}\right)^{2(1+\delta)} \int \int \int \left| \xi_h \left( \frac{y_1 - y_j}{h} \right) \xi_h \left( \frac{y_i - y_j}{h} \right) \right|^{2(1+\delta)} f(y_1, y_i, y_j) dy_1 dy_i dy_j \\
&= \left(\frac{1}{n^4h^2}\right)^{2(1+\delta)} h^2 \int \int \int |\xi_h(z)\xi_h(z')|^{2(1+\delta)} f(y_j + zh, y_j + z'h, y_j) dz dz' dy_j \\
&\leq C \left(\frac{1}{n^4h^2}\right)^{2(1+\delta)} h^2 \|\xi_h\|_{2(1+\delta)}^{4(1+\delta)},
\end{aligned}$$

where in the last step the integrability for the joint density function is used and  $C > 0$  is a generic large constant. With the same method, it can be shown that other quantities in the definition of  $M_{n2}$  share this same upper bound, such that this is also an upper bound for their maximum  $M_{n2}$ . Use Lemma 3, one has

$$\frac{n^{3/2}M_{n2}^{\frac{1}{2(1+\delta)}}}{\sigma_n^2} = O(n^{-1/2}) = o(1),$$

for  $\delta > 0$ .

Using the same strategy it can be shown that

$$\begin{aligned}
E|H_{1j}H_{ij}|^2 &\leq C \left(\frac{1}{n^4h^2}\right)^2 h^2 \|\xi_h\|_2^4, \\
E|H_{1i}H_{jk}|^{2(1+\delta)} &\leq C \left(\frac{1}{n^4h^2}\right)^{2(1+\delta)} h^2 \|\xi_h\|_{2(1+\delta)}^{4(1+\delta)},
\end{aligned}$$

where  $C > 0$  is a generic large constant. Such that  $n^{3/2}M_{n3}^{1/2}/\sigma_n^2 = O(n^{-1/2}) = o(1)$  and  $n^{3/2}M_{n4}^{1/2}/\sigma_n^2 = O(n^{-1/2}) = o(1)$ .

For  $M_{n5}$ , notice that

$$\begin{aligned}
& E \left| \int H_{1i} H_{1j} dP(Y_1) \right|^{2(1+\delta)} \\
&= E \left| \frac{4}{n^4 h^2} \int \xi_h \left( \frac{y_1 - Y_i}{h} \right) \xi_h \left( \frac{y_1 - Y_j}{h} \right) f(y_1) dy_1 \right|^{2(1+\delta)} (1 + o(1)) \\
&= \left( \frac{1}{n^4 h} \right)^{2(1+\delta)} \int \int \left| \int \xi_h(z) \xi_h \left( z + \frac{y_i - y_j}{h} \right) f(zh + y_i) dz \right|^{2(1+\delta)} \\
&\quad f(y_i, y_j) dy_i dy_j (1 + o(1)) \\
&\leq C \left( \frac{1}{n^4 h} \right)^{2(1+\delta)} \int \int \int \left| \xi_h(z) \xi_h \left( z + \frac{y_i - y_j}{h} \right) \right|^{2(1+\delta)} f(zh + y_i) dz f(y_i, y_j) dy_i dy_j \\
&= C \left( \frac{1}{n^4 h} \right)^{2(1+\delta)} h \int \int \int |\xi_h(z) \xi_h(z + u)|^{2(1+\delta)} f(zh + y_i) dz f(y_i, y_i - uh) dy_i du \\
&\leq C \left( \frac{1}{n^4 h} \right)^{2(1+\delta)} h \left( \int |\xi_h(z)|^{2(1+\delta)} dz \right)^2 \\
&= C \left( \frac{1}{n^4 h} \right)^{2(1+\delta)} h \|\xi_h\|_{2(1+\delta)}^{4(1+\delta)},
\end{aligned}$$

where  $C > 0$  is a generic large constant. In the third step the Jensen's inequality is used, in the fifth step the integrability of the joint density function of observed returns is used. With the same argument, the other quantities in the definition of  $M_{n5}$  also have this upper bound, so  $M_{n5}$  has this upper bound. Still use the result of Lemma 3, it is obtained that

$$\frac{n^2 M_{n5}^{\frac{1}{2(1+\delta)}}}{\sigma_n^2} = O(h^{1 - \frac{1}{2(1+\delta)}}) = o(1),$$

for  $\delta > 0$ . In the same fashion, it can be shown that

$$E \left| \int H_{1i} H_{1j} dP(Y_1) \right|^2 \leq C_6 n^{-8} h^{-1} \|\xi_h(x)\|_2^4,$$

and

$$\frac{n^2 M_{n6}^{\frac{1}{2}}}{\sigma_n^2} = O(h^{1/2}) = o(1).$$

Now, only the condition for  $M_{n1}$  is left for verifying. As above, it can be calculated that

$$\begin{aligned}
E |H_{1j}H_{ij}|^{1+\delta} &= \left(\frac{1}{n^4h^2}\right)^{1+\delta} E \left| \xi_h \left(\frac{Y_1 - Y_j}{h}\right) \xi_h \left(\frac{Y_i - Y_j}{h}\right) \right|^{1+\delta} \\
&\leq \left(\frac{1}{n^4h^2}\right)^{1+\delta} \|\xi_h\|_\infty^{2(1+\delta)} \\
&= O\left(\left(\frac{1}{n^4h^2}\right)^{1+\delta} h^{2(1+\delta)} \exp\left(\frac{2(1+\delta)\pi}{h}\right)\right) \\
&= O\left(\left(\frac{1}{n^4}\right)^{1+\delta} \exp\left(\frac{2(1+\delta)\pi}{h}\right)\right).
\end{aligned}$$

where again Lemma 3 is used. However, with this upper bound one can only show  $n^2M_{n1}^{1/(1+\delta)}/\sigma_n^2 = O(1)$ , instead of  $o(1)$ . It is now shown that the central limit theorem in Hjellvik et al. (1998) still holds.

Looking at the proof of the theorem of Hjellvik et al. (1998) (Page 25), notice that the condition  $n^2M_{n1}^{1/(1+\delta)}/\sigma_n^2 \rightarrow 0$  is used to show that for the variance of the U-statistic, the sum of the covariance terms are dominated by the sum of the variances. That is, for

$$S_2 = \frac{2}{n^2} \sum_{i < j} H_n(Y_i, Y_j),$$

the variance satisfies

$$\begin{aligned}
\text{Var}(S_2) &= \sum_{i < j} \text{Var}(H_n(Y_i, Y_j)) + 2 \sum_{i < j < k} E H_n(Y_i, Y_k) H_n(Y_j, Y_k) \\
&= \sum_{i < j} \text{Var}(H_n(Y_i, Y_j))(1 + o(1)).
\end{aligned} \tag{19}$$

It is shown next that (19) still holds when  $n^2M_{n1}^{1/(1+\delta)}/\sigma_n^2 = O(1)$ . Notice that the

sum of the covariance terms can be rewritten as follows,

$$\begin{aligned}
& \sum_{i < j < k} EH_n(Y_i, Y_k)H_n(Y_j, Y_k) \\
&= \sum_{k=3}^{n-1} \left( \sum_{i=1}^{k-1} \sum_{j=i+1}^{k-1} EH_n(Y_i, Y_k)H_n(Y_j, Y_k) \right) \\
&= \sum_{k=3}^{n-1} \left( \sum_{i=2}^{k-1} (k-i)EH_n(Y_1, Y_k)H_n(Y_i, Y_k) \right) \\
&= \sum_{i=2}^{n-2} \sum_{k=i+1}^{n-1} (k-i)EH_n(Y_1, Y_k)H_n(Y_i, Y_k) \\
&= \sum_{i=2}^{c_n} \sum_{k=i+1}^{n-1} (k-i)EH_n(Y_1, Y_k)H_n(Y_i, Y_k) + \sum_{i=c_n+1}^{n-1} \sum_{k=i+1}^{n-1} (k-i)EH_n(Y_1, Y_k)H_n(Y_i, Y_k)
\end{aligned}$$

where in the last step, the outer layer sum is splitter into two parts.<sup>4</sup>

First it is noticed that

$$|EH_n(Y_1, Y_k)H_n(Y_i, Y_k)| = O_p \left( \frac{1}{n^4} \exp \left( \frac{\pi}{h} \right) \right), \quad (20)$$

---

<sup>4</sup>In existing literature of nonparametric statistics for dependent data, this strategy had already been used in e.g. Fan and Yao (2002) Section 2.7.7, Proof of Theorem 2.22; or Masry (1993), Proof of Lemma 3.2 in similar contexts.

uniformly for all  $i, k$ . This is because

$$\begin{aligned}
& |EH_n(Y_1, Y_k)H_n(Y_i, Y_k)| \\
&= \left| \frac{4}{n^4 h^4} E \int \int \nu_h \left( \frac{u - Y_1}{h} \right) \nu_h \left( \frac{u - Y_k}{h} \right) \nu_h \left( \frac{v - Y_i}{h} \right) \nu_h \left( \frac{v - Y_k}{h} \right) dudv \right| \\
&= \left| \frac{4}{n^4 h^4} \int \dots \int \frac{\phi_K(t)\phi_K(t')\phi_K(t'')\phi_K(t''')}{\phi_k(t/h)\phi_k(t'/h)\phi_k(t''/h)\phi_k(t'''/h)} e^{-itu/h} e^{-it'u/h} e^{-it''v/h} e^{-it'''v/h} \right. \\
&\quad \left. E(e^{itY_1/h} e^{it'Y_k/h} e^{it''Y_i/h} e^{it'''Y_k/h}) dt dt' dt'' dt''' dudv \right| \\
&= \left| \frac{4}{n^4 h^4} \int \dots \int \frac{\phi_K(t)\phi_K(t')\phi_K(t'')\phi_K(t''')}{\phi_k(t/h)\phi_k(t'/h)\phi_k(t''/h)\phi_k(t'''/h)} e^{-itu/h} e^{-it'u/h} e^{-it''v/h} e^{-it'''v/h} \right. \\
&\quad \left. E(e^{it(X_1+\varepsilon_1)/h} e^{it'(X_k+\varepsilon_k)/h} e^{it''(X_i+\varepsilon_i)/h} e^{it'''(X_k+\varepsilon_k)/h}) dt dt' dt'' dt''' dudv \right| \\
&= \left| \frac{4}{n^4 h^4} \int \dots \int \frac{\phi_K(t)\phi_K(t')\phi_K(t'')\phi_K(t''')}{\phi_k(t/h)\phi_k(t'/h)\phi_k(t''/h)\phi_k(t'''/h)} e^{-itu/h} e^{-it'u/h} e^{-it''v/h} e^{-it'''v/h} \right. \\
&\quad \left. \phi_k(t/h)\phi_k(t''/h)\phi_k((t'+t''')/h) E(e^{itX_1/h} e^{it'X_k/h} e^{it''X_i/h} e^{it'''X_k/h}) dt dt' dt'' dt''' dudv \right| \\
&= \left| \frac{4}{n^4 h^4} \int \dots \int \frac{\phi_K(t)\phi_K(t')\phi_K(t'')\phi_K(t''')}{\phi_k(t'/h)\phi_k(t'''/h)} e^{-itu/h} e^{-it'u/h} e^{-it''v/h} e^{-it'''v/h} \right. \\
&\quad \left. \phi_k((t'+t''')/h) E(e^{itX_1/h} e^{it'X_k/h} e^{it''X_i/h} e^{it'''X_k/h}) dt dt' dt'' dt''' dudv \right| \\
&\leq \left| \frac{4}{n^4 h^4} \int \dots \int \frac{\phi_K(t)\phi_K(t')\phi_K(t'')\phi_K(t''')}{\phi_k(t'/h)\phi_k(t'''/h)} (e^{-itu/h})(e^{-it'u/h}) \left| (e^{-it''v/h})(e^{-it'''v/h}) \right| \right. \\
&\quad \left. |\phi_k((t'+t''')/h)| E \left| e^{itX_1/h} e^{it'X_k/h} e^{it''X_i/h} e^{it'''X_k/h} \right| dt dt' dt'' dt''' dudv \right| \\
&\leq C \left| \frac{4}{n^4 h^4} \int \dots \int \frac{\phi_K(t)\phi_K(t')\phi_K(t'')\phi_K(t''')}{\phi_k(t'/h)\phi_k(t'''/h)} (e^{-itu/h})(e^{-it'u/h}) dt dt' dt'' dt''' dudv \right| \\
&= \frac{C}{n^4 h^4} \left| \int \int K\left(\frac{u}{h}\right) K\left(\frac{v}{h}\right) dudv \int \int \frac{\phi_K(t')\phi_K(t''')}{\phi_k(t'/h)\phi_k(t'''/h)} dt' dt''' \right| \\
&= \frac{C}{n^4 h^4} \left( \int K\left(\frac{u}{h}\right) du \right)^2 \left( \left| \int \frac{\phi_K(t)}{\phi_k(t/h)} dt \right| \right)^2 \\
&\leq \frac{C}{n^4 h^2} \left( \int K(u) du \right)^2 \left( \int \left| \frac{\phi_K(t)}{\phi_k(t/h)} \right| dt \right)^2 = O\left(\frac{1}{n^4} \exp\left(\frac{\pi}{h}\right)\right).
\end{aligned}$$

On the other hand, from the proof of Theorem A in Hjellvik et al. (1998), it is known that

$$|EH_n(Y_i, Y_k)H_n(Y_j, Y_k)| \leq 4M_{n1}^{1/(1+\delta)} \beta(j-i)^{\delta/(1+\delta)}. \quad (21)$$

Use the bound (20) in the sum  $\sum_{i=2}^{c_n}$  and the bound (21) in the sum  $\sum_{i=c_n+1}^{n-1}$ , it holds

that

$$\begin{aligned}
& \sum_{i < j < k} EH_n(Y_i, Y_k)H_n(Y_j, Y_k) \\
&= \sum_{i=2}^{c_n} \sum_{k=i+1}^{n-1} (k-i)EH_n(Y_1, Y_k)H_n(Y_i, Y_k) + \sum_{i=c_n+1}^{n-1} \sum_{k=i+1}^{n-1} (k-i)EH_n(Y_1, Y_k)H_n(Y_i, Y_k) \\
&= n \sum_{i=2}^{c_n} \sum_{k=i+1}^{n-1} \frac{(k-i)}{n} EH_n(Y_1, Y_k)H_n(Y_i, Y_k) + n \sum_{i=c_n+1}^{n-1} \sum_{k=i+1}^{n-1} \frac{(k-i)}{n} EH_n(Y_1, Y_k)H_n(Y_i, Y_k) \\
&\leq c_n n^2 O_p \left( \frac{1}{n^4} \exp \left( \frac{\pi}{h} \right) \right) + 4n^2 M_{n1}^{1/(1+\delta)} \sum_{i=c_n+1}^{n-1} \beta(i-1)^{\delta/(1+\delta)}.
\end{aligned}$$

Since  $\sigma_n^2 = (4/n^2) \exp(2\pi/h)$ , choose, say  $c_n = \exp(\pi/(2h))$  the first term will have order  $O_p(\exp(3\pi/(2h))/n^2)$  and it is dominated by  $\sigma_n^2$ . The second term is also dominated by  $\sigma_n^2$  by noticing that  $n^2 M_{n1}^{1/(1+\delta)} = O(1)$  and  $\sum_{i=c_n+1}^{n-1} \beta(i-1)^{\delta/(1+\delta)} \rightarrow 0$  because of the exponential rate of decay of the  $\beta$ -mixing coefficients and  $c_n = \exp(\pi/(2h)) \rightarrow \infty$ .

Combining the expression for the asymptotic variance for the U-statistic in (18), and with the sufficient conditions for the central limit theorem in Hjellvik et al. (1998) satisfied, the limiting distribution for  $S_2$  is obtained and the central limit theorem of  $T_1$  is proved. The central limit theorem for  $(T_0 - \mu_{n,1})/\hat{\sigma}_{n,1}$  follows easily by the Slutsky's theorem because a consistent estimator  $\widehat{\|f\|_2}$  for  $\|f\|_2$  is used.  $\square$

**Proof** (of Theorem 2) The central limit theorem for  $T_1$  has already been proved in the proof of Theorem 1. Since there is no Integrated Squared Bias term in the definition of  $T_1$ , the condition  $\exp(\pi/h)/(nh^2) \rightarrow \infty$  is dropped from the assumption of Theorem 1. The central limit theorem for  $T_2$  follows easily.  $\square$

**Proof** (of Theorem 3) Notice that

$$\begin{aligned}
T_1 &= \int \left( \hat{g}(x) - K_h * g(x; \hat{\theta}) \right)^2 dx \\
&= \int \left[ \left( \hat{g}(x) - K_h * g_1(x) \right) + \left( K_h * g_1(x) - K_h * g(x; \hat{\theta}) \right) \right]^2 dx \\
&= \int \left( \hat{g}(x) - K_h * g_1(x) \right)^2 dx + \int \left( K_h * g_1(x) - K_h * g(x; \hat{\theta}) \right)^2 dx \\
&\quad + 2 \int \left( \hat{g}(x) - K_h * g_1(x) \right) \left( K_h * g_1(x) - K_h * g(x; \hat{\theta}) \right) dx.
\end{aligned}$$

Under the alternative that  $g(x) = g_1(x) \neq g(x; \theta)$  for all  $\theta \in \Theta$ , it can be shown in the same way as in Theorem 1 that the first term is of order  $O_p((4\pi^2 n^2 / (\exp(2\pi/h) \|f\|^2))^{-1/2})$ ; the second term is  $O(1)$  as this is the  $L^2$  distance between the alternative and the pseudo-true model in the limit; while the third term is of order  $O_p((4\pi^2 n^2 / (\exp(2\pi/h) \|f\|^2))^{-1/4})$  by applying the Cauchy-Schwarz inequality and using the orders of the first two terms. With these orders, it is easy to see that  $(4\pi^2 n^2 / (\exp(2\pi/h) \|f\|^2))^{1/2} T_1 \rightarrow \infty$  and

$P(T_1 > Z_\alpha) \rightarrow 1$  is true. The consistency result for  $T_2$  can be shown analogously.  $\square$

**Proof** (of Theorem 4) Notice that

$$\begin{aligned} T_1 &= \int \left( \hat{g}(x) - K_h * g(x; \hat{\theta}) \right)^2 dx \\ &= \int \left( \hat{g}(x) - K_h * g_n(x) \right)^2 dx + \int \left( K_h * g_n(x) - g(x; \hat{\theta}) \right)^2 dx \\ &\quad + 2 \int \left( \hat{g}(x) - K_h * g_n(x) \right) \left( K_h * g_n(x) - g(x; \hat{\theta}) \right) dx \\ &=: A + B + C, \end{aligned}$$

where  $A$ ,  $B$  and  $C$  are defined implicitly. Using similar method as in Theorem 1 it can be shown that under  $\mathcal{H}'_{1n}$ ,

$$\left( \frac{4\pi^2 n^2}{\exp(2\pi/h) \|f\|_2^2} \right)^{1/2} \left( A - \frac{1}{2\pi^2 n} e^{\pi/h} \right) \xrightarrow{d} N(0, 1).$$

With Assumptions (B1a) and (B2), one can obtain

$$B = \gamma_n^2 \int d^2(x) dx (1 + o(1)).$$

Next if it is shown that

$$C = o_p(\gamma_n^2),$$

then the results are proved.

For part C first notice that

$$\begin{aligned} C &= 2\gamma_n \int \left( \hat{g}(x) - K_h * g_n(x) \right) \left( d(x) + o(1) \right) dx \\ &= 2\gamma_n \int \left( \hat{g}(x) - K_h * g_n(x) \right) d(x) dx (1 + o(1)). \end{aligned}$$

So it can be calculated that

$$\begin{aligned} EC &= E \int \left( \hat{g}(x) - K_h * g_n(x) \right) d(x) dx \\ &= \int \left( K_h * g_n(x) - K_h * g_n(x) \right) d(x) dx \\ &= 0. \end{aligned}$$

For the variance, notice that

$$\begin{aligned}
\text{Var}(C) &= \text{Var} \left( \int (\hat{g}(x) - K_h * g_n(x)) d(x) dx \right) \\
&\leq E \left( \frac{1}{n} \sum_{j=1}^n \int (\omega_h(x - Y_j) - K_h * g_n(x)) d(x) dx \right)^2 \\
&= \frac{1}{n} E \left( \int (\omega_h(x - Y_1) - K_h * g_n(x)) d(x) dx \right)^2 + \\
&\quad \frac{2}{n^2} \sum_{j=1}^{n-1} \left( (n-j) E \int (\omega_h(x - Y_1) - K_h * g_n(x)) d(x) dx \right. \\
&\quad \left. \int (\omega_h(x - Y_{j+1}) - K_h * g_n(x)) d(x) dx \right).
\end{aligned}$$

The sum of diagonal terms satisfies

$$\begin{aligned}
&\frac{1}{n} E \left( \int (\omega_h(x - Y_1) - K_h * g_n(x)) d(x) dx \right)^2 \\
&= \frac{1}{n} E \left( \int \omega_h(x - Y_1) d(x) dx \right)^2 (1 + o(1)) \\
&= \frac{1}{n} E \left( \int \frac{1}{h} \nu_h \left( \frac{x - Y_1}{h} \right) d(x) dx \right)^2 (1 + o(1)) \\
&\leq \frac{1}{nh^2} E \left( \int d(x) dx \right)^2 \|\nu_h\|_\infty^2 (1 + o(1)) \\
&= o \left( \frac{1}{n} \exp \left( \frac{\pi}{h} \right) \right) = o(\gamma_n^2),
\end{aligned}$$

where the results in Lemma 3 and the assumption that  $\int d(x) dx = 0$  are used.

For the cross product terms,

$$\begin{aligned}
&\frac{2}{n} \sum_{j=1}^{n-1} \left( \frac{(n-j)}{n} E \int (\omega_h(x - Y_1) - K_h * g_n(x)) d(x) dx \int (\omega_h(x - Y_{j+1}) - K_h * g_n(x)) d(x) dx \right) \\
&\leq \frac{2}{n} \sum_{j=1}^{n-1} \left| E \int (\omega_h(x - Y_1) - K_h * g_n(x)) d(x) dx \int (\omega_h(x - Y_{j+1}) - K_h * g_n(x)) d(x) dx \right| \\
&= \frac{2}{n} \sum_{j=1}^{c_n} \left| E \int (\omega_h(x - Y_1) - K_h * g_n(x)) d(x) dx \int (\omega_h(x - Y_{j+1}) - K_h * g_n(x)) d(x) dx \right| \\
&\quad + \frac{2}{n} \sum_{j=c_n+1}^{n-1} \left| E \int (\omega_h(x - Y_1) - K_h * g_n(x)) d(x) dx \int (\omega_h(x - Y_{j+1}) - K_h * g_n(x)) d(x) dx \right| \quad (22)
\end{aligned}$$

where again the strategy of splitting the sum into two parts is used to evaluate the order of the cross product terms.

First, uniformly for all  $j$ ,

$$\begin{aligned}
& \left| E \int (\omega_h(x - Y_1) - K_h * g_n(x)) d(x) dx \int (\omega_h(x - Y_{j+1}) - K_h * g_n(x)) d(x) dx \right| \\
&= \left| E \int \int (\omega_h(x - Y_1) - K_h * g_n(x)) (\omega_h(y - Y_{j+1}) - K_h * g_n(y)) d(x) d(y) dx dy \right| \\
&= \left| \int \int E(\omega_h(x - Y_1) \omega_h(y - Y_{j+1})) d(x) d(y) dx dy (1 + o(1)) \right| \\
&= \left| \frac{1}{h^2} \int \int E \left( \int \int \frac{\phi_K(t) \phi_K(t')}{\phi_k(t/h) \phi_k(t'/h)} e^{-it \frac{x-Y_1}{h}} e^{-it' \frac{y-Y_{j+1}}{h}} dt dt' \right) d(x) d(y) dx dy (1 + o(1)) \right| \\
&= \left| \frac{1}{h^2} \int \int \int \int \frac{\phi_K(t) \phi_K(t')}{\phi_k(t/h) \phi_k(t'/h)} e^{-it \frac{x}{h}} e^{-it' \frac{y}{h}} E(e^{itY_1/h} e^{it'Y_{j+1}/h}) dt dt' d(x) d(y) dx dy (1 + o(1)) \right| \\
&= \frac{1}{h^2} \left| \int \int \int \int \phi_K(t) \phi_K(t') e^{-it \frac{x}{h}} e^{-it' \frac{y}{h}} E(e^{itX_1/h} e^{it'X_{j+1}/h}) dt dt' d(x) d(y) dx dy (1 + o(1)) \right| \\
&\leq \frac{1}{h^2} \left| \int \int \int \int \phi_K(t) \phi_K(t') e^{-it \frac{x}{h}} e^{-it' \frac{y}{h}} E|(e^{itX_1/h} e^{it'X_{j+1}/h})| dt dt' d(x) d(y) dx dy (1 + o(1)) \right| \\
&\leq \frac{1}{h^2} \left| \int \int \int \int \phi_K(t) \phi_K(t') e^{-it \frac{x}{h}} e^{-it' \frac{y}{h}} dt dt' d(x) d(y) dx dy (1 + o(1)) \right| \\
&= \frac{1}{h^2} \left| \int \int K \left( \frac{x}{h} \right) K \left( \frac{y}{h} \right) d(x) d(y) dx dy \right| (1 + o(1)) \\
&= \frac{1}{h^2} \left( \int K \left( \frac{x}{h} \right) d(x) dx \right)^2 \leq \frac{C}{h^2} \left( \int d(x) dx \right)^2 = o\left(\frac{1}{h^2}\right), \tag{23}
\end{aligned}$$

where in the tenth step the boundedness of the kernel  $K$  is used and in the final step the assumption that  $\int d(x) dx = 0$  is used.

On the other hand, using Lemma 1 of Yoshihara (1976),

$$\begin{aligned}
& \left| E \int (\omega_h(x - Y_1) - K_h * g_n(x)) d(x) dx \int (\omega_h(x - Y_{j+1}) - K_h * g_n(x)) d(x) dx \right| \\
&\leq \beta(j)^{\delta/(1+\delta)} \left( E \left| \int \omega_h(x - Y_1) d(x) dx \right|^{1+\delta} \right)^{1/(1+\delta)} (1 + o(1)) \\
&\leq \beta(j)^{\delta/(1+\delta)} \left( E \left| \left( \int \omega_h^2(x - Y_1) dx \right)^{1/2} \left( \int d^2(x) dx \right)^{1/2} \right|^{1+\delta} \right)^{1/(1+\delta)} (1 + o(1)) \\
&\leq \beta(j)^{\delta/(1+\delta)} \left( \int d^2(x) dx \right)^{1/2} \frac{1}{h^{1/2}} \|\nu_h\|_2^2 (1 + o(1)) \\
&= \beta(j)^{\delta/(1+\delta)} h^{1/2} \exp\left(\frac{\pi}{h}\right) (1 + o(1)), \tag{24}
\end{aligned}$$

where in the second step the Cauchy-Schwarz theorem is used, and in the final step the results from Lemma 3 is again used.

Plug the results in (24) and (23) back into (22), one has

$$\begin{aligned} & \frac{2}{n^2} \sum_{j=1}^{n-1} \left( (n-j) E \int (\omega_h(x - Y_1) - K_h * g_n(x)) d(x) dx \int (\omega_h(x - Y_{j+1}) - K_h * g_n(x)) d(x) dx \right) \\ &= O\left(\frac{c_n}{nh^2}\right) + \frac{2}{n} h^{1/2} \exp\left(\frac{\pi}{h}\right) \sum_{j=c_n+1}^{n-1} \beta(j)^{\delta/(1+\delta)}. \end{aligned}$$

Take, say  $c_n = \exp(\pi/(2h)) \rightarrow \infty$ , the first term  $c_n/n = o(\exp(\pi/h)/n)$ ; the second term is also  $o(\exp(\pi/h)/n)$  because of the exponential decay rate of the  $\beta$ -mixing coefficients, so one can have  $\text{Var}\left(\int (\hat{g}(x) - K_h * g_n(x)) d(x) dx\right) = o(\gamma_n^2)$ , this in turn implies that  $\text{Var}(C) = o(\gamma_n^4)$ . Combine the order results for the mean and the variance of  $C$ , use the Markov's inequality one can have  $C = o_p(\gamma_n^2)$ .

The local power properties for  $T_2$  can be proved analogously.  $\square$

**Proof** (or Proposition 1) The proof goes in the same way as that of Theorem 1, the only difference is that  $\phi_K(t) = I(|t| \leq 1)$  is replaced with  $\phi_{K_1}(t) = (1 - t^2)^3 I(|t| \leq 1)$ . The orders of different norms of the functions  $\nu_h(x)$  and  $\xi_h(x)$  when using the kernel  $K_1$  are given in Lemma 4 of Appendix C.

With the same notation as in the proof of Theorem 1, the proof proceeds along the following steps:

1.  $\frac{1}{\sigma_{n,2}} |S_1 - \mu_{n,2}| = o_p(1)$ ,
2.  $\frac{1}{\sigma_{n,2}} S_2 \xrightarrow{d} N(0, 1)$ .
3. Then it is proved that

$$\frac{1}{\sigma_{n,2}} (T_1 - \mu_{n,2}) \xrightarrow{d} N(0, 1),$$

by combining the above two results.

**Order of  $S_1$**  Similarly as in the proof of Theorem 1 one has

$$ES_1 = \frac{1}{nh} \|\nu_h\|_2^2 (1 + o(1)) = \frac{h^6}{n} \exp\left(\frac{\pi}{h}\right) (1 + o(1)),$$

$$\text{Var}(S_1) \leq \frac{C}{n^3 h^2} \|\nu_h\|_2^4 (1 + o(1)) = \frac{h^{12}}{n^3} \exp\left(\frac{2\pi}{h}\right) (1 + o(1)).$$

Combine the above results, one can obtain that

$$S_1 = O_p\left(\frac{h^6}{n} \exp\left(\frac{\pi}{h}\right)\right),$$

and

$$(S_1 - ES_1) = O_p\left(\frac{h^6}{n^{3/2}} \exp\left(\frac{\pi}{h}\right)\right).$$

**Limiting distribution of  $S_2$**  Similarly as in the proof of Theorem 1,

$$\begin{aligned} & \sigma_0^2 \\ &= \int H_n^2(Y_1, Y_2) dP(Y_1) dP(Y_2) \\ &= \frac{4}{n^4 h} \left[ \int \xi_h^2(x) dx \times \|f\|_2^2 + o\left(\int \xi_h^2(x) dx\right) \right]. \end{aligned}$$

Now, use the result in Lemma 4 one can obtain

$$\sigma_n^2 \sim \frac{n^2}{2} \sigma_0^2 = C_2 \frac{h^{12}}{n^2} \exp\left(\frac{2\pi}{h}\right) \times \|f\|_2^2 \times (1 + o(1)),$$

where  $C_2 = 720^2 \times 231/\pi^{14}$ .

Next, upper bounds for  $M_{in}$  for  $i = 1, \dots, 6$  are evaluated. I start with the quantity  $M_{n2}$ . Similarly as in proof of Theorem 1,

$$E|H_{1j}H_{ij}|^{2(1+\delta)} \leq C \left(\frac{1}{n^4 h^2}\right)^{2(1+\delta)} h^2 \|\xi_h\|_{2(1+\delta)}^{4(1+\delta)}.$$

Using the same method as before, but use the result of Lemma 4, it is easy to check that

$$\frac{n^{3/2} M_{n2}^{\frac{1}{2(1+\delta)}}}{\sigma_n^2} = O(n^{-1/2}) = o(1),$$

for  $\delta > 0$ .

In the same way, it can be shown that

$$\begin{aligned} E|H_{1j}H_{ij}|^2 &\leq C \left(\frac{1}{n^4 h^2}\right)^2 h^2 \|\xi_h\|_2^4, \\ E|H_{1i}H_{jk}|^{2(1+\delta)} &\leq C \left(\frac{1}{n^4 h^2}\right)^{2(1+\delta)} h^2 \|\xi_h\|_{2(1+\delta)}^{4(1+\delta)}. \end{aligned}$$

Still use the results of Lemma 4, one can check that  $n^{3/2} M_{n3}^{1/2}/\sigma_n^2 = o(1)$  and  $n^{3/2} M_{n4}^{1/2}/\sigma_n^2 = o(1)$ .

For  $M_{n5}$  and  $M_{n6}$ , first

$$E \left| \int H_{1i} H_{1j} dP(Y_1) \right|^{2(1+\delta)} \leq \frac{Ch}{n^{8(1+\delta)} h^{2(1+\delta)}} \|\xi_h\|_{2(1+\delta)}^{4(1+\delta)}.$$

Still use the result of Lemma 4, it can be obtained that

$$\frac{n^2 M_{n5}^{\frac{1}{2(1+\delta)}}}{\sigma_n^2} = O(h^{1-\frac{1}{2(1+\delta)}}) = o(1),$$

for  $\delta > 0$ . In the same fashion, it can be shown that

$$E \left| \int H_{1i} H_{1j} dP(Y_1) \right|^2 \leq C_6 n^{-8} h^{-1} \|\xi_h(x)\|_2^4,$$

and

$$\frac{n^2 M_{n6}^{\frac{1}{2}}}{\sigma_n^2} = O(h^{1/2}) = o(1).$$

Now, only the condition for  $M_{n1}$  is left for verifying. First notice that

$$E |H_{1j} H_{ij}|^{1+\delta} \leq C(n^2 h)^{-2(1+\delta)} h \|\xi_h\|_{2(1+\delta)}^{2(1+\delta)}.$$

Using the result of Lemma 4, it can be seen that  $n^2 M_{n1}^{1/(1+\delta)} / \sigma_n^2 = O(1)$ . I now use the same method to show that the asymptotic normality result still holds. As in the previous proof, one can show that

$$|EH_n(Y_1, Y_k) H_n(Y_i, Y_k)| = O_p \left( \frac{h^5}{n^4} \exp \left( \frac{\pi}{h} \right) \right), \quad (25)$$

uniformly for all  $i, k$  for kernel  $K_1$ .

Use the bound (25) in the sum  $\sum_{i=2}^{c_n}$  and the still valid inequality (21) in the sum  $\sum_{i=c_n+1}^{n-1}$ , one can obtain that

$$\begin{aligned} & \sum_{i < j < k} EH_n(Y_i, Y_k) H_n(Y_j, Y_k) \\ &= n \sum_{i=2}^{c_n} \sum_{k=i+1}^{n-1} \frac{(k-i)}{n} EH_n(Y_1, Y_k) H_n(Y_i, Y_k) + n \sum_{i=c_n+1}^{n-1} \sum_{k=i+1}^{n-1} \frac{(k-i)}{n} EH_n(Y_1, Y_k) H_n(Y_i, Y_k) \\ &\leq c_n n^2 O_p \left( \frac{h^5}{n^4} \exp \left( \frac{\pi}{h} \right) \right) + n^2 4 M_{n1}^{1/(1+\delta)} \sum_{i=c_n+1}^{n-1} \beta(i-1)^{\delta/(1+\delta)}. \end{aligned}$$

Since  $\sigma_n^2 = (4h^{12}/n^2) \exp(2\pi/h)$ , choose say  $c_n = \exp(\pi/(2h))$  the first term will be  $O_p(h^5 \exp(3\pi/(2h))/n^2)$  and it is dominated by  $\sigma_n^2$ . The second term is also dominated by  $\sigma_n^2$  by noticing that  $n^2 M_{n1}^{1/(1+\delta)} = O(1)$  and  $\sum_{i=c_n+1}^{n-1} \beta(i-1)^{\delta/(1+\delta)} \rightarrow 0$  because of the exponential decay of the beta-mixing coefficients and  $c_n = \exp(\pi/(2h)) \rightarrow \infty$ .

Combining the above results, the limiting distribution for  $S_2$  is obtained and the central limit theorem of  $T_1$  is proved. The central limit theorem for  $T_2$  follows easily by the Slutsky's theorem because  $\widehat{\|f\|_2}$  is a consistent estimator for  $\|f\|_2$ .  $\square$

## Appendix C: Technical lemmas

This section collects several results that will be useful in the proof of this paper.

The first result is the central limit theorem for U-statistics with  $\beta$ -mixing data by

Hjellvik et al. (1998). Let  $Y_1, Y_2, \dots$  be a strictly stationary  $\beta$ -mixing sequence. Assume  $H_n(x, y)$  to be a symmetric Borel function defined on  $\mathbb{R} \times \mathbb{R}$ , which may depend on sample size  $n$ . Assume further that there exists a sequence of  $\sigma$ -algebras  $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$  for which  $Y_j \in \mathcal{F}_j$ , and

$$E[H_n(x, Y_1)] = 0,$$

for any  $x \in \mathbb{R}$ , and

$$E[H_n(Y_i, Y_j) | \mathcal{F}_{j-1}] = 0, \forall i < j.$$

Let  $H_{ij} = H_n(Y_i, Y_j)$ ,  $\sigma_{ij}^2 = \text{Var}(H_{ij})$ , and  $\sigma_n^2 = \sum_{1 \leq i < j \leq n} \sigma_{ij}^2$ . For some constant  $\delta > 0$ , define

$$M_{n1} = \max_{1 < i < j \leq n} \max \left\{ E |H_{1j} H_{ij}|^{(1+\delta)}, \int |H_{1j} H_{ij}|^{(1+\delta)} dP_{Y_1} dP_{(Y_i, Y_j)} \right\},$$

$$M_{n2} = \max_{1 < i < j \leq n} \max \left\{ E |H_{1j} H_{ij}|^{2(1+\delta)}, \int |H_{1j} H_{ij}|^{2(1+\delta)} dP_{Y_1} dP_{(Y_i, Y_j)}, \right. \\ \left. \int |H_{1j} H_{ij}|^{2(1+\delta)} dP_{(Y_1, Y_i)} dP_{Y_j}, \int |H_{1j} H_{ij}|^{2(1+\delta)} dP_{Y_1} dP_{Y_i} dP_{Y_j}, \right\}$$

$$M_{n3} = \max_{1 \leq i < j \leq n} E |H_{1j} H_{ij}|^2,$$

$$M_{n4} = \max_{1 < i \neq j \neq k \leq n} \left\{ \max_P \int |H_{ij} H_{jk}|^{2(1+\delta)} dP \right\},$$

$$M_{n5} = \max_{1 < i < j} \max \left\{ E \left| \int H_{1i} H_{1j} dP_{Y_1} \right|^{2(1+\delta)}, \int \left| \int H_{1i} H_{1j} dP_{Y_1} \right|^{2(1+\delta)} dP_{Y_i} dP_{Y_j} \right\},$$

$$M_{n6} = \max_{1 < i < j} E \left| \int H_{1i} H_{1j} dP_{Y_1} \right|^2.$$

**Proposition 2** (Hjellvik et al. (1998) Theorem A in Appendix 1) *If for some  $\delta > 0$ ,*

$$\sum_{k=1}^{\infty} k^2 (\beta(k))^{1+\delta} < \infty,$$

and

$$\max \frac{1}{\sigma_n^2} \left\{ n^2 \left\{ M_{n1}^{\frac{1}{1+\delta}} + M_{n5}^{\frac{1}{2(1+\delta)}} + M_{n6}^{\frac{1}{2}} \right\}, n^{\frac{3}{2}} \left( M_{n2}^{\frac{1}{2(1+\delta)}} + M_{n3}^{\frac{1}{2}} + M_{n4}^{\frac{1}{2(1+\delta)}} \right) \right\} \rightarrow 0,$$

as  $n \rightarrow \infty$ , then

$$\sigma_n^{-1} \sum_{1 \leq i < j \leq n} H_n(Y_i, Y_j) \rightsquigarrow N(0, 1).$$

Define

$$\sigma_0^2 = \int H_n^2(Y_i, Y_j) dP(Y_i) dP(Y_j),$$

which is the variance of  $H_{ij}$  as if the data were independent,  $\sigma_n^2$  is asymptotically equivalent to  $n^2 \sigma_0^2 / 2$ .

The second result is by Yoshihara (1976), it is useful in the proof of Theorem 1. For a sequence of  $k$  random variables  $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ , with  $i_1 < i_2 < \dots < i_k$ , for any  $j (1 \leq j \leq k - 1)$ , define

$$P_j^{(k)}(B^{(j)} \times B^{(k-j)}) = P((X_{i_1}, \dots, X_{i_j}) \in B^{(j)}) P((X_{i_{j+1}}, \dots, X_{i_k}) \in B^{(k-j)})$$

and

$$P_0^{(k)}(B^{(k)}) = P((X_{i_1}, \dots, X_{i_k}) \in B^{(k)}),$$

where  $B^{(j)}$  is a Borel set in  $\mathbb{R}^j$ . These roughly imply  $P_j^{(k)}$  to be the probability measure corresponding to a  $k$ -dimensional joint distribution split in two independent groups of size  $j$  and  $(k - j)$ .

**Lemma 2** (Lemma 1 in Yoshihara (1976)) For any  $0 \leq j \leq k - 1$ , let  $h(x_{i_1}, \dots, x_{i_k})$  be a Borel function such that

$$\int \dots \int_{\mathbb{R}^k} |h(x_{i_1}, \dots, x_{i_k})|^{1+\delta} dP_j^{(k)} \leq M,$$

for some  $\delta > 0$ . Then

$$\begin{aligned} & \left| \int \dots \int_{\mathbb{R}^k} h(x_{i_1}, \dots, x_{i_k}) dP_j^{(k)} - \int \dots \int_{\mathbb{R}^k} h(x_{i_1}, \dots, x_{i_k}) dP_0^{(k)} \right| \\ & \leq 4M^{\frac{1}{1+\delta}} \beta^{\frac{\delta}{1+\delta}} (i_{j+1} - i_j). \end{aligned}$$

**Lemma 3** With the sinc kernel  $K$  defined in (7), let  $C > 0$  to be a generic large constant, it holds that for the function  $\nu_h$ ,

$$\|\nu_h\|_2 = \frac{1}{\sqrt{2\pi}} h^{1/2} \exp\left(\frac{\pi}{2h}\right).$$

$$\|\nu_h\|_\infty \leq Ch \exp\left(\frac{\pi}{2h}\right).$$

For  $p > 2$ ,

$$\|\nu_h\|_p \leq C \times h^{(p-1)/p} \exp\left(\frac{\pi}{2h}\right).$$

For the function  $\xi_h$ ,

$$\|\xi_h\|_2 = \frac{1}{2\sqrt{2\pi}} h^{1/2} \exp\left(\frac{\pi}{h}\right).$$

$$\|\xi_h\|_\infty \leq Ch \exp\left(\frac{\pi}{h}\right).$$

For  $p > 2$ ,

$$\|\xi_h\|_p \leq C' \times h^{(p-1)/p} \exp\left(\frac{\pi}{h}\right).$$

**Proof** For the  $L^2$  norm, by the Parseval's equality

$$\begin{aligned} \int \nu_h^2(x) dx &= \frac{1}{2\pi} \int |\phi_{\nu_h}(u)|^2 du \\ &= \frac{1}{2\pi} \int \left| \frac{\phi_K(u)}{\phi_k(u/h)} \right|^2 du \\ &= \frac{1}{2\pi} \int \left| \frac{I(|u| \leq 1)}{\sqrt{2} \exp(-\pi|u|/(2h))} \right|^2 du \\ &= \frac{1}{2\pi^2} h \exp\left(\frac{\pi}{h}\right), \end{aligned}$$

and the first result for  $\nu_h$  is proved.

For the  $L_\infty$  norm, notice that

$$\begin{aligned} \|\nu_h\|_\infty &= \sup_x \left| \frac{1}{2\pi} \int \frac{\phi_K(t)}{\phi_k(t/h)} e^{-itx} dt \right| \\ &\leq \frac{1}{2\pi} \int \left| \frac{\phi_K(t)}{\phi_k(t/h)} \right| dt \\ &\leq \frac{\sqrt{2}}{\pi^2} h \exp\left(\frac{\pi}{2h}\right), \end{aligned}$$

and the second result for  $\nu_h$  is proved.

For  $p > 2$ , it is known from Van Es et al. (2003) and Masry (1991) that

$$\|\nu_h\|_p \leq \|\nu_h\|_\infty^{1-2/p} \|\nu_h\|_2^{2/p}.$$

This is simple to show by noticing  $\int |\nu_h(x)|^p dx \leq \int |\nu_h(x)|^2 \sup_x |\nu_h(x)|^{p-2} dx$ . Therefore,

$$\begin{aligned} \|\nu_h\|_p &\leq \|\nu_h\|_\infty^{(p-2)/p} \|\nu_h\|_2^{2/p} \\ &\leq C \times h^{(p-2)/p} \exp\left(\frac{\pi(p-2)/p}{2h}\right) \times h^{1/p} \exp\left(\frac{\pi(2/p)}{h}\right) \\ &= C \times h^{(p-1)/p} \exp\left(\frac{\pi}{2h}\right), \end{aligned}$$

and the last result for  $\nu_h$  is proved.

For the  $L^2$  norm of  $\xi_h$ , using Parseval's equality,

$$\begin{aligned}
\int \xi_h^2(x) dx &= \frac{1}{2\pi} \int |\phi_{\xi_h}(t)|^2 dt \\
&= \frac{1}{2\pi} \int |\phi_{\nu_h}(t)\phi_{\nu_h}(-t)|^2 dt \\
&= \frac{1}{2\pi} \int_{|u| \leq 1} \frac{1}{|\phi_k(u/h)|^2 |\phi_k(-u/h)|^2} du \\
&= \frac{1}{8\pi^2} h \exp\left(\frac{2\pi}{h}\right),
\end{aligned}$$

and the first result for  $\xi_h$  is proved.

The upper  $L_\infty$  norm of  $\xi_h(x)$  can be evaluated with the help of Young's inequality for the convolution of two functions, recall that  $\xi_h(x)$  is the convolution of  $\nu_h(x)$  with itself.

$$\begin{aligned}
\|\xi_h\|_\infty &\leq C \|\nu_h\|_2^2 \\
&= C \times h \exp\left(\frac{\pi}{h}\right),
\end{aligned}$$

and the second result for  $\xi_h$  is proved.

For the upper bound of  $L^p$  norm with  $p > 2$ , still use the inequality

$$\begin{aligned}
\|\xi_h\|_p &\leq \|\xi_h\|_\infty^{(p-2)/p} \|\xi_h\|_2^{2/p} \\
&\leq C h^{1-1/p} \exp\left(\frac{\pi}{h}\right),
\end{aligned}$$

and the last result for  $\xi_h$  is proved. □

**Lemma 4** *With the kernel  $K_1$  defined in (8), let  $C > 0$  be a generic large constant, it holds that:*

$$\begin{aligned}
\|\nu_h\|_2 &= \sqrt{\frac{23040}{\pi^8}} \exp\left(\frac{\pi}{2h}\right) h^{7/2}, \\
\|\nu_h\|_\infty &\leq C \exp\left(\frac{\pi}{2h}\right) h^4, \\
\|\nu_h\|_p &\leq C \exp\left(\frac{\pi}{2h}\right) h^{4-1/p}, \\
\|\xi_h\|_2 &= \frac{360\sqrt{462}}{\pi^7} \exp\left(\frac{\pi}{h}\right) h^{13/2}, \\
\|\xi_h\|_\infty &\leq C \exp\left(\frac{\pi}{h}\right) h^7, \\
\|\xi_h\|_p &\leq C \exp\left(\frac{\pi}{h}\right) h^{7-1/p},
\end{aligned}$$

with  $p > 2$ .

**Proof** For the  $L^2$  norm of  $\nu_h$ , use the Parseval's equality

$$\begin{aligned}\int \nu_h^2(x)dx &= \frac{1}{2\pi} \int \left| \frac{\phi_K(u)}{\phi_k(u/h)} \right|^2 du \\ &= \frac{1}{2\pi} \int \left| \frac{I(|u| \leq 1)(1-u^2)^3}{\sqrt{2} \exp(-\pi|u|/(2h))} \right|^2 du \\ &= \frac{23040}{\pi^8} \exp\left(\frac{\pi}{h}\right) h^7,\end{aligned}$$

and the  $L^2$  norm of  $\nu_h$  is obtained.

$$\begin{aligned}\|\nu_h\|_\infty &\leq \frac{1}{2\pi} \int \left| \frac{\phi_K(u)}{\phi_k(u/h)} \right| du \\ &= \frac{1}{2\pi} \int \left| \frac{I(|u| \leq 1)(1-u^2)^3}{\sqrt{2} \exp(-\pi|u|/(2h))} \right| du \\ &= C \times \exp\left(\frac{\pi}{2h}\right) h^4,\end{aligned}$$

and the upper bound for the  $L_\infty$  norm of  $\nu_h$  is obtained.

Then for  $p > 2$ , since

$$\begin{aligned}\|\nu_h\|_p &\leq \|\nu_h\|_\infty^{1-2/p} \|\nu_h\|_2^{2/p} \\ &= C \times \left(\exp\left(\frac{\pi}{2h}\right) h^4\right)^{1-2/p} \left(\exp\left(\frac{\pi}{2h}\right) h^{7/2}\right)^{2/p} \\ &= C \times \exp\left(\frac{\pi}{2h}\right) h^{4-1/p},\end{aligned}$$

and the upper bound for the  $L_p$  norm ( $p > 2$ ) of  $\nu_h$  is obtained.

For the  $L^2$  norm of  $\xi_h$ , still use the Parseval's equality

$$\begin{aligned}\int \xi_h^2(x)dx &= \frac{1}{2\pi} \int |\phi_{\xi_h}(t)|^2 dt \\ &= \frac{1}{2\pi} \int |\phi_{\nu_h}(t)\phi_{\nu_h}(-t)|^2 dt \\ &= \frac{1}{2\pi} \int \left| \frac{I(|u| \leq 1)(1-u^2)^3}{\sqrt{2} \exp(-\pi|u|/(2h))} \right|^4 du \\ &= \frac{467775 \times 128}{\pi^{14}} \exp\left(\frac{2\pi}{h}\right) h^{13},\end{aligned}$$

and the  $L^2$  norm of  $\xi_h$  is obtained.

With the help of Young's inequality for the integral of convoluted functions, one has

$$\begin{aligned}\|\xi_h\|_\infty &\leq C \|\nu_h\|_2^2 \\ &= C \exp\left(\frac{\pi}{h}\right) h^7,\end{aligned}$$

and the upper bound for the  $L_\infty$  norm of  $\xi_h$  is obtained.

Then for  $p > 2$ , one has

$$\begin{aligned}\|\xi_h\|_p &\leq \|\xi_h\|_\infty^{1-2/p} \|\xi_h\|_2^{2/p} \\ &= C \left( \exp\left(\frac{\pi}{h}\right) h^7 \right)^{1-2/p} \left( \exp\left(\frac{\pi}{h}\right) h^{13/2} \right)^{2/p} \\ &= C \exp\left(\frac{\pi}{h}\right) h^{7-1/p},\end{aligned}$$

and the upper bound for the  $L_p$  norm ( $p > 2$ ) of  $\xi_h$  is obtained. □

## Appendix D: A heuristic approximation for the density $g(x)$ : small- $\Delta$ approximation

In this appendix, a heuristic approximation method for the parametric estimation of the density  $g(x)$  is proposed. To be specific, the stationary volatility density function  $\pi(\cdot)$  is used to approximate the integrated volatility density function of  $\eta_i = \int_{t_{i-1}}^{t_i} \sigma_s^2 ds / \Delta$ . Applying the standard change of variable formula to the log transformation, the stationary density of  $\log \eta_i$  is then approximated by  $e^x \pi(e^x)$  — this is defined as the small- $\Delta$  approximation for  $g(x)$  and can be denoted as follows:

$$g^{\text{sd}}(x) := e^x \pi(e^x).$$

The formula for calculating  $\pi(x)$  from the functions  $a(x)$  and  $b(x)$  are given in (12) in Appendix A.

To give two examples. The GARCH diffusion model, which has the volatility process specified as:

$$d\sigma_t^2 = \alpha(\beta - \sigma_t^2)dt + \gamma\sigma_t^2 dW_t.$$

has the stationary distribution of the volatility process exists when  $2\alpha/\gamma^2 > -1$ , it is an inverse Gamma distribution with the following density function:

$$\pi(u) = \frac{\lambda^\iota}{\Gamma(\iota)} u^{-\iota-1} \exp\left(-\frac{\lambda}{u}\right) \mathbb{1}_{\{u>0\}},$$

where  $\iota = 1 + 2\alpha/\gamma^2$  and  $\lambda = 2\alpha\beta/\gamma^2$ .

The Heston model, which has the volatility process specified as:

$$d\sigma_t^2 = \alpha(\beta - \sigma_t^2)dt + \gamma\sqrt{\sigma_t^2} dB_t.$$

has the stationary distribution exists when  $2\alpha\beta/\gamma^2 > 1$ , it is a Gamma distribution with

the following density function:

$$\pi(u) = \frac{\lambda^\iota}{\Gamma(\iota)} u^{\iota-1} e^{-\lambda u} \mathbb{1}_{\{u>0\}},$$

where  $\iota = 2\beta\alpha/\gamma^2$  and  $\lambda = 2\alpha/\gamma^2$ .

Because  $\int_{t_{i-1}}^{t_i} \sigma_s^2 ds/\Delta$  involves integrating the volatility path over a small interval and performing normalization, the density of  $\eta_i$  is expected to close to the stationary log density of the volatility process. The performance of this approximation is studied in parallel with the simulation based tests in Monte Carlo simulations in Section 6. It is found that the performance of the small- $\Delta$  approximation based tests is very similar to that of the simulation based tests for all the scenarios considered. Studying the theoretical properties of this approximation error is beyond the scope of this paper and thus left for future research.

## References

- Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9, 385–426.
- Aït-Sahalia, Y., P. J. Bickel, and T. M. Stoker (2001). Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics* 105(2), 363–412.
- Aït-Sahalia, Y., J. Fan, and J. Jiang (2010). Nonparametric tests of the markov hypothesis in continuous-time models. *The Annals of Statistics* 38(5), 3129–3163.
- Aït-Sahalia, Y., J. Fan, and H. Peng (2009). Nonparametric transition-based tests for jump diffusions. *Journal of the American Statistical Association* 104, 1102–1116.
- Aït-Sahalia, Y. and J. Y. Park (2012). Stationarity-based specification tests for diffusions when the process is nonstationary. *Journal of Econometrics* 169(2), 279–292.
- Andersen, T. G., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time equity return models. *The Journal of Finance* 57(3), 1239–1284.
- Andrews, D. W. (1997). A conditional kolmogorov test. *Econometrica: Journal of the Econometric Society*, 1097–1128.
- Bickel, P. J. and M. Rosenblatt (1973). On some global measures of the deviations of density function estimates. *Annals of Statistics* 1, 1071–1095.
- Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* 2, 107–144.
- Butucea, C. (2004). Asymptotic normality of the integrated square error of a density estimator in the convolution model. *Sort* 28, 9–26.
- Butucea, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Annals of Statistics* 35, 1907–1930.

- Carrasco, M. and X. Chen (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.
- Carroll, R. and P. Hall (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* 83, 1184–1186.
- Chen, S., J. Gao, and C. Tang (2008). A test for model specification of diffusion processes. *Annals of Statistics* 36, 167–198.
- Chen, X., L. P. Hansen, and M. Carrasco (2010). Nonlinearity and temporal dependence. *Journal of Econometrics* 155, 155 – 169.
- Comte, F. (2004). Kernel deconvolution of stochastic volatility models. *Journal of Time Series Analysis* 25(4), 563–582.
- Comte, F., V. Genon-Catalot, and Y. Rozenholc (2010). Nonparametric estimation for a stochastic volatility model. *Finance and Stochastics* 14, 49–80.
- Corradi, V. and W. Distaso (2006). Semi-parametric comparison of stochastic volatility models using realized measures. *Review of Economic Studies* 73, 635–667.
- Corradi, V. and N. Swanson (2005). Bootstrap specification tests for diffusion processes. *Journal of Econometrics* 124, 117–148.
- Corradi, V. and N. R. Swanson (2011). Predictive density construction and accuracy testing with multiple possibly misspecified diffusion models. *Journal of Econometrics* 161, 304 – 324.
- Delaigle, A. and I. Gijbels (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Computational statistics & data analysis* 45(2), 249–267.
- Delaigle, A. and I. Gijbels (2007). Frequent problems in calculating integrals and optimizing objective functions: a case study in density deconvolution. *Statistics and Computing* 17(4), 349–355.
- Delaigle, A. and P. Hall (2010). Kernel methods and minimum contrast estimators for empirical deconvolution. *Arxiv preprint arXiv:1003.0315*.
- Efron, B. and R. J. Tibshirani (1994). *An introduction to the bootstrap*, Volume 57. CRC press.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics* 19, 1257–1272.
- Fan, J. and Q. Yao (2002). *Nonlinear time series*, Volume 2. Springer.
- Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theory* 10, 316–356.
- Fan, Y. (1995). Bootstrapping a consistent nonparametric goodness-of-fit test. *Econometric Reviews* 14(3), 367–382.
- Fan, Y. and Q. Li (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics* 10(3), 245–271.
- Fan, Y. and A. Ullah (1999). On goodness-of-fit tests for weakly dependent processes

- using kernel method. *Journal of Nonparametric Statistics* 11, 337–360.
- Franke, J., W. Härdle, and J.-P. Kreiß (2003). Nonparametric estimation in a stochastic volatility model. In M. G. Akritas and D. N. Politis (Eds.), *Recent Advances and Trends in Nonparametric Statistics*, pp. 302–313. Amsterdam: North Holland.
- Franke, J., J.-P. Kreiss, and E. Mammen (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* 8(1), 1–37.
- Gallant, A. R., D. Hsieh, and G. Tauchen (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics* 81, 159–192.
- Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric Theory* 12, 657–681.
- Gao, J. and I. Gijbels (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association* 103, 1584–1594.
- Gao, J. and M. King (2004). Adaptive testing in continuous-time diffusion models. *Econometric Theory* 20, 844–882.
- Genon-Catalot, V., T. Jeantheau, and C. Laredo (1998). Limit theorems for discretely observed stochastic volatility models. *Bernoulli* 4, 283–303.
- Genon-Catalot, V., T. Jeantheau, and C. Laredo (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* 6, 1051–1079.
- Glad, I. K., N. L. Hjort, and N. Ushakov (2007). Density estimation using the sinc kernel. *Preprint Statistics* (2).
- Glasserman, P. (2004). *Monte Carlo methods in financial engineering*, Volume 53. Springer.
- Hall, P. and J. S. Marron (1987). Estimation of integrated squared density derivatives. *Statistics & Probability Letters* 6(2), 109–115.
- Härdle, W. and E. Mammen (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1926.
- Hjellvik, V., Q. Yao, and D. Tjøstheim (1998). Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* 68, 295–321.
- Holzmann, H., N. Bissantz, and A. Munk (2007). Density testing in a contaminated sample. *Journal of Multivariate Analysis* 98, 57–75.
- Hong, Y. and H. Li (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies* 18, 37.
- Kanaya, S. and D. Kristensen (2015). Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory*.
- Kloeden, P. E. and E. Platen (1992). *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer.
- Lambert, P. and S. Laurent (2001). Modelling financial time series using garch-type models with a skewed student distribution for the innovations. Technical report, Dis-

- cussion paper 0125, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
- Masry, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Transactions on Information Theory* 37, 1105–1115.
- Masry, E. (1993). Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *Journal of Multivariate Analysis* 44(1), 47–68.
- Meddahi, N. (2002). Moments of continuous time stochastic volatility models. *Unpublished paper: University of Montreal*.
- Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. Berlin: Springer.
- Neumann, M. H. and E. Paparoditis (2000). On bootstrapping l2-type statistics in density testing. *Statistics & Probability Letters* 50(2), 137 – 147.
- Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics*. Cambridge: Cambridge University Press.
- Pham, T. D. and L. T. Tran (1985). Some mixing properties of time series models. *Stochastic Processes and Their Applications* 19, 297–303.
- Reno, R. (2006). Nonparametric estimation of stochastic volatility models. *Economics Letters* 90, 390–395.
- Reno, R. (2008). Nonparametric estimation of the diffusion coefficient of stochastic volatility models. *Econometric Theory* 24, 1174–1206.
- Stefanski, L. and R. Carroll (1990). Deconvoluting kernel density estimators. *Statistics* 21, 169–184.
- Van der Sluis, P. J. (1997). EmmPack 1.01: C/C++ code for use with Ox for estimation of univariate stochastic volatility models with the efficient method of moments. *Studies in Nonlinear Dynamics and Econometrics* 2, 77–94.
- Van Es, B. and P. Spreij (2011). Estimation of a multivariate stochastic volatility density by kernel deconvolution. *Journal of Multivariate Analysis* 102(3), 683 – 697.
- Van Es, B., P. Spreij, and H. Van Zanten (2003). Nonparametric volatility density estimation. *Bernoulli* 9, 451–465.
- Van Es, B., P. Spreij, and H. Van Zanten (2005). Nonparametric volatility density estimation for discrete time models. *Nonparametric Statistics* 17, 237–251.
- Wasserman, L. (2004). *All of statistics: a concise course in statistical inference*. Springer.
- Yoshihara, K. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Probability Theory and Related Fields* 35, 237–252.
- Zu, Y. and H. P. Boswijk (2009). Nonparametric return based tests for stochastic volatility models. Technical report, University of Amsterdam.

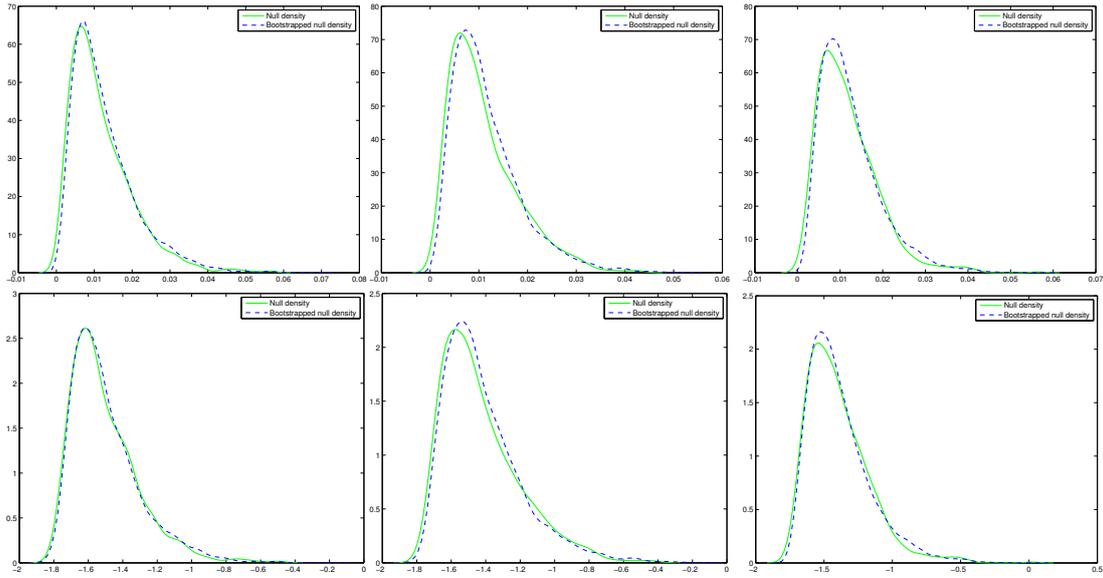


Figure 1: Null distribution and bootstrap null distribution of  $T_1$  (top panel) and  $T_2$  (bottom panel), for 5 years (left), 10 years (middle) and 15 years (right) of daily observations. Parameter value  $\alpha = 10$ ,  $\beta = -3$ ,  $\gamma = 3$ .

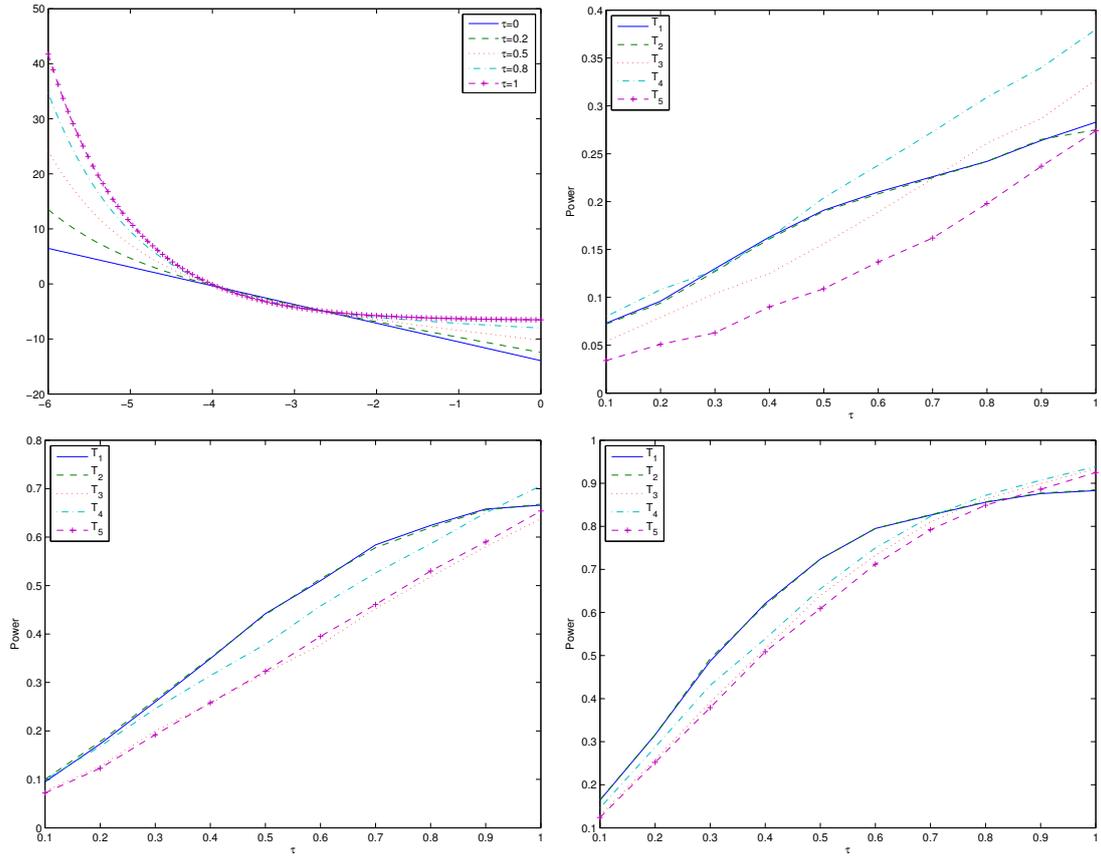


Figure 2: Top left panel: drift functions in (10) for different value of  $\tau$ . Top right panel: Power functions for the five competing test statistics evaluated at the sequence of alternative models (10), 5 years of daily observations. Bottom left panel: Power functions for the five competing test statistics evaluated at the sequence of alternative models (10), 10 years of daily observations. Bottom right panel: Power functions for the five competing test statistics evaluated at the sequence of alternative models (10), 15 years of daily observations.

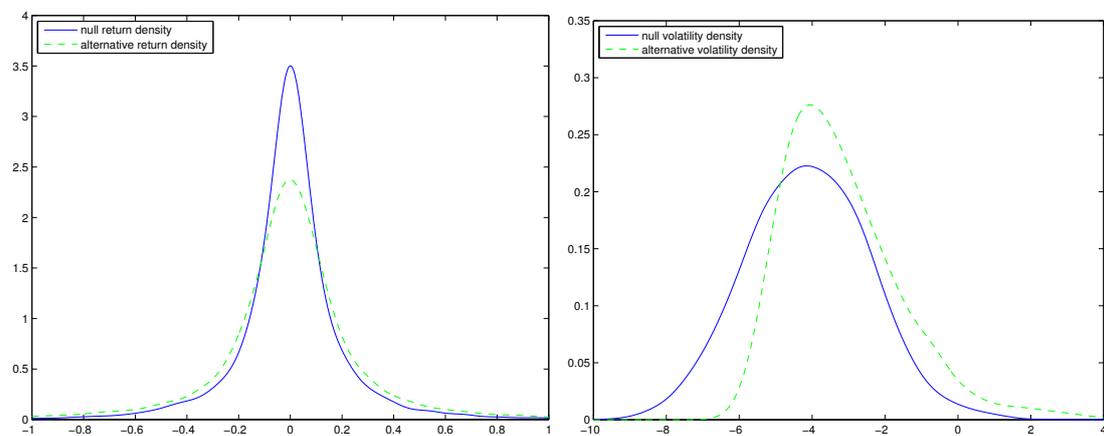


Figure 3: Left panel: Return density under the null model and the alternative model (10) for  $\tau = 1$ . Right panel: Volatility density under the null model and the alternative model (10) for  $\tau = 1$ .

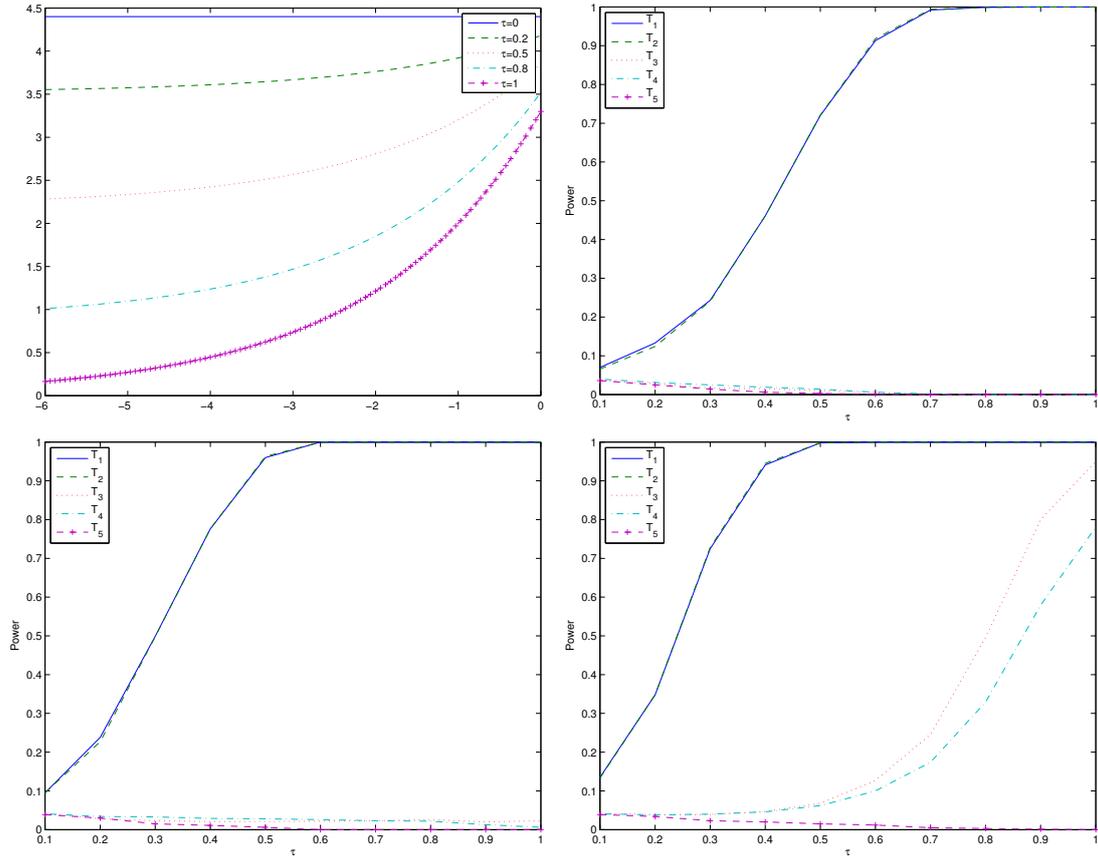


Figure 4: Top left panel: diffusion functions in (11) for different value of  $\tau$ . Top right panel: Power functions for the five competing test statistics evaluated at the sequence of alternative models (11), 5 years of daily observations. Bottom left panel: Power functions for the five competing test statistics evaluated at the sequence of alternative models (11), 10 years of daily observations. Bottom right panel: Power functions for the five competing test statistics evaluated at the sequence of alternative models (11), 15 years of daily observations.

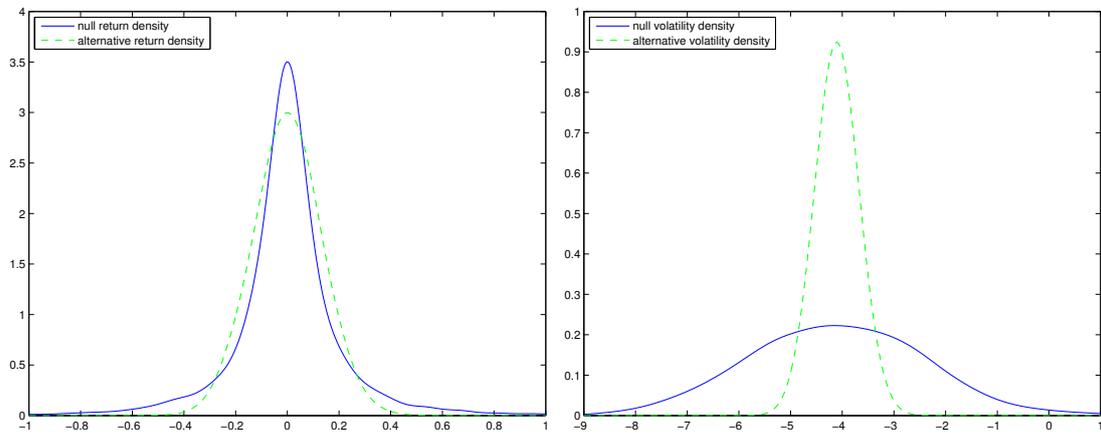


Figure 5: Left panel: Return density under the null model and the alternative model (11) for  $\tau = 1$ . Right panel: Volatility density under the null model and the alternative model (11) for  $\tau = 1$ .

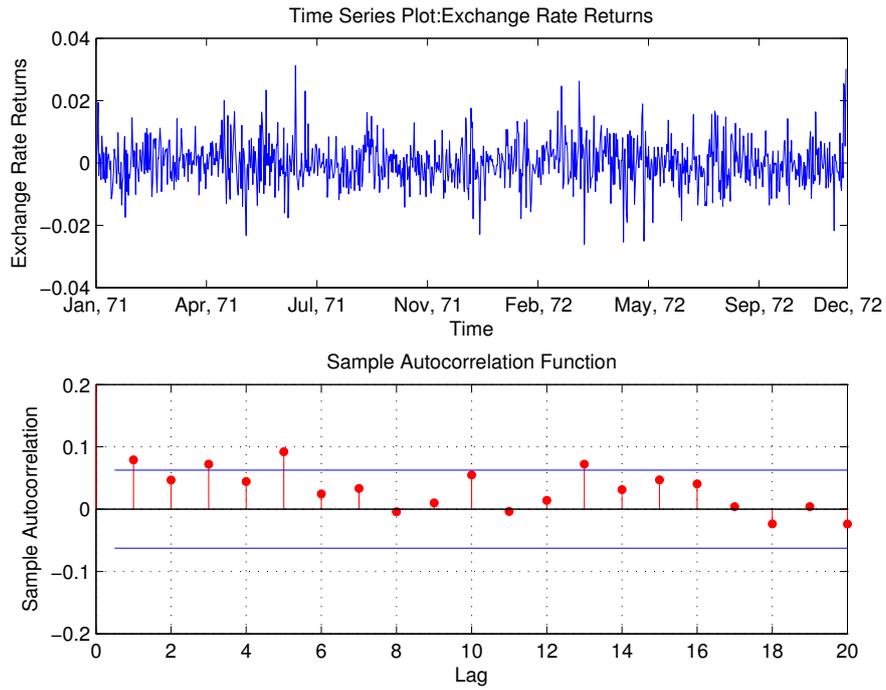


Figure 6: Time series plot and the Autocorrelation Function of the squared returns of daily British pound/Canadian dollar exchange rate from January 1971 to August 1996.

	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$
$n = 5 \times 252$					
<b>10% level</b>	0.1126	0.1082	0.1194	0.1190	0.1206
<b>5% level</b>	0.0612	0.0570	0.0704	0.0714	0.0684
<b>1% level</b>	0.0090	0.0084	0.0108	0.0106	0.0044
$n = 10 \times 252$					
<b>10% level</b>	0.1020	0.0982	0.1222	0.1208	0.1242
<b>5% level</b>	0.0536	0.0528	0.0738	0.0724	0.0628
<b>1% level</b>	0.0184	0.0188	0.0202	0.0206	0.0084
$n = 15 \times 252$					
<b>10% level</b>	0.1204	0.1156	0.1318	0.1290	0.1248
<b>5% level</b>	0.0684	0.0646	0.0722	0.0714	0.0612
<b>1% level</b>	0.0086	0.0084	0.0266	0.0270	0.0080

Table 1: Simulated size with parameter value  $\alpha = 10$ ,  $\beta = -3$ ,  $\gamma = 3$ .

	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$
<b>p value</b>	0.0030	0.0050	0.0010	0.0010	0.0660

Table 2:  $p$ -value of the test statistics