Essays on nonparametric econometrics of stochastic volatility

Zu, Y.

Publication date
2012

Chapter 4

Nonparametric specification tests for stochastic volatility models based on the volatility density

4.1 Introduction

Consider the following stochastic volatility model,

$$\mathrm{d}Y_t = \sigma_t \mathrm{d}B_t, \tag{4.1}$$

$$\mathrm{d}\sigma_t^2 = b(\sigma_t^2)\mathrm{d}t + a(\sigma_t^2)\mathrm{d}W_t, \tag{4.2}$$

where B and W are two independent standard Brownian motion processes. σ^2 is the unobservable diffusion parameter of the Y process, which is assumed to follow a stationary diffusion process. Y is assumed to be observed discretely at $t_i = i\Delta$, $i = 0, 1, \ldots, n$ and Δ is assumed to be fixed.

(4.1)–(4.2) is often used to model stock prices in financial econometrics. Y is often the efficient logarithmic price, and σ^2 is the *latent* volatility process. Specifying a stochastic structure for the volatility process allows the model to reproduce important stylized facts observed in realistic financial returns, such as fat-tailed return distributions and volatility clustering (Shephard (1996)).

Model (4.1)–(4.2) includes many popular models such as the Heston model (Heston (1993)) and the GARCH diffusion model (Nelson (1990)), among others. Although significant progress has been made in estimating stochastic volatility models in the past decade, specification testing for stochastic volatility models is still an underdeveloped area. As

an alternative to the previous chapter, this chapter provide another approach for such testing problem.

Assume that in the stochastic volatility model (4.1)–(4.2), the true process is characterized by the functions $\{b_0(.), a_0(.)\}$, and let a parameterization of the model be

$$\{b(.;\theta), a(,;\theta), \theta \in \Theta \subseteq \mathbb{R}^k\}. \tag{4.3}$$

As the previous chapter, this chapter studies the problem of testing the null hypothesis

$$\mathcal{H}_0 : \{\exists \theta_0 \in \Theta, b(.;\theta_0) = b_0(.), a(.;\theta_0) = a_0(.)\}$$

against the alternative

$$\mathcal{H}_1 : \{b(.;\theta) \neq b_0(.), a(.;\theta_0) \neq a_0(.), \forall \theta \in \Theta\}.$$

If σ^2 were observable (in discrete time), this problem would be reduced to a specification test problem for diffusion process. For example, Aït-Sahalia (1996) proposes a specification test statistic for diffusion processes by comparing the nonparametric kernel density estimate of the stationary density of the process with its parametric counterpart. Letting $\pi(.)$ be the stationary density of the volatility process, the test statistic is

$$T_n = \sum_{i=1}^{n} \left(\hat{\pi}\left(\sigma_i^2\right) - \pi\left(\sigma_i^2; \hat{\theta}_n\right) \right)^2, \tag{4.4}$$

where $\hat{\pi}(x) = \sum_{i=1}^{n} K\left(\left(x - \sigma_i^2\right)/h\right)/(nh)$ is the nonparametric kernel density estimator of the volatility, and

$$\pi(x;\theta) = \exp\left(2 \int_{x_0}^{x} \frac{b(u;\theta)}{a^2(u;\theta)} \mathrm{d}u \right) \frac{1}{a^2(x;\theta)} \frac{1}{M} \tag{4.5}$$

is the stationary volatility density under the null hypothesis, where $x_0 > 0$, M is the normalizing constant to make π a probability density function, and $\hat{\theta}_n$ is an estimator of the parameters. The sum in (4.4) is over the observations. However, the unobservability of σ^2 in the stochastic volatility model (4.1)–(4.2) makes this test not applicable to the problem in this chapter.

Although the volatility process is not observable in model (4.1)–(4.2), 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 return data (first difference of the observed logarithmic prices). This is because the discretized stochastic

volatility model can be rewritten into a convolution model. Let the logarithmic price Y be observed at equally Δ-spaced times $t_0, \ldots, t_n$, and define the return sequence[1]

$$\begin{aligned} y_i &:= \frac{1}{\Delta^{1/2}}\left(Y_{t_i} - Y_{t_{i-1}}\right) \\ &= \frac{1}{\Delta^{1/2}}\int_{t_{i-1}}^{t_i} \sigma_s \mathrm{d}B_s \\ &\sim \left(\frac{1}{\Delta}\int_{t_{i-1}}^{t_i} \sigma_s \mathrm{d}s\right)^{1/2} \times \frac{1}{\sqrt{\Delta}}\left(B_{t_i} - B_{t_{i-1}}\right), \end{aligned}$$

for $i = 1, \ldots, n$. Denote $\eta_i := \int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s/\Delta$ and ε_i to be independent and identically distributed standard normal variables for each i. Then the above equation can be written as

$$y_i = \eta_i^{1/2}\varepsilon_i, i = 1, \ldots, n. \tag{4.6}$$

Taking squares and logarithms on both sides,

$$\log y_i^2 = \log \eta_i + \log \varepsilon_i^2, i = 1, \ldots, n, \tag{4.7}$$

such that the variable $\log y_i^2$ is the convolution of $\log \eta_i$ with a completely known distribution, called the logarithmic chi-square distribution. Notice here the distribution of η_i is not the same as the stationary density of the volatility process.

In statistics, model (4.7) is known as a measurement error model — the signal $\log \eta_i$ of interest 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 $\log y_i^2$, $i = 1, \ldots, n$, is called deconvolution, which can be done in several ways, see Meister (2009) for a review. Denote the density function of $\log \eta_i$ as $g(x)$. Van Es et al. (2003) use a kernel type deconvolution estimator to estimate $g(x)$ as follows

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

where ϕ_K and ϕ_k are defined in Section 2. Letting $n \to \infty$ and $\Delta \to 0$ at the same time, these authors show that $\hat{g}(x)$ is a consistent estimator for the stationary volatility density. Heuristically, when $n \to \infty$, $\hat{g}(x)$ will converge to the density function of $\log \eta_i = \log\left(\int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s/\Delta\right)$, and when $\Delta \to 0$, this density function will become the stationary density of the volatility process (after appropriate change of variables).

Motivated by these results, this chapter proposes to test the hypothesis $\mathcal{H}_0$ against $\mathcal{H}_1$ by comparing the parametric estimate and the nonparametric kernel deconvolution

[1]Technical conditions on this sequence are discussed in Appendix 4.A.

estimate of the stationary density function of $\log \eta_i$. Notice that the context in this chapter will be different from that of Van Es et al. (2003), because the sampling interval Δ in this chapter is fixed — such that the object of comparing is not exactly the stationary volatility density, but the density function of $\log \eta_i := \log\left(\int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s/\Delta\right)$, the *integrated volatility density*. Denote $g_0(x)$ as the true integrated volatility density function, and $g(x;\theta)$ as the parametric integrated volatility density function. The test statistic I_0 is formulated by calculating the L_2 distance between $\hat{g}(x)$ and $g(x;\hat{\theta})$,

$$I_0 = \int_{\mathbb{R}} \left(\hat{g}(x) - g(x;\hat{\theta})\right)^2 \mathrm{d}x \tag{4.8}$$

and what is essentially tested is,

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

As discussed in the previous Chapter, the reformulated testing problem is not equivalent to the original one, which test $\mathcal{H}_0$ against $\mathcal{H}_1$ - there are certain deviations in $\mathcal{H}_1$ can not be detected by the reformulated testing problem. However, testing the volatility density itself also consists of an important problem to investigate.

The idea of comparing parametric and nonparametric estimates for specification testing is not new, see Bickel and Rosenblatt (1973), Härdle and Mammen (1993), among others for early contributions. For the case of weakly dependent data, see Fan (1994), Fan and Ullah (1999) among others. Nonparametric specification testing for diffusion process was investigated by Aït-Sahalia (1996), Hong and Li (2005), Corradi and Swanson (2005) and Chen, Gao, and Tang (2008), among others. Specification testing involving a kernel deconvolution estimator is relatively new, see Butucea (2007) and Holzmann, Bissantz, and Munk (2007).

The test studied in this chapter is not the only way to test the specification of stochastic volatility models. As already discussed in the previous chapter, early specification tests for stochastic volatility are mostly based on the moment restrictions used in estimating the model (such as Gallant et al. (1997)), which are not able to test the correctness of the full parametric specification as in the present chapter. Recent contributions include Corradi and Distaso (2006), who propose to test the specification of stochastic volatility models based on the moment information of realized volatility measures; and Corradi and Swanson (2011), who propose to base the test on one-step predictive density of observed returns. Zu and Boswijk (2009) (see also Chapter 2 of this thesis) propose tests which compare the nonparametric and parametric estimates of the density function and distribution function of the return sequence (y_i).

The reason to consider the test based on volatility density is, when one is concerned with the misspecification about the volatility process, it is more natural to formulate the test based on the volatility process itself. Local deviations in the volatility density could be smoothed out and not easily detectable if one only looks at the distributional properties of returns, which is a convolution of the volatility density with another random variable. This is related to some known results. In a similar context of specification testing for diffusion models, Kristensen (2011) shows that deviations in the drift and diffusion functions are in general not easily detected when the test statistic is formulated based on a function (transition density in his case) that involves integration of the drift and diffusion functions, because these deviations are likely to be smoothed out. In the literature of nonparametric tests with indirect observations, Holzmann, Bissantz, and Munk (2007) notice that "*the deconvolution problem is ill-posed, i.e. the inverse of the convolution operator is unbounded. Thus it can happen that the true* f_0 *is at an arbitrarily large* L_2*-distance to the parametric model in the domain of* f*, whereas, the corresponding* $g_0 = f_0 * \psi$ *is very close to the parametric model in the domain of* g*. Hence direct application of tests to the observable data ... will result in an inefficient procedure for those alternatives which can hardly be distinguished from the null in terms of* g*.*" In the present context, this means that two volatility densities with large L_2 distance will be very close to each other after the convolution operation. In this case, it will be more efficient to construct a test statistic based on the integrated volatility density, rather than the return density after convolution.

A weak point of the test in this chapter is, it relies on the (implied) assumption that the distribution of ε_t is standard normal, it cannot test the misspecification for this distribution. However, the test developed in Chapter 3 is able to test the misspecification in the volatility process $\{\sigma_t\}$ as well as the distribution of ε_t simultaneously. The tests studied in Chapter 3 (also Zu and Boswijk (2009)) and the current chapter will be useful complements to each other.

The test in this chapter is developed entirely from a statistical point of view. Stochastic volatility models are proposed in the context of financial modeling, *e.g.* for option pricing models. In that case, more aspects need to be considered than the goodness of fit, such as model simplicity and tractability, and sometimes these are more important in practical financial modeling. So it is important that the test developed in this chapter be interpreted from a statistical point of view, and it should be used with caution to invalidate a stochastic volatility model.

It is remarked further that the tests developed in Chapter 3 and Chapter 4 do not

need high-frequency data, which is reflected in the assumption of fixed sampling interval Δ, and the use of long-span asymptotics; while high-frequency data literature often use in-fill asymptotics. However, these tests can be applied to intraday high-frequency data (for example, 5-minute data), as long as the market microstructure noise is not an issue.[2]

The plan of the chapter is as follows. Deconvolution kernel density estimation for the integrated volatility density is discussed in Section 4.2. The parametric estimate of the integrated volatility density is discussed in Section 4.3. The integrated volatility density usually doesn't have an explicit formula, this section also discusses methods of approximating the integrated volatility density. Test statistic, its asymptotic null distribution, and the asymptotic power of the test under fixed and local alternatives are studied in Section 4.4, a parametric bootstrap procedure is also proposed in this section to obtain the finite sample null distribution. Section 4.5 discusses some possible extensions to discrete time models. Section 4.6 performs Monte Carlo simulations, where the performance of the approximation methods of integrated volatility density and the power of the test are studied in finite samples. The test is applied to a real example in Section 4.7. Section 4.8 concludes. Technical assumptions are stated in Appendix 4.A, the proofs are collected in Appendix 4.B, and some (known) technical lemmas used often in the proofs are also collected in Appendix 4.B for reference.

In this chapter, we follow the convention to denote $\phi_g(t) = \int_{-\infty}^{\infty} e^{\mathrm{i}tx} g(x)\mathrm{d}x$ as the Fourier transform of function $g(x)$, thus its inverse Fourier transform is $g(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} e^{-\mathrm{i}tx}\phi_g(t)\mathrm{d}t$. We use $\rightsquigarrow$ to denote convergence in distribution and $\sim$ to denote equality in distribution. Other notation is standard.

4.2 Nonparametric integrated volatility density

Recall from equation (4.7) that

$$\log y_i^2 = \log \eta_i + \log \varepsilon_i^2, \quad i = 1, \ldots, n.$$

Assume that the density functions of $\log y_i^2$, $\log \eta_i$ and $\log \varepsilon_i^2$ exist, denoted as $f(.)$, $g(.)$, and $k(.)$, respectively. Equation (4.7) also implies the following convolution relationship,

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

[2]With high-frequency data, alternative test statistics could be proposed and studied in the future. For example, using the methods developed recently in Kanaya and Kristensen (2010), nonparametric estimates of the drift and diffusion functions of the volatility process could be obtained, and test statistics could be formulated based on comparing them with their parametric counterparts respectively.

where $g * k(x) = \int g(x-y)k(y)\mathrm{d}y$ denotes convolution of $g(.)$ and $k(.)$.

When $\log y_i^2$, $i = 1, \ldots, n$ are observable, and the distributions of $\log \varepsilon_i^2$, $i = 1, \ldots, n$ are i.i.d. and fully known, the density function $g(x)$ can be estimated by the kernel density deconvolution estimator,

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

where we denote $X_j := \log y_j^2$, ϕ_K is the Fourier transform of a kernel function K and $\phi_k(t)$ is the characteristic function of logarithmic chi-square distribution. The kernel deconvolution estimator was first proposed for a measurement error model by Carroll and Hall (1988) and Stefanski and Carroll (1990).

If we define

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

the estimator can be written in kernel form

$$\hat{g}(x) = \frac{1}{nh}\sum_{j=1}^{n}\nu_h\left(\frac{x-X_j}{h}\right),$$

with $\nu_h(.)$ the deconvolution kernel function, it is not difficult to verify that it is a real-valued function (see e.g. Van Es et al. (2005)).

It is known in the literature that the convergence rate of the kernel deconvolution estimator depends heavily on the distributional properties of the measurement error. To be specific, the tail decay rate of the modulus of the characteristic function of the error density matters most — the higher the decay rate, the slower the convergence rate of the estimator. For independent and identically distributed observations, the convergence rate of the kernel deconvolution estimator was studied by Fan (1991), the extensions to allow various forms of dependent observations were studied by Masry (1991).

The deconvolution problem here has logarithmic chi-square errors, its characteristic function is $\phi_k(t) = 2^{it}\Gamma(1/2+it)/\sqrt{\pi}$, and by Lemma 5.1 of Van Es et al. (2003), the tail decay rate of its modulus function is described by

$$|\phi_k(t)| = \sqrt{2}\exp\left(-\frac{\pi|t|}{2}\right)\left(1+O\left(\frac{1}{|t|}\right)\right), |t| \to \infty.$$

According to Fan (1991), this error belongs to the so-called super-smooth errors, because the tail of the modulus function decays exponentially fast.

In this chapter, I consider a particular kernel, namely the so called *sinc kernel*

$$K(x) = \frac{\sin(x)}{\pi x},$$

if $x \neq 0$ and $K(x) = 1$ if $x = 0$, with Fourier transform $\phi_K(t) = I\{|t| \leqslant 1\}$. Although $\int K(x) = 1$, it is not a proper density as it takes negative values.[3]

Using one particular kernel function seems to be rather restrictive. However, unlike the classical kernel density estimator, where the choice of kernel function has no big influence on the quality of estimator, the kernel deconvolution estimator has to put more restrictions on the kernel function, see the discussion in Delaigle and Gijbels (2004). The *sinc* kernel has optimality properties: the goodness of fit test in Butucea (2007), which uses this kernel, is shown to be a rate optimal (asymptotic) minimax test. 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. For numerical implementation, there are better choices of kernel functions, to give one example, Butucea (2007) uses a smoothed version (into a continuous trapezoidal-shaped function) of $\phi_K(t)$,

$$\phi_K(t) = I\{|t| \leqslant 1\} + \exp(1 - (|t|(2 - |t|))^{-2}) I\{1 \leqslant |t| \leqslant 2\}.$$

However, in the simple Monte Carlo experiments and empirical example in this chapter, the *sinc* kernel is still used.

For the objective volatility density function, we need the following smoothness assumption.

(C1) $g(x)$ and its second order derivatives are bounded and uniformly continuous on $\mathbb{R}$.

4.3 Parametric integrated volatility density

4.3.1 Parametric estimation of stochastic volatility model

To evaluate the parametric integrated volatility density, one first need an estimator for the stochastic volatility model. Stochastic volatility models are known to be difficult to estimate, however, significant progresses have been made in the past decade. Important works include the quasi-maximum likelihood (QML) estimator for the discrete-time stochastic volatility model in Ruiz (1994) and Harvey and Shephard (1996); the simulated methods of moments approach in Duffie and Singleton (1993); indirect inference method by Gourieroux, Monfort, and Renault (1993); the efficient methods of moments (EMM)

[3]One should notice that for the definition of the deconvolution kernel density estimator, only the Fourier transform of the function is needed, there is no need to refer to the form of the kernel function itself.

by Gallant and Tauchen (1996); and the Monte Carlo maximum likelihood methods advanced by Kim et al. (1998) and Sandmann and Koopman (1998). See also the review by Shephard (1996) and Renault (2009).

Here no specific method is chosen, it is assumed that the parametric estimator is $\sqrt{n}$-consistent under the null hypothesis, and that the parameterization is smooth, as follows:

(B1) Under the null hypothesis,

$$|\hat{\theta} - \theta_0| = O_p\left(n^{-1/2}\right).$$

(B2) $g(x, \theta)$ is Lipschitz in θ.

(B1) is satisfied by many of the above mentioned methods, such as the efficient method of moments (EMM) by Gallant and Tauchen (1996). With (B1) and (B2), the stochastic error made by the parametric estimation of the model is small and negligible relative to the nonparametric estimation part in the asymptotic analysis of the test statistics.

4.3.2 Approximating the parametric marginal density

Under the hypothesis that the parameterization $\{b(.;\theta), a(,;\theta), \theta \in \Theta \subseteq \mathbb{R}^k\}$ is correctly specified, the density function $g(x;\theta)$ of $\log \eta_i = \log\left(\int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s/\Delta\right), i = 1, \ldots, n$, usually doesn't have a closed form expression in terms of functions $a(.;\theta)$ and $b(.;\theta)$ because of the fixed interval sampling scheme used in this chapter.

Two methods of approximating the function $g(x;\hat{\theta})$ are proposed in the following. The first is based on an Euler scheme simulation of the volatility process, its precision can be made arbitrarily high by simulating the path fine and long enough. The second is heuristic: the stationary density, which is known in closed form, is used as an approximation of integrated volatility density, where the approximation error is unknown.

Approximation by Euler scheme simulation

Given the estimated parameter values $\hat{\theta}$, one can simulate paths of the estimated model and evaluate functionals of it. To be specific, given an estimate $\hat{\theta}$, the parameterization $b(.;\theta)$ and $a(.;\theta)$, and the step size Δ, taking $\delta = \Delta/N$ as a finer step, we can simulate M consecutive blocks of N observations — making $M \times N$ observations of $\sigma_1^2, \ldots, \sigma_{MN}^2$ over *one path* of the model with step length δ. Then averaging in each M blocks and taking logarithms produces a sequence of M realizations for the sequence $\log\left(\int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s/\Delta\right)$,

$i = 1, \ldots, M$. Denote the resulting simulated sample by $X_1^*, \ldots, X_M^*$. 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

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

where $K^*(.)$ is a kernel function, h_M is the bandwidth parameter.

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

Small-Δ approximation

Although the stationary volatility density, denoted as $\pi(.)$, is not the density function of the integrated volatility $\eta_i = \int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s/\Delta$, denoted as $\pi^\Delta(.)$, it can still be used as an approximation to π^Δ. Heuristically, since π^Δ converges to π when $\Delta \to 0$, the approximation error of using π to approximate π^Δ should be small for a small Δ, which leads to the name *small Δ approximation*. Aït-Sahalia et al. (2009) use a similar argument to justify the appropriateness of their approximation to the parametric transition density under a fixed sampling scheme in a jump diffusion model.

If the stationary volatility density is $\pi(x)$, the $\log \sigma^2$ process has stationary density $e^x \pi(e^x)$ by the standard change of variable formula. Denote

$$g^{\mathrm{sd}}(x) := \exp(x) \times \pi(\exp(x)), \tag{4.9}$$

and define it to be the small-Δ approximation for $g(x)$.

Studying the size of this approximation error is beyond the scope of this chapter, I only present this approximation as a heuristic method. The performance of this approximation is studied in a specific model through Monte Carlo simulation later in Section 4.6.

The formula of calculating $\pi(x)$ from functions $a(x)$ and $b(x)$ are given in (4.18) in Appendix 4.B. The following are examples of $\pi(x)$ in two popular models, the small-Δ approximation $g^{\mathrm{sd}}(x)$ can then be obtained with formula (4.9).

Example 4.1 The GARCH diffusion was found by Nelson (1990) as a diffusion approximation to the GARCH(1,1) model:

$$\mathrm{d}Y_t = \sigma_t \mathrm{d}W_t,$$

$$\mathrm{d}\sigma_t^2 = \alpha(\beta - \sigma_t^2)\mathrm{d}t + \gamma\sigma_t^2 \mathrm{d}W_t,$$

The volatility process belongs to the Chan-Karolyi-Longstaff-Sanders (CKLS) family of parametric diffusion processes (see Chan, Karolyi, Longstaff, and Sanders (1992)) with the specification for the power in the diffusion parameter to be 1. The model does not have an analytical solution but if $2\alpha/\gamma^2 > -1$, the stationary distribution is the inverse Gamma distribution, obtained from (4.5),

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

where $a = 1 + 2\alpha/\gamma^2$, $\lambda = 2\alpha\beta/\gamma^2$.

Example 4.2 Heston (1993)'s model for option pricing is:

$$\begin{aligned} \mathrm{d}X_t &= \sigma_t \mathrm{d}W_t, \\ \mathrm{d}\sigma_t^2 &= \alpha(\beta - \sigma_t^2)\mathrm{d}t + \gamma\sqrt{\sigma_t^2}\mathrm{d}B_t, \end{aligned}$$

The volatility process is the Cox-Ingersoll-Ross (CIR) process (Cox et al. (1985)). Again by formula (4.5), if $2\alpha\beta/\gamma^2 \geq 1$, its stationary distribution is the Gamma distribution, with density

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

where $a = 2\beta\alpha/\gamma^2$, $\lambda = 2\alpha/\gamma^2$.

4.4 Test statistics and asymptotic properties

4.4.1 Test statistic and asymptotic null distribution

The test statistic

$$I_0 = \int_{\mathbb{R}} \left(\hat{g}(x) - g(x;\hat{\theta})\right)^2 \mathrm{d}x,$$

calculates the L_2 distance between the parametric and nonparametric density estimates of the volatility process. However, it will be clear from Appendix 4.B that $E\hat{g}(x) = K_h * g(x;\hat{\theta})$ under $\mathcal{H}_0$, and using $K_h * g(x;\hat{\theta})$ instead of $g(x;\hat{\theta})$ will correct the bias of

the test and simplify the derivation. Notice that K here is not the deconvolution kernel function ν_h. So the following bias-corrected statistic is used instead,

$$I_1 = \int \left(\hat{g}(x) - K_h * g(x;\hat{\theta})\right)^2 \mathrm{d}x.$$

In the literature of nonparametric specification testing, using bias-corrections in the definition of test statistics has already been seen in Härdle and Mammen (1993) and Fan (1994), among others.

Theorem 4.1 *Assume conditions (SV0)–(SV5) in Appendix 4.A, and Assumptions B1, B2, and C1, such that $\{X_i\}$ is a stationary, β-mixing sequence with coefficients $\beta(k) = e^{-\lambda k}$, for some $\lambda > 0$; if $\exp(\pi/h)/n \to 0$, then*

$$\frac{1}{\sigma_1}(I_1 - \mu_1) \rightsquigarrow N(0,1),$$

where $\mu_1 = \exp(\pi/h)/(2\pi^2 n)$, $\sigma_1 = \exp(\pi/h)\|g\|_2/(2\pi n)$, and $\|g\|_2 = \left(\int g^2(x)\mathrm{d}x\right)^{1/2}$.

The above convergence rate is very slow, for example, when the bandwidth $h = O(\pi/\log(n/\log n))$, the convergence rate is $(\log n)^{-1}$.

I_1 still centers at a vanishing mean. From the proof of the theorem it is obvious that the following type of test statistic:

$$\begin{aligned} I_2 &= \frac{2}{n(n-1)h^2} \\ &\quad \times \sum_{j<k} \int \left(\nu_h\left(\frac{x - X_j}{h}\right) - K_h * g(x;\hat{\theta})\right)\left(\nu_h\left(\frac{x - X_k}{h}\right) - K_h * g(x;\hat{\theta})\right)\mathrm{d}x, \end{aligned}$$

will center at zero, and it is not difficult to show $I_2/\sigma_1 \rightsquigarrow N(0,1)$. However, for ease of implementation, we stick to the test statistic I_1 in the following.

4.4.2 Power properties

For the asymptotic power of the test, a fixed alternative is firstly considered,

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

Theorem 4.2 *Assume Conditions (SV0)–(SV5) in Appendix 4.A, and Assumptions B1, B2, and C1, let $\alpha \in (0,1)$ be the level of significance, and $Z_{1-\alpha}$ the $1-\alpha$ quantile of the standard normal distribution, then under $\mathcal{H}_1'$,*

$$P\left(\frac{1}{\sigma_1}(I_1 - \mu_1) > Z_{1-\alpha}\right) \to 1,$$

as $n \to \infty$.

As with most nonparametric tests, this test is consistent.

Power of the test under local alternatives is considered next:

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

where γ_n and $\Delta(x)$ have to be chosen such that g_n is a density, with $\gamma_n \to 0$ as $n \to \infty$, $\int \Delta(x)\mathrm{d}x = 0$ and $\int \Delta^2(x)\mathrm{d}x < \infty$. This is a sequence of alternative models converging to the null density at a certain rate γ_n.

Theorem 4.3 *Assume Conditions (SV0)–(SV5) in Appendix 4.A, and Assumptions B1, B2, and C1, then under $\mathcal{H}'_{1n}$, for $\gamma_n^2 = \exp(\pi/h)/n \to 0$*

$$\left(\frac{4\pi^2 n^2}{\exp(2\pi/h)\|g\|_2^2}\right)^{1/2} \left(I_1 - \frac{1}{2\pi^2 n} e^{\pi/h}\right) \rightsquigarrow N\left(\frac{2\pi}{\|g\|_2^2}\int \Delta^2(x)\mathrm{d}x, 1\right),$$

and it follows that

$$P\left(\frac{1}{\sigma_1}(I_1 - \mu_1) > Z_{1-\alpha}\right) \to \Phi\left(Z_{1-\alpha} - \frac{2\pi}{\|g\|_2^2}\int \Delta^2(x)\mathrm{d}x\right).$$

Although the test is consistent against fixed alternatives, it only has power against local alternatives that converge at the rate slower than $\exp(\pi/(2h))/n^{1/2}$. This is very slow, for example when $h = O(\pi/\log(n/\log n))$, this is $(\log n)^{-1/2}$.

4.4.3 Bootstrap null distribution

The slow rate of convergence in Theorem 1 will make the asymptotic null distribution a poor approximation to the null distribution in finite samples. This happens in most nonparametric testing procedures, where bootstrap methods are often used to approximate the null distribution. I use a parametric bootstrap procedure as follows:

Step 1 Given a parametric estimate $\hat{\theta}$, step size Δ, simulate n (original sample size) discretely observed Δ-returns, which is called one bootstrap sample. This step has to be done over a fine grid like in section 4.3.2.

Step 2 With this bootstrap sample, compute the test statistic I_1, call it I_1^*.

Step 3 Repeat step 1 and 2 for B times to get a bootstrap sample $I_1^{*1}, \ldots, I_1^{*B}$ for the statistic I_1.

When B is big, the empirical distribution of $I_1^{*1}, \ldots, I_1^{*B}$ estimates the finite sample null distribution.

Using similar methods as in the previous section, it is possible to show the consistency and power properties of the bootstrap test. These results and proofs are omitted here to avoid repetitions.

Notice when computing I_1^* in each bootstrap sample, it is not necessary to re-estimate the parametric part $g(x, \hat{\theta}^*)$ each time, the re-estimation only applies to the nonparametric estimator $\hat{g}(x)$. The procedure should still be consistent, see Fan (1994) for a discussion in a similar context. This is crucial for the stochastic volatility models, as the estimation of stochastic volatility models usually requires simulation and numerical maximization, the bootstrap will be very slow if these have to be repeated in each replication of the bootstrap.

4.5 Extensions

4.5.1 Discrete-time model

Continuous-time models have been considered so far, but the approach can be adapted to discrete-time stochastic volatility model easily. Consider the classical discrete-time stochastic volatility model

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

where the specification is directly on the observed return y_{t_i}, $i = 1, \ldots, n$. When $\gamma < 1$, and the volatility process $\log \sigma_t^2$ is initiated from the stationary distribution of $\{\log \sigma_{t_i}^2\}_{i=1}^n$, which is $N(\omega/(1-\gamma), \sigma_\eta^2/(1-\gamma^2))$, the volatility process is strictly stationary. It is also β-mixing with exponentially decaying coefficients, see Pham and Tran (1985), and thus ergodic.

Similarly as before, one can compare the kernel deconvolution volatility density estimate with its parametric counterpart. A particularly nice aspect of *this* discrete-time model is that the parametric logarithm volatility density is known to be normal and needs no approximation. In more general cases, however, the stationary densities of discrete time volatility models are usually difficult to obtain.

4.6 Monte Carlo simulations

The finite sample properties of the test are studied in this section through Monte Carlo experiments. In the context of a continuous time stochastic volatility model, the null distribution of the test and the performance of the small-delta approximation are firstly studied. Simulation results show that, for this specific example, the null distribution obtained through small-Δ approximation is almost identical to that obtained through Euler scheme simulation over a very fine grid. To reduce the computation cost in Monte Carlo experiment, a discrete-time model is considered explicitly as the null model, the null distribution is plotted and the power of the test I_1 is compared with several other tests when the alternative model is a Heston model.

The null model considered in this section is the continuous time Stochastic AutoRegressive Volatility (SARV) model

$$\mathrm{d}Y_t = \sigma_t \mathrm{d}B_t, \tag{4.10}$$

$$\mathrm{d}\log\sigma_t^2 = \alpha(\beta - \log\sigma_t^2)\mathrm{d}t + \gamma \mathrm{d}W_t, \tag{4.11}$$

with annulized parameter value $\alpha = 10.080$, $\beta = 6.5294$, $\gamma = 3.6511$, $\Delta = 1/252$, so the sampling interval $\Delta = 1/252$ corresponds to daily data, the sample size used is $n = 3000$. It will be explained how these parameters are determined in Section 4.6.3.

4.6.1 Accuracy of small-delta approximation

In this section I investigate the accuracy of the small delta approximation for this specific model. The stationary distribution of $\log\sigma^2$ here is $N(\beta, \gamma^2/(2\alpha))$, the small-$\Delta$ approximation proposed in Section 4.3.2 uses this density to approximate the density of the integrated volatility $\int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s/\Delta$, because $\Delta = 1/252$ is small. On the other hand, the density of $\int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s/\Delta$ can be approximated through an Euler scheme simulation of the stochastic differential equation system over a finer grid; this approximation can be made arbitrarily well if one chooses the grid to be fine enough and the simulation path to be long enough.

First, the model is simulated over a finer grid with step size $\delta = 1/(252 \times 100)$ for $3000 \times 100 \times 100$ δ-steps, this is 3000×100 days with 100 steps within each day. The the volatilities are summed and normalized within each day to get 3000×100 integrated volatility realizations over a day, and these are taken as the realizations of the integrated volatility $\int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s$. Since the simulation sample size is large (3000×100), using classical kernel density estimation with these simulated realizations will give an approximation of

the integrated volatility density, this is taken as the *true* integrated volatility density. On the other hand, the density of $N(\beta, \gamma^2/(2\alpha))$ is the small-Δ approximation of the integrated volatility density.

These two densities could be compared directly, but here a different approach of comparison is taken — for the purpose of obtaining the null distribution of the test, the null distributions of the test under two approximation schemes are compared. With the same method of simulation over $\delta = 1/(252 \times 100)$, I simulate 1000 paths of 3000 daily returns, these and the true density obtained in the previous paragraph are used to compute 1000 realizations of I_1. Furthermore, these and the small-Δ approximated density are used to compute 1000 realizations of the small-delta approximated version of I_1, denoted as I_1^s. The following table gives the quantiles of the true density with the small delta approximated density, for different bandwidths. The table shows that in the present context, the small delta approximation, although heuristic, works quite well in approximating the null distribution, especially its right tail, where critical values are often found, of the test statistic for a wide range of bandwidths. Bandwidths $0.1, 0.2, \ldots 0.9$ are used in the Monte Carlo but only results for 4 different bandwidths are given below to save space.

Quantiles of true distribution at	$h = 0.3$	$h = 0.5$	$h = 0.7$	$h = 0.9$
0.99	0.010	0.010	0.013	0.011
0.95	0.053	0.055	0.053	0.056
0.90	0.105	0.105	0.107	0.113
0.80	0.210	0.213	0.209	0.218
0.70	0.299	0.306	0.318	0.313
0.60	0.403	0.404	0.407	0.415
0.50	0.500	0.506	0.518	0.517
0.40	0.607	0.607	0.609	0.606
0.30	0.703	0.709	0.697	0.700
0.20	0.807	0.800	0.794	0.805
0.10	0.910	0.899	0.903	0.900

Table 4.1: Exceedance probabilities of small delta approximated distribution at different true quantiles, with different bandwidths

4.6.2 Null distribution of the test

In view of the good performance of the small-delta approximation in the previous subsection, and also to reduce the numerical burden in the Euler scheme approximation over a finer grid, the following model (4.12)–(4.14) is considered from now on,

$$y_t = \sigma_t \varepsilon_t, \tag{4.12}$$

$$\log \sigma_t^2 = \omega + \gamma \log \sigma_{t-1}^2 + \sigma_\eta \eta_t, \tag{4.13}$$

$$(\varepsilon_t, \eta_t) \sim \text{i. i. d. } N(0, I_2), t = 1, \ldots, n, \tag{4.14}$$

with parameter values $\omega = 0.04$, $\gamma = 0.96$, and $\sigma_\eta = 0.23$, which are implied by the parameters in model (4.10)–(4.11). This model is often directly referred to as Stochastic AutoRegressive Volatility (SARV) model mentioned above. It can be viewed as a discretization of the continuous-time SARV model, but can also be used as a null model by itself. Considering this model avoids the difficulties in the approximation of the stationary volatility density, and keeps the focus on the properties of the test.

Simulating this model for 1000 times, and calculating the test values, 1000 realizations of the null distribution can be obtained, from which the null distribution, or more precisely the bootstrapped null distribution can be obtained using a classical kernel density estimator. Figure 4.1 gives a plot of the bootstrapped null distribution of the test (in red) when the bandwidth used is 0.9; it has a high peak and is slightly skewed to the left. The green line is the density of a normal distribution that has the same mean and variance as the bootstrapped null distribution, while the blue squares denotes the density given by the asymptotic approximation of Theorem 4.1. From Figure 4.1 it is clear that at the present sample size 3000 and bandwidth 0.9, although the asymptotic theory gives a good estimate of the mean and variance of the null distribution (because of the close match between the calibrated normal density and the asymptotic density), the shape of the null density is far away from a normal density, which also justifies the usage of the bootstrap in obtaining the null distribution in the present nonparametric test context.

4.6.3 Power of the test

The alternative model considered is the Heston model,

$$\begin{aligned} \mathrm{d}Y_t &= \sigma_t \mathrm{d}W_t, \\ \mathrm{d}\sigma_t^2 &= \alpha(\beta - \sigma_t^2)\mathrm{d}t + \gamma\sqrt{\sigma_t^2}\mathrm{d}B_t. \end{aligned}$$

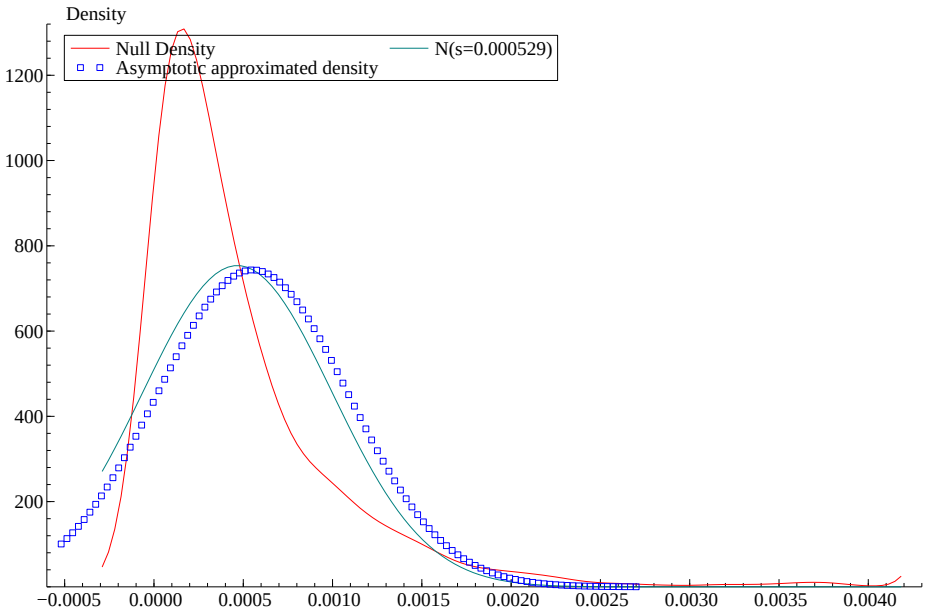

Figure 4.1: Bootstrapped null distribution of I_1, together with the normal calibrated distribution and asymptotic approximated null distribution.

The parameter values are $\alpha = 5$, $\beta = 0.1$, $\gamma = 1$ with step size $\Delta = 1/252$ and sample size 3000. These are annualized parameter values from Aït-Sahalia and Yu (2009), so $\Delta = 1/252$ corresponds to one day. The null model is estimated by fitting the SARV model with the simulated Heston model data using the Efficient Method of Moment (EMM) by Gallant and Tauchen (1996), the implementation in OxMetrics is provided by EmmPack 1.08 by Van der Sluis. The model parameters used in Section 4.6.2 are actually the estimated parameters in the null model.

A plot of the return density and the volatility density for the two models may be instructive on why the test I_1 is useful. Figure 2 and 3 plot the stationary volatility densities of the two models (for $\log \sigma^2$). The volatility density for Heston model is skewed to the right and has a fat left tail, but the volatility density of the SARV model is a normal density. But after taking the variance mixture with a normal distribution, the return densities for both models become symmetric and fat-tailed, as showed in Figure 4 and 5. At least visually, it's easier to distinguish the two models by comparing the volatility density, instead of the return density.

The test requires a bandwidth parameter. Bandwidth selection is important in the implementation of the kernel deconvolution density estimator. See Delaigle and Gijbels (2004) for a comprehensive discussion of practical bandwidth selection for kernel deconvo-

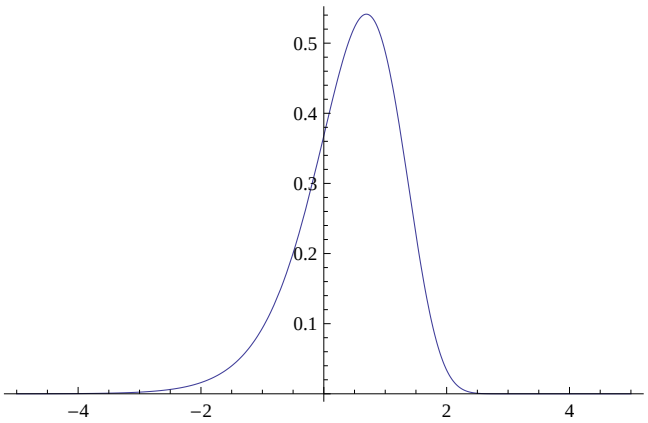

Figure 4.2: Volatility density of Heston model

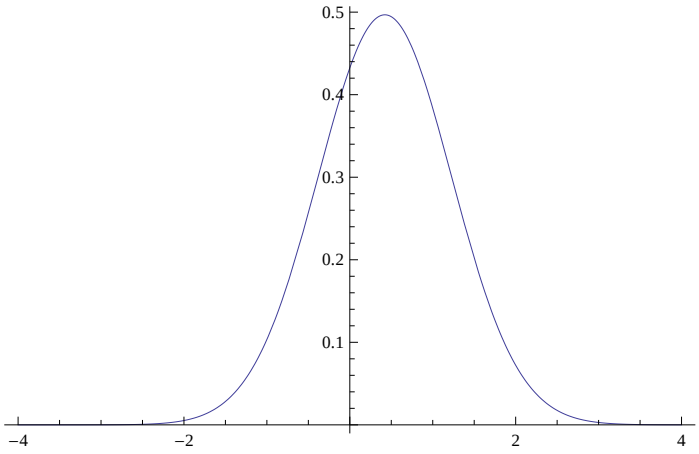

Figure 4.3: Volatility density of SAR1V model

lution estimators. In the context of nonparametric tests with direct observations available, bandwidth selection is also important for the power of the test, see Gao and Gijbels (2008). This chapter does not deal with bandwidth selection problem explicitly; I perform the Monte Carlo experiment for a set of bandwidths $h = 0.1, 0.2, \ldots, 0.9$, and the bandwidth is chosen to maximize the Monte Carlo power. This is also the reason I report the null distribution for $h = 0.9$ in Section 4.6.2.

1000 paths of the alternative model are simulated over finer grid and the corresponding 1000 realizations of test I_1 are calculated. The empirical distribution from these 1000 realizations is taken as the alternative distribution. The null distribution is already given in the previous section. The Monte Carlo power obtained in this way can be viewed as the power of the bootstrap test. The simulation was conducted with OxMetrics 5.10 with random seed 20101215.

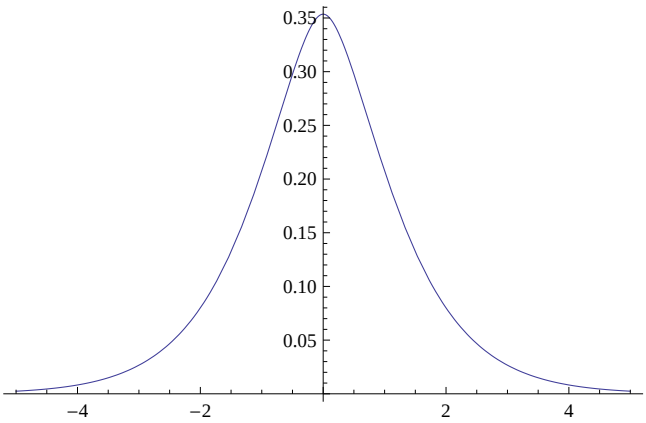

Figure 4.4: Return density of Heston model

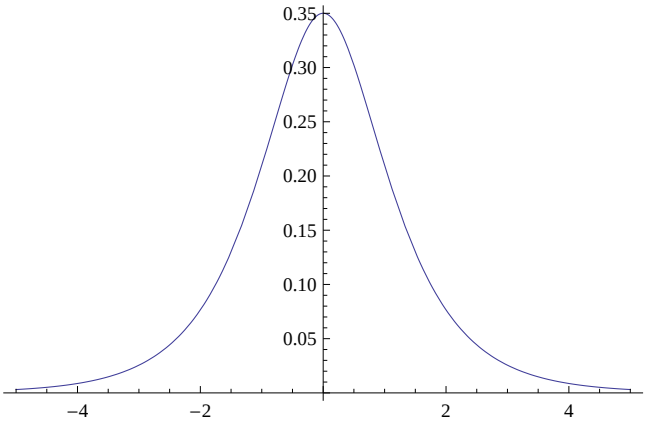

Figure 4.5: Return density of SAR1V model

The test I_1 is not the only nonparametric specification test for stochastic volatility models. One can compare the nonparametric estimate and the parametric estimate of the density function, denoted as $g(x)$, of the return sequence $\{y_i\}$; see Chapter 2. If the L_2 distance is used as in the present chapter, the test statistic is

$$T_1 = \int \left(\hat{q}(x) - K_h * q(x;\hat{\theta})\right)^2 \mathrm{d}x,$$

where $\hat{q}(x) = \sum_{i=1}^{n} K\left((x-y_i)/h\right)/(nh)$ is the classical kernel density estimator (with direct observations) of the return density, $K_h * q(x;\hat{\theta}) = \int_{\mathbb{R}} K_h(x-y)q(y;\hat{\theta})\mathrm{d}y$ is the convolution of the function $K_h(x) = K(x/h)/h$ with function $q(x;\hat{\theta})$ and the function $K(.)$ and bandwidth h here are the same that used in the definition of $\hat{q}(x)$.

Another way to construct test statistics is to compare the distribution function of the return sequence $\{y_i\}$, denoted as $Q(x)$. Using the L_2 distance gives the Cramer von Mises

type test for the return distribution function,

$$T_2 = n \int \left(\widehat{Q}(x) - Q(x;\hat{\theta})\right)^2 \mathrm{d}Q(x;\hat{\theta}),$$

where $\hat{Q}(x) = \sum_{i=1}^{n} I(y_i \leqslant x)/n$ is the empirical distribution function of the return sequence, and $Q(x;\hat{\theta})$ is the parametric counterpart. Similarly, the L_∞ distance gives a Kolmogorov-Smirnov type test

$$T_3 = \sup_{x\in\mathbb{R}} \sqrt{n} \left|\widehat{Q}(x) - Q(x;\hat{\theta})\right|.$$

Zu and Boswijk (2009) studied these tests. To show the power of I_1, I compare the Monte Carlo power of these tests with the power of I_1. The Monte Carlo power of these tests are evaluated using the same simulation method described early in this section.

	1% level	5% level	10% level
I_1 $(h = 0.9)$	0.400	0.670	0.784
I_1 $(h = 0.8)$	0.339	0.617	0.755
I_1 $(h = 0.7)$	0.328	0.534	0.699
I_1 $(h = 0.6)$	0.303	0.483	0.623
I_1 $(h = 0.5)$	0.243	0.433	0.555
I_1 $(h = 0.4)$	0.128	0.301	0.390
T_1	0.245	0.405	0.538
T_2	0.157	0.274	0.365
T_3	0.189	0.303	0.427

Table 4.2: Comparison of the power of the bootstrap tests

The power of the test I_1 depends on the bandwidth parameter. The first 6 rows of Table 4.2 give the power of the test when the bandwidth used varies 0.4 to 0.9. Notice that the power of the test increases with bandwidth value: when $h = 0.5$, the power of I_1 is roughly the same as the return density based test T_1, and when h is greater than 0.5, the power of I_1 strictly dominates the rest of the tests. Theorem 1 requires the bandwidth to satisfy $h \to 0$ and $\exp(\pi/h)/n \to 0$, so h has to goes to 0 slower than $\pi/\log(n)$. In practice, it is difficult to specify a number that satisfies this condition, but heuristically, since $n = 3000$ here and $\pi/\log(3000) \approx 0.39$, an oversmoothed bandwidth should be greater than this value. From the table it is clear that all the bandwidths having good power satisfy these requirements. So Table 4.2 shows that when the bandwidth is appropriately chosen, the volatility density based tests have superior power compared to

the return density function and return distribution based tests in the present scenario. This confirms the claim that the volatility density based test is a useful alternative to the nonparametric tests for stochastic volatility models.

4.7 Empirical example

In this section, the test I_1 is applied to the discrete time stochastic volatility model estimated in Van der Sluis (1997),

$$y_t = \sigma_t \varepsilon_t, \tag{4.15}$$

$$\log \sigma_t^2 = \omega + \gamma \log \sigma_{t-1}^2 + \sigma_\eta \eta_t, \tag{4.16}$$

$$(\varepsilon_t, \eta_t) \sim \text{i. i. d. } N(0, I_2), \quad t = 1, \ldots, n. \tag{4.17}$$

This model is fitted to daily British pound/Canadian dollar exchange rate data from January 1971 to August 1996, with 1019 observations in total. Figure 4.6 gives a plot of the exchange rate series, from which volatility clustering is observed.

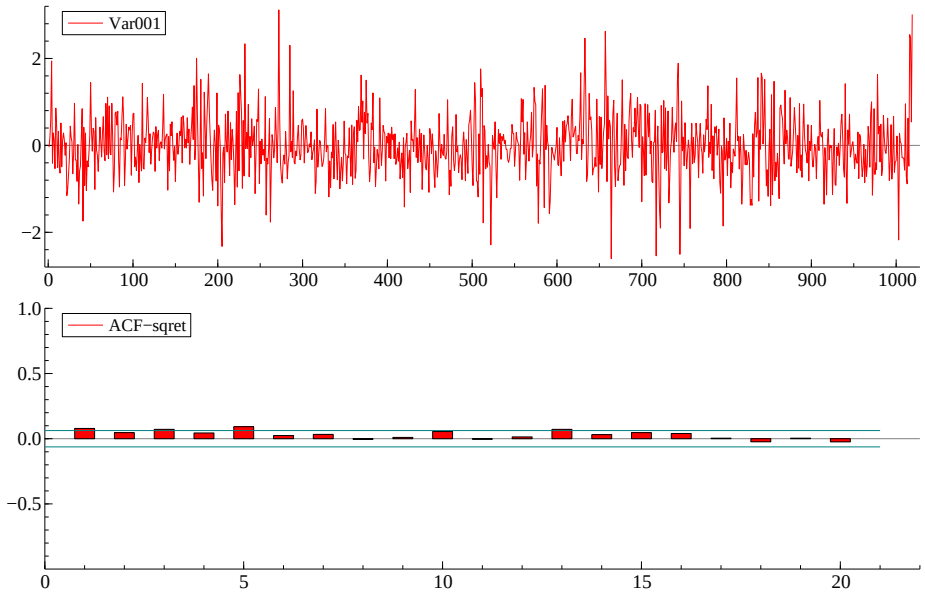

Figure 4.6: Plot of the British Pound/Canadian Dollar exchange rate

Using the kernel deconvolution estimator, the empirical volatility density can be obtained and it is plotted in Figure 4.7. The bandwidth is chosen using the same heuristic argument as in the previous section — since $n = 1019$ and $\pi/\log(1019) \approx 0.453$, I use a slightly oversmoothed bandwidth 0.46 here. It will be seen later that, at this bandwidth

the model can be rejected. Figure 4.7 shows that the deconvolution density is slightly skewed to the left. The estimator generates negative values at both tails, which is a common problem among many kernel deconvolution estimators. The negative estimates can be understood by noticing that the kernel deconvolution density estimator is an unbiased estimator of a convoluted version of the objective density function (in finite sample, recall the result in Section 4.4.1 and the proof in Appendix 4.B), which could take negative values if the kernel function is not a proper density function. By making a forward reference to Figure 4.9 one can see that the current choice of kernel function indeed leads to a convoluted volatility density taking negative values. To make a direct comparison between the estimates and the convoluted kernel function, I don not truncate the negative estimates.[4]

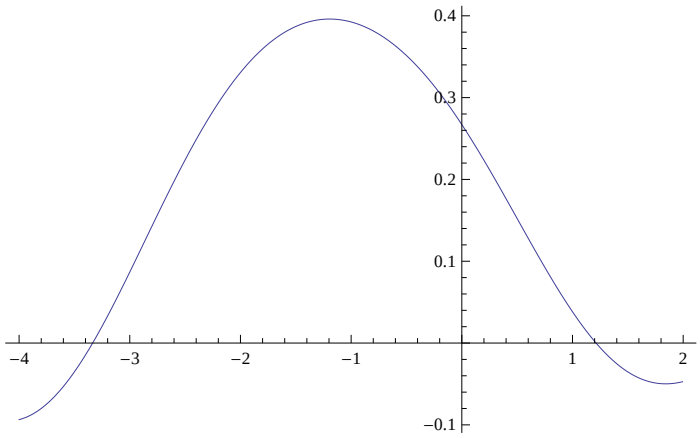

Figure 4.7: Empirical volatility density

The model (4.15)–(4.17) is estimated by the Efficient Method of Moment developed by Gallant and Tauchen (1996). The data and code are provided for free public access in EmmPack 1.08. The estimated parameter values are $\hat{\omega} = -0.083688$, $\hat{\gamma} = 0.89630$, $\hat{\sigma_\eta} = 0.19583$. The volatility process $\log \sigma_t^2$ follows an AR(1) process, with stationary distribution $N(\omega/(1-\gamma), \sigma_\eta^2/(1-\gamma^2))$. With the estimated value, the volatility density implied by the model is plotted in Figure 4.8. Because the test I_1 uses a convoluted parametric volatility density, Figure 4.9 gives a plot of the convoluted volatility with the kernel K. Notice that after the convolution operation, the density can take negative values. This is the reason why the negative values in Figure 4.7 have not been truncated.

[4]It should be pointed out that possibly negative estimates *per se* is not a serious drawback of the *sinc* kernel, as many kernels used in kernel deconvolution estimators are not proper densities. Getting negative estimates in the tail is a usual problem encountered in kernel deconvolution estimation.

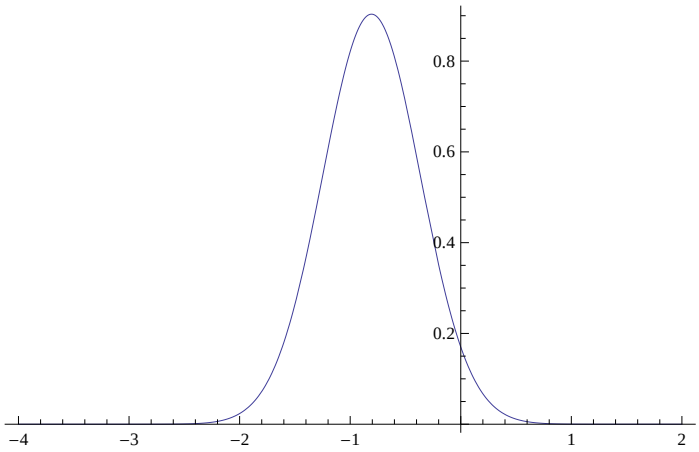

Figure 4.8: Volatility density implied by the null model

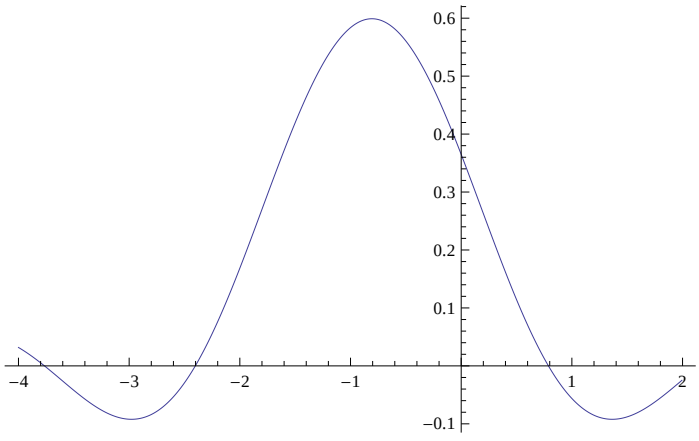

Figure 4.9: Convoluted volatility density implied by the null model

A quick inspection of Figures 4.7 to 4.9 shows that the empirical volatility density differs a lot from the convoluted volatility density implied by the null model. One may doubt if this is the consequence of the poor performance of the deconvolution estimator or a bad choice of bandwidth. Figure 4.10 gives a plot of the deconvolution volatility density (using the same bandwidth 0.46) with the simulated returns from the estimated null model. As compared to the empirical volatility density in Figure 4.7, it looks more similar to the convoluted volatility density in Figure 4.9. This is a sign that the estimated model may not show a good fit with data in terms of the volatility density.

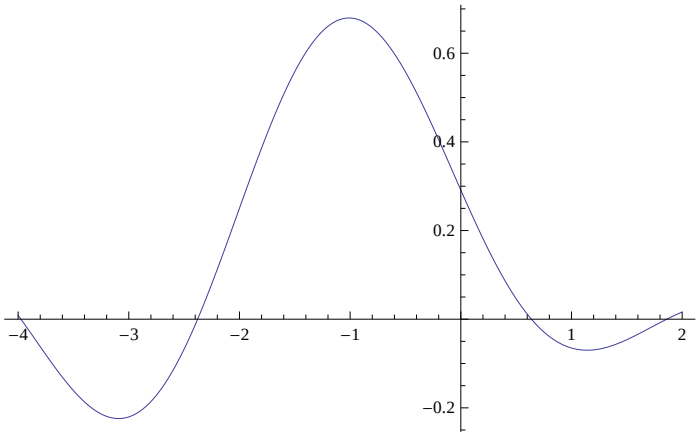

Figure 4.10: Deconvoluted volatility density of the null model

Using the parametric bootstrap method, the estimated null model is simulated (bootstrapped) 5000 times with random seed 20110602 to get the 5000 realizations of the null distribution, the empirical distribution of these realization is taken as the null distribution. The test statistic I_1 is calculated and compared with the null distribution to get the p-value of the test. Still the power of the test depends on the bandwidth used — bandwidths greater than or equal to 0.42 lead to reject the null model at the 5% level, and bandwidths greater or equal than 0.44 lead to reject the null model at the 1% level. That is to say, oversmoothing bandwidths larger than 0.46 all lead to rejection of the model, and this is in line with the Monte Carlo simulation evidence in Section 4.6.3, that oversmoothing bandwidths lead to better power.

4.8 Conclusion

This chapter studies nonparametric specification tests for stochastic volatility models. A test statistic is proposed based on integrated volatility density. The asymptotic null

p-value	bandwidth
0.0702	0.40
0.0266	0.42
0.0100	0.44
0.0050	0.46
0.0038	0.48
0.0026	0.50

Table 4.3: p-values of the test for different bandwidths

distribution of the test and its asymptotic power properties are derived, and a parametric bootstrap procedure is proposed to obtain the critical values. The finite sample properties of the method are studied in Monte Carlo simulations. When compared with existing nonparametric tests for stochastic volatility models, the test proposed in this chapter shows superior power properties.

It is well known that an important property of stochastic volatility is its ability to capture the volatility clustering phenomenon existing in financial data. The tests in this chapter only use information on the marginal integrated volatility density; extending the test in this chapter to exploit the information contained in the dependence structure in the data is an interesting future direction to work on.

Appendix 4.A: Assumptions and probabilistic properties of stochastic volatility models[5]

Assumptions

In this section, the basic setup and assumptions of the model (4.1)–(4.2) are established. These assumptions are made for the observed return sequence to be stationary, ergodic and β-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 made on the $b(x;\theta)$ and $a(x;\theta)$ functions, which can imply the probablistic properties needed for the observed return sequence.

[5]This section is largely a repitition of Appendix 3.A, it is kept here for the completeness of the present chapter.

In the parametric stochastic volatility model (4.1)–(4.2), 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 σ_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

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

and

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

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

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

and the speed measure

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

and assume

(SV2)

$$\int_{0+} s(x;\theta)\mathrm{d}x = +\infty, \quad \int^{+\infty} s(x;\theta)\mathrm{d}x = +\infty, \quad \int_0^{+\infty} m(x;\theta)\mathrm{d}x = M < +\infty,$$

where the indefinite part of the integral means an arbitrary point in the domain of the integrand.

(SV3) The initial random variable σ_0^2 has distribution $\pi(\mathrm{d}x) = \pi(x;\theta)\mathrm{d}x$.

(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

$$\pi(x;\theta) = \frac{m(x;\theta)}{M} I\{x > 0\}. \tag{4.18}$$

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 β-mixing. These are standard conditions 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 ρ-mixing. From Theorem 3.6 in Chen et al. (2010), a sufficient condition (together with (SV1) and (SV2)) for exponential decay β-mixing coefficients is the process to be ρ-mixing — this in particular is not a very strong assumption, because as discussed in the same paper, β-mixing and ρ-mixing with exponential decay are almost equivalent concepts for scalar diffusion. We also note the result that if a diffusion process is ρ-mixing, its ρ-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 β-mixing with exponentially decaying coefficients.

The conditions above only deliver properties for the volatility process, the following lemma shows that the return sequence $y_i = \int_{t_{i-1}}^{t_i} \sigma_s \mathrm{d}B_s/\sqrt{\Delta}$, $i = 1, \ldots, n$, 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 4.1 *In model (4.1)–(4.2), if the volatility process is stationary, ergodic and β mixing with a certain decay rate, then the normalized return sequence $y_i, i = 1, \ldots, n$, is also stationary, ergodic and β 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 $U_i := (\int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s, \sigma_{t_i}^2)$, $i = 1, \ldots, n$, 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 β-mixing with at least the rate in the Hidden chain U_i, $i = 1, \ldots, n$. As noted in the proof of proposition 3.2 in Genon-Catalot et al. (2000),

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

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

Appendix 4.B: Proofs of the theorems

Proof (the claim that $E\hat{g}(x) = K_h * g(x)$) Notice that

$$\begin{aligned}
E\hat{g}(x) &= E\frac{1}{h}\nu_h\left(\frac{x - X_1}{h}\right) \\
&= \int \frac{1}{h}\nu_h\left(\frac{x - x_1}{h}\right) f(x_1)\mathrm{d}x_1 \\
&= \frac{1}{2\pi}\int e^{-\mathrm{i}tx}\phi_{\nu_h}(th)\phi_g(t)\phi_k(t)\mathrm{d}t \\
&= \frac{1}{2\pi}\int e^{-\mathrm{i}tx}\frac{\phi_K(th)}{\phi_k(t)}\phi_g(t)\phi_k(t)\mathrm{d}t \\
&= \frac{1}{2\pi}\int e^{-\mathrm{i}tx}\phi_K(th)\phi_g(t)\mathrm{d}t \\
&= \frac{1}{2\pi}\int e^{-\mathrm{i}tx}\phi_{K_h * g}(t)\mathrm{d}t \\
&= K_h * g(x),
\end{aligned}$$

where I use Parseval's equality for the third equality, and the inverse Parseval's equality and Fourier inversion formula in the last two equalities. □

The derivation of Theorem 1 is closely related to the derivation of asymptotic normality of the integrated squared error (ISE) of the kernel deconvolution estimator. For i.i.d. data, the asymptotic distribution of ISE for different classes of signal densities and noise distributions were derived by Butucea (2004). This chapter extends the result by Butucea to the logarithmic chi-square noise distribution and β-mixing observations.

Proof (of Theorem 4.1)

The exponential rate decay of β-mixing coefficients is a sufficient condition to use the limiting theorem for U-statistics; it could be reduced to a polynomial rate.

Notice that because of the assumption of $\sqrt{n}$-consistency of the parametric estimator under the null hypothesis, and the Lipschitz continuity of the parameterization, the parametric estimator $g(x, \hat{\theta})$ will converge to $g(x)$ pointwise at rate $\sqrt{n}$. Taking into account

that the convergence rate of the nonparametric part is slower than $\sqrt{n}$, replacing $g(x,\hat{\theta})$ by the true density $g(x)$ will not alter the asymptotic result, and I will do that in the following to simplify the proof.

First the estimator can be written as

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

where ω_h is defined implicitly. The test statistic T_1 can then be written and decomposed as follows,

$$\begin{aligned}
T_1 &= \int\left(\frac{1}{n}\sum_{i=1}^{n}\omega_h\left(x-X_i\right) - K_h * g(x)\right)^2 \mathrm{d}x \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\int\left(\omega_h(x-X_i) - K_h * g(x)\right)^2 \mathrm{d}x \\
&\quad + \frac{2}{n^2}\sum_{1\leqslant i<j\leqslant n}\int\left(\omega_h(x-X_i) - K_h * g(x)\right)\left(\omega_h(x-X_j) - K_h * g(x)\right)\mathrm{d}x \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\|U_i\|^2 + \frac{2}{n^2}\sum_{1\leqslant i<j\leqslant n}\langle U_i, U_j\rangle
\end{aligned}$$

where $U_i(x) := \omega_h(x-X_i) - K_h * g(x)$ and $\|.\|$ and $\langle .,.\rangle$ are the norm and the scalar product in L_2 space, respectively. Taking the expectation of T_1,

$$ET_1 = \frac{1}{n^2}nE\|U_1\|^2 + \frac{2}{n^2}\sum_{1\leqslant i<j\leqslant n}E\langle U_i, U_j\rangle,$$

then

$$\begin{aligned}
T_1 - ET_1 &= \frac{1}{n^2}\sum_{i=1}^{n}\left(\|U_i\|^2 - E\|U_i\|^2\right) + \frac{2}{n^2}\sum_{1\leqslant i<j\leqslant n}\left(\langle U_i, U_j\rangle - E\langle U_i, U_j\rangle\right) \\
&:= S_1 + S_2,
\end{aligned}$$

where the two quantities S_1 and S_2 are defined implicitly.

With this decomposition, the proof proceeds along the following steps:

1. It is shown that $ET_1 = e^{\pi/h}/(2\pi^2 n)$.

2. It is shown that $\left(\frac{4\pi^2 n^2}{\exp(2\pi/h)\|g\|_2^2}\right)^{1/2} S_2 \rightsquigarrow N(0,1)$.

3. It is shown that $S_1 = o_p\left(\left(\frac{1}{n^2}\exp\left(\frac{2\pi}{h}\right)\right)^{1/2}\right)$.

4. Then it is proved that

$$\left(\frac{4\pi^2 n^2}{\exp(2\pi/h)\|g\|_2^2}\right)^{1/2}(T_1 - ET_1) = \left(\frac{4\pi^2 n^2}{\exp(2\pi/h)\|g\|_2^2}\right)^{1/2}(S_1 + S_2) \rightsquigarrow N(0,1),$$

as $n \to \infty$, $h \to 0$ by combining the above three results.

Order of ET_1 First,

$$\begin{aligned} E\|U_1\|^2 &= E\int (\omega_h(x - X_1) - K_h * g(x))^2 \,\mathrm{d}x \\ &= \int \left(\int (\omega_h(x - x_1) - K_h * g(x))^2 \,\mathrm{d}x\right) g(x_1)\mathrm{d}x_1 \\ &= \int \left(\int (\omega_h(x - x_1))^2 \,\mathrm{d}x\right) g(x_1)\mathrm{d}x_1 - \left(\int (K_h * g(x))^2 \,\mathrm{d}x\right)^2. \end{aligned}$$

The first term is

$$\begin{aligned} \int \left(\int (\omega_h(x - x_1))^2 \,\mathrm{d}x\right) g(x_1)\mathrm{d}x_1 &= \frac{1}{h}\int \nu_h^2(x)\mathrm{d}x \int g(x_1)\mathrm{d}x_1 \\ &= \frac{1}{h}\int \nu_h^2(x)\mathrm{d}x \\ &= \frac{1}{2\pi h}\int |\phi_{\nu_h}(u)|^2 \,\mathrm{d}u \\ &= \frac{1}{2\pi h}\int \left|\frac{\phi_K(u)}{\phi_k(u/h)}\right|^2 \mathrm{d}u \\ &= \frac{1}{2\pi h}\int \left|\frac{I(|u| \leqslant 1)}{\sqrt{2}\exp\left(-\pi|u|/(2h)\right)}\right|^2 \mathrm{d}u \\ &= \frac{1}{2\pi^2}\exp\left(\frac{\pi}{h}\right), \end{aligned} \tag{4.19}$$

where in the third equality Parseval's equality is used. In the second to last equality, I substitute the leading part of $\phi_K(u)$ in the integral, so the equality should be understood as an asymptotic one, I also do this in the following without notice to simplify proof. The second term is $O(1)$ as $h \to 0$, such that it's dominated by the first term.

Next, it is shown that the sum of cross products satisfies

$$\sum_{1 \leqslant i < j \leqslant n} E\langle U_i, U_j\rangle = o_p\left(\frac{n}{2\pi^2}\exp\left(\frac{\pi}{h}\right)\right), \tag{4.20}$$

by applying the covariance inequality for β-mixing sequence in Yoshihara (1976) (Lemma 8 in the Appendix). First I check the condition in the Lemma 1 of Yoshihara (1976): let $\widetilde{U}_i$ and $\tilde{U}_j$ be two independent random variables having the same distribution as U_i

and U_j, respectively; and let $\widetilde{X}_i$ and $\tilde{X}_j$ be two independent random variables having the same distribution as X_i and X_j, respectively. In the following, to simplify the proof, I only keep the dominant term in the expansion. Notice that for some $\delta > 0$,

$$\begin{aligned} & E\left|\left\langle \tilde{U}_i, \tilde{U}_j \right\rangle\right|^{1+\delta} \\ = \; & E\left|\int \left(\omega_h\left(x-\tilde{X}_i\right) - K_h * g\left(x\right)\right)\left(\omega_h\left(x-\tilde{X}_j\right) - K_h * g\left(x\right)\right) \mathrm{d}x\right|^{1+\delta} \\ = \; & \frac{1}{h^{2(1+\delta)}} E\left|\int \nu_h\left(\frac{x-\tilde{X}_i}{h}\right)\nu_h\left(\frac{x-\tilde{X}_j}{h}\right) \mathrm{d}x\right|^{1+\delta}. \end{aligned}$$

Define

$$\xi_h\left(x\right) := \int \nu_h\left(x+z\right)\nu_h\left(z\right)\mathrm{d}z,$$

then

$$\begin{aligned} E\left|\left\langle \tilde{U}_i, \tilde{U}_j \right\rangle\right|^{1+\delta} &= \frac{1}{h^{1+\delta}} E\left|\xi_h\left(\frac{\widetilde{X}_i - \tilde{X}_j}{h}\right)\right|^{1+\delta} \\ &= \frac{1}{h^{1+\delta}} \int\int \left|\xi_h\left(\frac{x_i - x_j}{h}\right)\right|^{1+\delta} g\left(x_i\right) g\left(x_j\right) \mathrm{d}x_i \mathrm{d}x_j \\ &= \frac{1}{h^{\delta}} \int\int \left|\xi_h\left(u\right)\right|^{1+\delta} g\left(x_j + uh\right) g\left(x_j\right) \mathrm{d}u \mathrm{d}x_j \left(1 + o\left(1\right)\right) \\ &= \frac{C}{h^{\delta}} \int \left|\xi_h\left(u\right)\right|^{1+\delta} \mathrm{d}u \left(1 + o\left(1\right)\right) \\ &\leqslant \frac{C}{h^{\delta}} \int \left|\nu_h\left(u\right)\right|^{2(1+\delta)} \mathrm{d}u, \end{aligned}$$

where the last inequality is obtained by simple change of variable and the application of Jensen's inequality with function $|\cdot|^{1+\delta}$.

Using the result of Lemma 4.3, one can obtain

$$E\left|\left\langle \tilde{U}_i, \tilde{U}_j \right\rangle\right|^{1+\delta} \leqslant C \times h^{1+\delta} \exp\left(\frac{\pi\left(1+\delta\right)}{h}\right),$$

where C is a constant. Similarly, one can show that this upper bound also holds for $E\left|\left\langle U_i, U_j \right\rangle\right|^{1+\delta}$. Then Lemma 1 of Yoshihara (1976) gives

$$\left|E\left\langle U_i, U_j \right\rangle\right| \leqslant C' \times h \exp\left(\frac{\pi}{h}\right) \beta\left(i-j\right)^{\frac{\delta}{1+\delta}} \left(i-j\right),$$

where C' is another constant. Plug this result into the sum

$$\begin{aligned}\sum_{1\leqslant i<j\leqslant n} E\left\langle U_i, U_j\right\rangle &= \sum_{k=2}^{n} E\left\langle U_1, U_k\right\rangle (n-k+1), \\ &\leqslant C' \times h \exp\left(\frac{\pi}{h}\right) \sum_{k=2}^{n} (n-k+1)\, k\beta\,(k)^{\frac{\delta}{1+\delta}}, \\ &= o_p\left(\frac{n}{2\pi^2}\exp\left(\frac{\pi}{h}\right)\right),\end{aligned}$$

where the last equality follows easily from the exponential decay rate of the β-mixing coefficients.

Combining the results in (4.19) and (4.20), it is shown that

$$ET_1 = \frac{1}{n^2} nE\left\|U_1\right\|^2 (1+o_p(1)) = \frac{1}{2\pi^2 n}\exp\left(\frac{\pi}{h}\right)(1+o_p(1)).$$

Limiting theorem for S_2 Recall from above that

$$S_2 = \frac{2}{n^2}\left(\sum_{1\leqslant i<j\leqslant n}\left\langle U_i, U_j\right\rangle - E\sum_{1\leqslant i<j\leqslant n}\left\langle U_i, U_j\right\rangle\right).$$

The order of $E\sum_{1\leqslant i<j\leqslant n}\langle U_i, U_j\rangle$ is known when calculating the order of T_1, so we focus on the other part: define

$$S_{21} := \frac{2}{n^2}\sum_{1\leqslant i<j\leqslant n}\left\langle U_i, U_j\right\rangle;$$

the sum $\sum_{1\leqslant i<j\leqslant n}\langle U_i, U_j\rangle$ is a U-statistic with kernel function

$$H_n(x,y) = \int (\omega_h(u-x) - K_h * g(u))\,(\omega_h(u-y) - K_h * g(u))\,\mathrm{d}u.$$

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

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

$\forall x \in \mathbb{R}$ holds. Now I check the conditions in the central limit theorem from Hjellvik, Yao, and Tjøstheim (1998) (collected in the Appendix 4.B as Proposition 4.1) with the U-statistic $\sum_{1\leqslant i<j\leqslant n}\langle U_i, U_j\rangle$, indexed by n.[6]

[6]Notice that the assumption $E[H_n(X_i, X_j)|\mathcal{F}_{j-1}] = 0$, 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.

First, the asymptotic variance σ_n^2 is calculated. This is done by calculating the asymptotically equivalent quantity $n^2\sigma_0^2/2$:

$$\begin{aligned}
& \sigma_0^2 \\
= & \int H_n^2(X_i, X_j)\mathrm{d}P(X_i)\mathrm{d}P(X_j) \\
= & \int H_n^2(X_1, X_2)\mathrm{d}P(X_1)\mathrm{d}P(X_2) \\
= & \int\int \left(\int \left(\omega_h(u - x_1) - K_h * g(u)\right)\left(\omega_h(u - x_2) - K_h * g(u)\right)\mathrm{d}u \right)^2 g(x_1)g(x_2)\mathrm{d}x_1\mathrm{d}x_2 \\
= & \int\int \left(\int \omega_h(u - x_1)\omega_h(u - x_2)\mathrm{d}u \right)^2 g(x_1)g(x_2)\mathrm{d}x_1\mathrm{d}x_2 \left(1 + o(1)\right)) \\
= & A(1 + o(1)),
\end{aligned}$$

where A is defined implicitly.

To evaluate the term A, define

$$\xi_h(x) = \int \nu_h(x + z)\nu_h(z)\mathrm{d}z.$$

Then

$$\begin{aligned}
A &= \frac{1}{h^4}\int\int \left(\int \nu_h\left(\frac{u - x_1}{h}\right)\nu_h\left(\frac{u - x_2}{h}\right)\mathrm{d}u \right)^2 g(x_1)g(x_2)\mathrm{d}x_1\mathrm{d}x_2 \\
&= \frac{1}{h^2}\int\int \left[\xi_h^2\left(\frac{x_1 - x_2}{h}\right)\right] g(x_1)g(x_2)\mathrm{d}x_1\mathrm{d}x_2 \\
&= \frac{1}{h}\left[\int \xi_h^2(x)\,\mathrm{d}x \times \|g\|^2 + o\left(\int \xi_h^2(x)\,\mathrm{d}x\right)\right],
\end{aligned}$$

where $\|g\|^2 = \int g^2(x)\mathrm{d}x$, and the last approximation is valid because

$$\begin{aligned}
& \left| \int\int \left[\frac{1}{h}\xi_h^2\left(\frac{x_1-x_2}{h}\right)\right] g(x_1)g(x_2)\mathrm{d}x_1\mathrm{d}x_2 - \int \xi_h^2(x)\,\mathrm{d}x \times |g|^2 \right| \\
= & \left| \int\int \xi_h^2(u)\, g(uh+x_2)\, g(x_2)\mathrm{d}u\mathrm{d}x_2 - \int \xi_h^2(x)\,\mathrm{d}x \int g^2(x)\mathrm{d}x \right| \\
= & \left| \int\int \xi_h^2(u)\,(g(uh+x_2)-g(x_2))\, g(x_2)\mathrm{d}u\mathrm{d}x_2 \right| \\
\leqslant & \int\int \xi_h^2(u)\,|g(uh+x_2)-g(x_2)|\, g(x_2)\mathrm{d}u\mathrm{d}x_2 \\
= & \int\int_{|uh|\leqslant\varepsilon} \xi_h^2(u)\,|g(uh+x_2)-g(x_2)|\, g(x_2)\mathrm{d}u\mathrm{d}x_2 \\
& + \int\int_{|uh|>\varepsilon} \xi_h^2(u)\,|g(uh+x_2)-g(x_2)|\, g(x_2)\mathrm{d}u\mathrm{d}x_2 \\
\leqslant & \int\int_{|uh|\leqslant\varepsilon} \xi_h^2(u)\, g(x_2)\mathrm{d}u\mathrm{d}x_2\, |uh|\, L + \\
& + \int\int_{|uh|>\varepsilon} \xi_h^2(u)\, g(x_2)\mathrm{d}u\mathrm{d}x_2 \times 2\sup_g |g|_\infty \\
\leqslant & \; o\left(\int \xi_h^2(x)\,\mathrm{d}x\right) + o(1),
\end{aligned}$$

by letting $\varepsilon \to 0$.

Now, since

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

so

$$A = \frac{1}{8\pi^2}\exp\left(\frac{2\pi}{h}\right) \times \|g\|^2,$$

and so

$$\begin{aligned}
\sigma_0^2 &= A(1+o(1)) \\
&= \frac{1}{8\pi^2}\exp\left(\frac{2\pi}{h}\right) \times \|g\|^2 \times (1+o(1)),
\end{aligned}$$

such that

$$\sigma_n^2 \sim \frac{n^2}{2}\sigma_0^2 = \frac{n^2}{16\pi^2}\exp\left(\frac{2\pi}{h}\right) \times \|g\|^2 \times (1+o(1))\,. \tag{4.21}$$

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

$$\begin{aligned} H_{ij}H_{kl} &= H_n(X_i, X_j)H_n(X_k, X_l) \\ &= \int (\omega_h(u - X_i) - K_h * g(u))\,(\omega_h(u - X_j) - K_h * g(u))\,\mathrm{d}u \\ &\quad \times \int (\omega_h(u - X_k) - K_h * g(u))\,(\omega_h(u - X_l) - K_h * g(u))\,\mathrm{d}u \\ &= \int \omega_h(u - X_i)\omega_h(u - X_j)\mathrm{d}u \int \omega_h(u - X_i)\omega_h(u - X_j)\mathrm{d}u(1 + o(1)), \end{aligned}$$

where only the dominating term is kept. Using the notation ξ_h defined above, one has

$$H_{ij}H_{kl} = \frac{1}{h^2}\xi_n\left(\frac{X_i - X_j}{h}\right)\xi_n\left(\frac{X_k - X_l}{h}\right).$$

First

$$M_{n1} = \max_{1<i<j\leqslant n} \max\left\{E\left|H_{1j}H_{ij}\right|^{(1+\delta)}, \int \left|H_{1j}H_{ij}\right|^{(1+\delta)} \mathrm{d}P_{X_1}P_{(X_i,X_j)}\right\}.$$

Denoting the joint density function of (X_1, X_i, X_j) as $g_{1,i,j}(y_1, y_i, y_j)$, one can write

$$\begin{aligned} & E\left|H_{1j}H_{ij}\right|^{1+\delta} \\ = & \left(\frac{1}{h^2}\right)^{1+\delta} \int\int\int \left|\xi_n\left(\frac{y_1 - y_j}{h}\right)\xi_n\left(\frac{y_i - y_j}{h}\right)\right|^{1+\delta} g_{1,i,j}(y_1, y_i, y_j)\mathrm{d}y_1\mathrm{d}y_i\mathrm{d}y_j \\ \leqslant & \left(\frac{1}{h^2}\right)^{1+\delta} h^2 \int\int\int \left|\xi_n(x)\,\xi_n(x+z)\right|^{1+\delta} g_{1,i,j}(xh + y_j, zh + y_j, y_j)\mathrm{d}x\mathrm{d}z\mathrm{d}y_j \\ \leqslant & Ch^{-2\delta} \int\int \left|\xi_n(x)\,\xi_n(x+z)\right|^{1+\delta} \mathrm{d}x\mathrm{d}z \\ = & Ch^{-2\delta} \left(\int \left|\xi_n(x)\right|^{1+\delta} \mathrm{d}x\right)^2 \\ \leqslant & Ch^{-2\delta} \int \left|\xi_n(x)\right|^{2(1+\delta)} \mathrm{d}x. \\ = & Ch^{-2\delta} \int \left|\int \nu_h(x+z)\nu_h(z)\mathrm{d}z\right|^{2(1+\delta)} \mathrm{d}x \\ \leqslant & Ch^{-2\delta} \int\int \left|\nu_h(x+z)\nu_h(z)\right|^{2(1+\delta)} \mathrm{d}z\mathrm{d}x \\ = & Ch^{-2\delta} \left(\int \left|\nu_h(z)\right|^{2(1+\delta)} \mathrm{d}z\right)^2. \end{aligned} \tag{4.22}$$

Using the result in Lemma 4.3, and continuing from (4.22),

$$\begin{aligned} E\left|H_{1j}H_{ij}\right|^{1+\delta} &\leqslant Ch^{-2\delta}\left(\int\left|\nu_h(z)\right|^{2(1+\delta)}\mathrm{d}z\right)^2 \\ &\leqslant Ch^{-2\delta}\left(h^{2\delta+1}\exp\left(\frac{\pi(1+\delta)}{h}\right)\right)^2 \\ &= Ch^{2\delta+2}\exp\left(\frac{\pi 2(1+\delta)}{h}\right). \end{aligned}$$

Similarly, it can be shown that the other term in the definition of M_{n1} also has this upper bound, such that

$$M_{n1} = O\left(h^{2\delta+2}\exp\left(\frac{\pi(1+\delta)}{h}\right)\right).$$

In the same fashion, it can be shown that

$$\begin{aligned} E\left|H_{1j}H_{ij}\right|^{2(1+\delta)} &\leqslant C_2h^{-2-4\delta}\left(\int\left|\xi_n\left(x\right)\right|^{2(1+\delta)}\mathrm{d}x\right)^2, \\ E\left|H_{1j}H_{ij}\right|^{2} &\leqslant C_3h^{-2}\left(\int\left|\xi_n\left(x\right)\right|^{2}\mathrm{d}x\right)^2, \\ E\left|H_{1i}H_{jk}\right|^{2(1+\delta)} &\leqslant C_4h^{-2-4\delta}\left(\int\left|\xi_n\left(x\right)\right|^{2(1+\delta)}\mathrm{d}x\right)^2, \end{aligned}$$

such that

$$\begin{aligned} M_{n2} &= O\left(h^{2+4\delta}\exp\left(\frac{4\pi(1+\delta)}{h}\right)\right), \\ M_{n3} &= O\left(\exp\left(\frac{4\pi}{h}\right)\right), \\ M_{n4} &= O\left(h^{2+4\delta}\exp\left(\frac{4\pi(1+\delta)}{h}\right)\right). \end{aligned}$$

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

$$\begin{aligned}
& E\left|\int H_{1i}H_{1j}\mathrm{d}P(X_1)\right|^{2(1+\delta)} \\
= \; & E\left|\frac{1}{h^2}\int \xi_n\left(\frac{y_1-X_i}{h}\right)\xi_n\left(\frac{y_1-X_j}{h}\right)g(y_1)\mathrm{d}y_1\right|^{2(1+\delta)} \\
\leqslant \; & E\left|\frac{1}{h}\int \xi_n\left(z\right)\xi_n\left(z+\frac{X_i-X_j}{h}\right)g(zh+X_i)\mathrm{d}z\right|^{2(1+\delta)} \\
\leqslant \; & C\times\left(\frac{1}{h}\right)^{2(1+\delta)} E\int\left|\xi_n\left(z\right)\xi_n\left(z+\frac{X_i-X_j}{h}\right)\right|^{2(1+\delta)}\mathrm{d}z \\
\leqslant \; & C\times\left(\frac{1}{h}\right)^{2(1+\delta)} h\int\int\left|\xi_n\left(z\right)\xi_n\left(z+x\right)\right|^{2(1+\delta)}\mathrm{d}z\mathrm{d}x \\
= \; & C\times h^{-1-2\delta}\left(\int\left|\xi_n(z)\right|^{2(1+\delta)}\mathrm{d}z\right)^2,
\end{aligned}$$

such that

$$M_{n5}=O\left(h^{3+6\delta}\exp\left(\frac{4\left(1+\delta\right)}{h}\pi\right)\right).$$

In the same fashion,

$$E\left|\int H_{1i}H_{1j}\mathrm{d}P(X_1)\right|^{2}\leqslant C_6h^{-1}\left(\int\left|\xi_n(z)\right|^{2}\mathrm{d}z\right)^2,$$

and

$$M_{n6}=O\left(h\exp\left(\frac{4\pi}{h}\right)\right).$$

Combining these upper bounds together with the variance order calculated in (4.8), we see that

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

as $n\to\infty$ and $h\to 0$.

Hence the limiting distribution for S_{21} is obtained:

$$\frac{2\pi n}{\exp(\pi/h)\|g\|}S_{21}\rightsquigarrow N\left(0,1\right).$$

This together with the result (4.20) gives the asymptotic distribution of S_2.

Order of S_1 One can expand

$$\mathrm{Var}\left(\sum_{i=1}^{n}\|U_i\|^2\right) = n\,\mathrm{Var}\left(\|U_1\|^2\right)+2\sum_{k=2}^{n}\mathrm{Cov}\left(\|U_1\|^2,\|U_k\|^2\right)\left(n-k+1\right),$$

using the stationarity of U_i.

For the first part, notice that

$$\begin{aligned}
\operatorname{Var}\left(\|U_1\|^2\right) &\leqslant E\left(\|U_1\|^4\right) \\
&= \int\left(\int\left(\omega_h\left(x-y\right)-K_h*g(x)\right)^2\mathrm{d}x\right)^2 g(y)\mathrm{d}y \\
&= \int\left(\int\left(\omega_h\left(x-y\right)\right)^2\mathrm{d}x\right)^2 g(y)\mathrm{d}y\left(1+o(1)\right) \\
&\leqslant \frac{C}{h^2}\left\|\nu_h\right\|_2^4 \\
&= C\times\exp\left(\frac{2\pi}{h}\right),
\end{aligned}$$

such that the first term

$$n\operatorname{Var}\left(\|U_1\|^2\right)=O\left(ne^{2\pi/h}\right).$$

For the second part, similarly to evaluating the sum of cross product terms in T_1, using Lemma 1 in Yoshihara (1976) one can obtain that

$$\begin{aligned}
\sum_{k=2}^{n}\operatorname{Cov}\left(\|U_1\|^2,\|U_k\|^2\right)(n-k+1) &\leqslant \exp\left(\frac{2\pi}{h}\right)M\sum_{k=2}^{n}\beta(k)^{\frac{\delta}{1+\delta}}(k-1)(n-k+1) \\
&= o\left(ne^{2\pi/h}\right).
\end{aligned}$$

The last equality is because of the exponential decay rate of β-mixing coefficients.

Summarizing the above results,

$$\begin{aligned}
\operatorname{Var}S_1 &= \frac{1}{n^4}\operatorname{Var}\left(\sum_{i=1}^{n}\|U_i\|^2\right) \\
&= O\left(\frac{1}{n^3}e^{2\pi/h}\right),
\end{aligned}$$

this together with the fact that $ES_1=0$ implies that,

$$S_1=O_p\left(\left(\frac{1}{n^3}\exp\left(\frac{2\pi}{h}\right)\right)^{1/2}\right)=o_p\left(\left(\frac{1}{n^2}\exp\left(\frac{2\pi}{h}\right)\right)^{1/2}\right).$$

□

Proof (of Theorem 4.2) Notice that

$$\begin{aligned} I_1 &= \int \left(\hat{g}(x) - K_h * g(x;\hat{\theta})\right)^2 \mathrm{d}x \\ &= \int \left[(\hat{g}(x) - K_h * g_1(x)) + \left(K_h * g_1(x) - K_h * g(x;\hat{\theta})\right)\right]^2 \mathrm{d}x \\ &= \int (\hat{g}(x) - K_h * g_1(x))^2 \,\mathrm{d}x + \int \left(K_h * g_1(x) - K_h * g(x;\hat{\theta})\right)^2 \mathrm{d}x \\ &\quad +2\int (\hat{g}(x) - K_h * g_0(x)) \left(K_h * g_0(x) - K_h * g(x;\hat{\theta})\right) \mathrm{d}x. \end{aligned}$$

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 it has the central limit theorem as in that theorem, the second part is nondegenerate, while the third part cannot dominate the other two because of Hölder's inequality, so when the second part is blow up by the rate $\left(4\pi^2 n^2 / \left(\exp(2\pi/h)\widehat{\|g\|_2^2}\right)\right)^{1/2}$, it will tend to ∞, such that $P(T_1 > Z_\alpha) \to 1$ is true. □

Proof (of Theorem 4.3) Notice that

$$\begin{aligned} I_1 &= \int \left(\hat{g}(x) - K_h * g(x;\hat{\theta})\right)^2 \mathrm{d}x \\ &= \int (\hat{g}(x) - K_h * g_n(x))^2 \,\mathrm{d}x + \int \left(K_h * g_n(x) - K_h * g(x;\hat{\theta})\right)^2 \mathrm{d}x \\ &\quad +2\int (\hat{g}(x) - K_h * g_n(x)) \left(K_h * g_n(x) - K_h * g(x;\hat{\theta})\right) \mathrm{d}x \\ &= A + B + C, \end{aligned}$$

where A, B and C are defined implicitly. Using similar method as in Theorem 4.1 it can be shown that under $\mathcal{H}'_{1n}$,

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

With Assumptions (B1) and (B2), it's not difficult to show that

$$B = \gamma_n^2 \left(\int \Delta^2(x)\mathrm{d}x + o(1)\right).$$

For part C, first notice that

$$\begin{aligned} C &= 2\gamma_n \int (\hat{g}(x) - K_h * g_n(x)) (\Delta(x) + o(1)) \,\mathrm{d}x \\ &= 2\gamma_n \int (\hat{g}(x) - K_h * g_n(x)) \,\Delta(x)\mathrm{d}x + \int (\hat{g}(x) - K_h * g_n(x)) \,\mathrm{d}x \times o(\gamma_n). \end{aligned}$$

It is not difficult to calculate that

$$E\int\left(\hat{g}(x)-K_h*g_n\left(x\right)\right)\Delta(x)\mathrm{d}x=0,$$

$$\begin{aligned}
&\mathrm{Var}\left(\int\left(\hat{g}(x)-K_h*g_n\left(x\right)\right)\Delta(x)\mathrm{d}x\right)\\
=\;& E\left(\frac{1}{n}\sum_{j=1}^{n}\int\left(\omega_h\left(x-X_j\right)-K_h*g(x)\right)\Delta(x)\mathrm{d}x\right)^2\\
=\;& \frac{1}{n}E\left(\int\left(\omega_h\left(x-X_1\right)-K_h*g(x)\right)\Delta(x)\mathrm{d}x\right)^2+\\
&\frac{2}{n^2}\sum_{j=1}^{n-1}\left(\left(n-j\right)E\int\left(\omega_h\left(x-X_1\right)-K_h*g(x)\right)\Delta(x)\mathrm{d}x\right.\\
&\qquad\left.\int\left(\omega_h\left(x-X_{j+1}\right)-K_h*g(x)\right)\Delta(x)\mathrm{d}x\right).
\end{aligned}$$

The sum of diagonal terms satisfies

$$\begin{aligned}
&\frac{1}{n}E\left(\int\left(\omega_h\left(x-X_1\right)-K_h*g(x)\right)\Delta(x)\mathrm{d}x\right)^2\\
=\;& \frac{1}{n}E\left(\int\omega_h\left(x-X_1\right)\Delta(x)\mathrm{d}x\right)^2\left(1+o(1)\right)\\
=\;& \frac{1}{n}E\left(\int\nu_h\left(x\right)\mathrm{d}x\Delta(X_1)+o(1)\right)^2\left(1+o(1)\right)\\
=\;& O\left(\frac{1}{n}\left(\int\nu_h\left(x\right)\mathrm{d}x\right)^2\right)\\
=\;& O\left(\frac{h}{n}\exp\left(\frac{\pi}{h}\right)\right).
\end{aligned}$$

For the cross product terms, using a similar approach as in earlier proofs to apply Lemma 1 of Yoshihara (1976), one can show that, for some $\delta>0$,

$$\begin{aligned}
&\frac{2}{n^2}\sum_{j=1}^{n-1}\left(\left(n-j\right)E\int\left(\omega_h\left(x-X_1\right)-K_h*g(x)\right)\Delta(x)\mathrm{d}x\right.\\
&\qquad\left.\int\left(\omega_h\left(x-X_{j+1}\right)-K_h*g(x)\right)\Delta(x)\mathrm{d}x\right)\\
\leqslant\;& C\frac{2}{n^2}h^{\frac{2\delta}{1+\delta}}\exp\left(\frac{\pi}{h}\right)\sum_{j=1}^{n-1}\left(n-j\right)\beta(j)^{\delta/(1+\delta)}.
\end{aligned}$$

This is $O\left((h/n)\exp\left(\pi/h\right)\right)$ for an appropriately chosen δ (for example $\delta = 1$), and noticing the exponential decay rate of the β-mixing coefficient. Summarizing the above calculations, the variance satisfies

$$\mathrm{Var}\left(\int \left(\hat{g}(x) - K_h * g_n\left(x\right)\right)\Delta(x)\mathrm{d}x\right) \;=\; O\left(\frac{h}{n}\exp\left(\frac{\pi}{h}\right)\right),$$

such that

$$C = O\left(\gamma_n\sqrt{\frac{h}{n}}\exp\left(\frac{\pi}{2h}\right)\right).$$

When h satisfies $h \to 0$ and $\exp\left(\pi/h\right)/n \to 0$, take $\gamma_n^2 = \exp\left(\pi/h\right)/n$, so it follows that

$$C = o\left(\gamma_n^2\right),$$

and the result is proved.

□

Technical lemmas

This section records several existing results that will be useful in the proof of this chapter.

The first result is the central limit theorem for U-statistics with β-mixing data by Hjellvik et al. (1998). Let $X_1, X_2, \ldots$ be a strictly stationary β-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 σ-algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$ for which $X_j \in \mathcal{F}_j$, and

$$E[H_n(x, X_1)] = 0,$$

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

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

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

$$M_{n1} = \max_{1<i<j\leqslant n}\max\left\{E\left|H_{1j}H_{ij}\right|^{(1+\delta)}, \int\left|H_{1j}H_{ij}\right|^{(1+\delta)}\mathrm{d}P_{X_1}\mathrm{d}P_{(X_i,X_j)}\right\},$$

$$\begin{aligned} M_{n2} \;=\; \max_{1<i<j\leqslant n}\max\Bigg\{&E\left|H_{1j}H_{ij}\right|^{2(1+\delta)}, \int\left|H_{1j}H_{ij}\right|^{2(1+\delta)}\mathrm{d}P_{X_1}\mathrm{d}P_{(X_i,X_j)}, \\ &\int\left|H_{1j}H_{ij}\right|^{2(1+\delta)}\mathrm{d}P_{(X_1,X_i)}\mathrm{d}P_{X_j}, \int\left|H_{1j}H_{ij}\right|^{2(1+\delta)}\mathrm{d}P_{X_1}\mathrm{d}P_{X_i}\mathrm{d}P_{X_j},\Bigg\} \end{aligned}$$

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

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

$$M_{n5} = \max_{1<i<j} \max \left\{ E\left| \int H_{1i}H_{1j}\mathrm{d}P_{X_1} \right|^{2(1+\delta)}, \int \left| \int H_{1i}H_{1j}\mathrm{d}P_{X_1} \right|^{2(1+\delta)} \mathrm{d}P_{X_i}\mathrm{d}P_{X_j} \right\},$$

$$M_{n6} = \max_{1<i<j} E\left| \int H_{1i}H_{1j}\mathrm{d}P_{X_1} \right|^2.$$

Proposition 4.1 *(Hjellvik et al. (1998) Theorem A in Appendix 1) If for some $\delta > 0$,*

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

and

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

as $n \to \infty$, then

$$\sigma_n^{-1} \sum_{1 \leqslant i < j \leqslant n} H_n(X_i, X_j) \rightsquigarrow N(0,1).$$

Define

$$\sigma_0^2 = \int H_n^2(X_i, X_j)\mathrm{d}P(X_i)\mathrm{d}P(X_j),$$

which is the variance of H_{ij} as if the data were independent, σ_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 4.1. For a sequence of k random variables $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$, with $i_1 < i_2 < \ldots < i_k$, for any $j (1 \leqslant j \leqslant k-1)$, define

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

and

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

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 splitted in two independent groups of size j and $(k-j)$.

Lemma 4.2 *(Lemma 1 in Yoshihara (1976)) For any $0 \leqslant j \leqslant k-1$, let $h(x_{i_1}, \ldots, x_{i_k})$ be a Borel function such that*

$$\int \ldots_{\mathbb{R}^k} \int |h(x_{i_1}, \ldots, x_{i_k})|^{1+\delta} \mathrm{d}P_j^{(k)} \leqslant M,$$

for some $\delta > 0$. Then

$$\begin{aligned} & \left| \int \ldots_{\mathbb{R}^k} \int h(x_{i_1}, \ldots, x_{i_k}) \mathrm{d}P_j^{(k)} - \int \ldots_{\mathbb{R}^k} \int h(x_{i_1}, \ldots, x_{i_k}) \mathrm{d}P_0^{(k)} \right| \\ \leqslant \;& 4M^{\frac{1}{1+\delta}} \beta(k)^{\frac{\delta}{1+\delta}} \left(i_{j+1} - i_j\right). \end{aligned}$$

Lemma 4.3 *For $p > 2$, the p-norm of the deconvolution kernel satisfies*

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

and

$$\int |\nu_h(z)|^p \, \mathrm{d}z \leqslant C \times h^{p-1} \exp\left(\frac{\pi p}{2h}\right)$$

where C is a constant.

Proof In general for $p > 2$, we known from Van Es et al. (2003) and Masry (1991) that

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

This is simple to show by noticing $\int |\nu_h(x)|^p \, \mathrm{d}x \leqslant \int |\nu_h(x)|^2 \left|\sup_x \nu_h(x)\right|^{p-2} \mathrm{d}x$. Therefore,

$$\begin{aligned} \int |\nu_h(z)|^p \, \mathrm{d}z &\leqslant \|\nu_h\|_\infty^{p-2} \|\nu_h\|_2^2 \\ &\leqslant C \times h^{p-2} \exp\left(\frac{\pi(p-2)}{2h}\right) \times h \exp\left(\frac{\pi}{h}\right) \\ &= C \times h^{p-1} \exp\left(\frac{\pi p}{2h}\right). \end{aligned}$$

This is because

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

and

$$\begin{aligned} \|\nu_h\|_2^2 &= \int |\nu_h(x)|^2 \, \mathrm{d}x \\ &= \frac{h}{2\pi^2} \exp\left(\frac{\pi}{h}\right), \end{aligned}$$

which is known from (4.19). □