School of Arts and Social Sciences
CITY UNIVERSITY LONDON

Department of Economics

Consistent nonparametric specification tests for stochastic volatility models based on the return distribution

Yang Zu[1]
City University London

H. Peter Boswijk[2]
University of Amsterdam

Department of Economics
Discussion Paper Series
No. 15/02

[1] Corresponding author: 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.

[2] Corresponding author: Email: H.P.Boswijk@uva.nl. Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE Amsterdam, The Netherlands. Telephone: +31 (0)20 525 4316.

Consistent nonparametric specification tests for stochastic volatility models based on the return distribution

Yang Zu[1]
City University London

H. Peter Boswijk[2]
University of Amsterdam

April 29, 2015

[1]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.

[2]Email: H.P.Boswijk@uva.nl. Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE Amsterdam, The Netherlands. Telephone: +31 (0)20 525 4316.

Abstract

This paper develops nonparametric specification tests for stochastic volatility models by comparing the nonparametically estimated return density and distribution functions with their parametric counterparts. Asymptotic null distributions of the tests are derived and the tests are shown to be consistent. Extensive Monte Carlo experiments are performed to study the finite sample properties of the tests. The tests are applied to an empirical dataset and we find the estimated stochastic volatility model is misspecified.

JEL Classification: C58, C12, C14.

Keywords: nonparametric test, stochastic volatility models.

1 Introduction

In this paper we consider specification tests for a class of parametric stochastic volatility models, given by

$$\begin{aligned} \mathrm{d}Y_t &= \sigma_t \mathrm{d}B_t, \\ \mathrm{d}\sigma_t^2 &= b(\sigma_t^2;\theta)\mathrm{d}t + a(\sigma_t^2;\theta)\mathrm{d}W_t, \end{aligned} \tag{1}$$

where $(B_t, W_t)_{t\geq 0}$ is a bivariate standard Brownian motion process, where b and a are known functions, and where θ is an unknown parameter vector. The model is tested within a larger class of nonparametric stochastic volatility models

$$\mathrm{d}Y_t = \sigma_t \mathrm{d}B_t, \tag{2}$$

where $(\sigma_t)_{t\geq 0}$ is a stochastic process satisfying certain regularity conditions. The model (2) is nonparametric in the sense that there is no parametric structure specified for the volatility process. Model (1) is often used in financial econometrics to describe a logarithmic stock price process $(Y_t)_{t\geq 0}$, where $(\sigma_t)_{t\geq 0}$ is an unobserved spot volatility process. It includes popular models such as the Hull-White model, the Heston model and the GARCH diffusion model, which motivates the development of specification tests for this class of models.

Let Y be observed discretely at times $t_i = i\Delta$, $i = 0, 1, \ldots, n$. Consider the re-scaled Δ-period return sequence

$$X_i = \frac{1}{\sqrt{\Delta}}(Y_{t_i} - Y_{t_{i-1}}) = \frac{1}{\sqrt{\Delta}}\int_{t_{i-1}}^{t_i} \sigma_s \mathrm{d}B_s, \qquad i = 1, \ldots, n. \tag{3}$$

Let $(X_i)_{i=1}^n$ having a stationary density, denoted by $q(x)$, and let $q(x;\theta)$ be its specification implied by the parametric model (1). In this paper we propose to test the specification (1) by comparing the estimated parametric return density to its nonparametrically estimated counterpart. Stated formally, we are testing

$$\mathcal{H}_0 : q(x) = q_0(x) \in \{q(x;\theta), \theta \in \Theta\},$$

where $\Theta \subseteq \mathbb{R}^k$ is the parameter space, and define θ_0 to be the true parameter under the null hypothesis: that is, it satisfies $q(x;\theta_0) = q_0(x)$.

Specification tests based on the stationary marginal return distribution have their empirical justifications — as discussed in Section 3.3 of Aït-Sahalia, Hansen, and Scheinkman (2010), reproducing the stationary distribution is an important aspect of structural economic modeling. Return distribution is also widely used as the basis to formulate specification tests for continuous-time diffusion processes, see e.g. Aït-Sahalia (1996) and Gao and King (2004). Admittedly, formulating the test based on $q(.;\theta)$ will limit its power in detecting certain deviations in the functions $\{b(.;\theta), a(.;\theta)\}$[1]; however, tests constructed this way would still be an important "first check" because of its empirical significance in any structural modelling. The problem could be solved by defining test statistics based on the transition distribution of the observed returns, we discuss this issue in Section 7.

To formulate the test statistic, one can compare either the density functions or the cumulative distribution functions. It is known from the literature that generally speaking, density-based tests are more sensitive to local deviations, whereas distribution-based tests are more sensitive to global deviations (see e.g. Eubank and LaRiccia (1992), Escanciano (2009) and Aït-Sahalia, Fan, and Peng (2009)), we thus consider both in this paper.

A long-span asymptotic scheme is used in this paper. That is, we consider the asymptotics when $n \to \infty$ with fixed Δ. This is because model (1) is often used to describe price processes observed at relatively low frequencies (usually daily), prominent microstructure noise effects in prices observed at higher frequencies make the model unsuitable for such data. Throughout, we assume that $(X_i)_{i=1}^n$ is a stationary and ergodic sequence, and that it is β-mixing with exponentially decaying coefficients. In Appendix A, checkable sufficient conditions for these properties to hold in the parametric model are given.

The stochastic volatility model we consider here is essentially a (partially observed) two dimensional diffusion process, so our test is related to the vast literature of nonparametric test for diffusion models, such as Aït-Sahalia (1996), Hong and Li (2005), Corradi and Swanson (2005), Li (2007), Chen et al. (2008), Aït-Sahalia et al. (2010), Kristensen (2011) and Aït-Sahalia and Park (2012), among others. However, the unobservability of the volatility process in Model (1) makes the aforementioned research not applicable. Closely related to this paper is Corradi and Swanson (2011), who consider a conditional distribution based nonparametric test for stochastic volatility models. Other than the test statistics we consider are different from that of Corradi and Swanson, the model we consider is also different. Corradi and Swanson consider the stochastic volatility model where the observed series is assumed to be strictly stationary; while in our model we assume the observed return series (first difference of the observed series) to be stationary and we allow the observed series to exhibit say unit-root type of dynamics. From a practical perspective, the stochastic volatility model considered in Corradi and Swanson in more appropriate to be used with interest rate data, where mean-reversion is often observed; while our model is more appropriate for equity and exchange rate price data, where unit-root behaviour is often observed. Zu (2015) analyzes an alternative approach to a similar testing problem, by comparing the nonparametric kernel deconvolution estimator of the volatility density with its parametric counterpart.

The structure of this paper is as follows. Sections 2 and 3 discuss nonparametric and parametric estimation of the return density and distribution functions, respectively.

[1] That is, there might exist two different specifications $\{b(.;\theta), a(.;\theta)\}$ and $\{\tilde{b}(.;\theta), \tilde{a}(.;\theta)\}$ leading to the same return distribution with density $q(x)$.

Section 4 defines the test statistics, derives their asymptotic null distributions and consistency, and discusses using the bootstrap to approximate the null distribution. In section 5 Monte Carlo evidence for the size and power properties of the tests are given. In Section 6 we study an empirical application. Section 7 discusses possible extensions and concludes. Technical assumptions are collected in Appendix A. The proofs for the theorems are collected in Appendix B.

2 Nonparametric estimation

We now discuss the nonparametric estimation of density and distribution functions. In the nonparametric model (2), estimation of the stationary marginal return density and distribution functions is considered under the direct assumption that the sequence $(X_i)_{i=1}^n$ is stationary, ergodic and β-mixing with exponentially decaying coefficients.

To estimate $q(x)$, the stationary return density function of $(X_i)_{i=1}^n$. Let h_n be a bandwidth, $K(.)$ be a kernel function. It is well known that the density function $q(x)$ can be estimated by the kernel density estimator

$$\hat{q}(x) = \frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right).$$

Under appropriate conditions on the bandwidth parameter and the kernel function, the consistency and asymptotic distribution of the kernel density estimator are classical results, we refer the readers to e.g. Pagan and Ullah (1999).

Denote the distribution function of the sequence $(X_i)_{i=1}^n$ to be $Q(x)$. Letting $I(.)$ denote the indicator function, the distribution function $Q(x)$ can be estimated by the empirical distribution function

$$\hat{Q}(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leqslant x).$$

The consistency and asymptotic normality of the empirical distribution function are classical results in statistics, see e.g. Chapter 19 of Van der Vaart (2000) for the results with independent and identically distributed data. For stationary dependent data, such properties still hold with the application of Ergodic Law of Large Numbers and Central Limit Theorem for dependent data, see Appendix A.5 in Pagan and Ullah (1999) for a summary.

3 Parametric estimation

Given a parameterization $\{b(x;\theta), a(x;\theta)\}$, to obtain the parametric estimate of the functions $q(x;\theta)$ and $Q(x;\theta)$, we first need an estimate for the parameter vector, denoted as $\hat{\theta}$, then evaluate the two functions given $\hat{\theta}$.

On the one hand, parametric estimation of stochastic volatility model is by no means an easy task; substantial research efforts were devoted to it in the past decades. Here we first briefly review the existing methods and just assume we have a parametric estimator satisfying certain conditions. On the other hand, evaluating the two functions given $\hat{\theta}$ is also not trivial, because the density and distribution functions of the observed returns usually do not have closed-form expressions and one needs to resort to approximation methods to evaluate them.

3.1 Parametric estimation of stochastic volatility models

Many efforts have been devoted to the estimation of stochastic volatility models in the past decades. For a review, see e.g. Renault (2009). Here we do not confine to any particular parametric estimation method, but only give conditions that a parametric estimator should satisfy. We will need different assumptions for the density function based test and the distribution function based test. For the density function based test, we only need to assume the parametric estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent. We will also need the parametrization to be smooth.

(P1a) Under the null hypothesis,

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

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

For the distribution function based test, however, stronger assumptions are needed — the estimator has to satisfy a certain first order asymptotic expansion, which will be a non-vanishing part of the asymptotic distribution. We also need the parameterization to be differentiable.

(P1b) Under the null hypothesis,

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\theta_0}(X_i) + o_p(1),$$

with $P_{\theta_0}\psi_{\theta_0} = 0$ and $P_{\theta_0}\|\psi_{\theta_0}\|^2 < \infty$, and $Q(x;\theta)$ is differentiable with respect to θ.

3.2 Approximating parametric density and distribution function

When no closed-form expressions for the density and distribution functions exist, we can in principle use an Euler scheme to simulate the process and hence evaluate intractable functionals of the process.

Given an estimate $\hat{\theta}$, the parameterization $b(.,\theta)$ and $a(.,\theta)$, the observation interval Δ, and the objective variables $X_i = \frac{1}{\sqrt{\Delta}} \int_{t_{i-1}}^{t_i} \sigma_s \mathrm{d}B_s, i = 1, \ldots, n$, to be approximated, we first choose an integer m as the steps to simulate *within* the interval Δ, and another integer M as the number of Δ-interval returns, such that we simulate the process Y with step size $\delta = \Delta/m$ for $m \times M$ steps. Then take first differences to get δ-returns, and aggregate and rescale over every m returns to get M simulated Δ-returns, X_i^*, $i = 1, \ldots, M$. Using a kernel density estimator we can approximate $q(x;\hat{\theta})$ from the simulated sample with

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

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

Standard consistency results for the kernel density estimator and convergence theorems for the Euler scheme simulation of stochastic differential equations imply that when $M \to \infty, h_M \to 0$ and $m \to \infty$, $q^*(x;\hat{\theta}) \to q(x;\hat{\theta})$ pointwise in $x \in \mathbb{R}$. The convergence should be understood as in the probability space of Monte Carlo simulation. For the technical conditions on the kernel function $K(.)$, bandwidth h_M and the consistency result for the kernel density estimator, we refer to, e.g. Section 2.6.2 of Pagan and Ullah (1999). For the convergence result of the Euler simulation method, we refer to Chapter 9 of Kloeden and Platen (1992). The accuracy of this approximation is determined by the number M and m that we choose. Because these numbers do not have to be bounded by the sample size n, they can be chosen very large to make the approximation error arbitrarily small. The parametric distribution function $\hat{Q}(x,\hat{\theta})$ can be approximated analogously using the empirical distribution function with the simulated data. In this reason, in the following we take the approximated $q^*(x;\theta)$ and $Q^*(x;\theta)$ to be equal to the corresponding exact ones $q(x;\theta)$ and $Q(x;\theta)$ to avoid complication.

Remark 1 With the purpose to approximate the conditional distribution in a stochastic volatility model, Bhardwaj, Corradi, and Swanson (2008) propose to simulate multiple paths of fixed length starting from a common initial value and take average across these paths. We remark that when the simulated process is stationary, our simulation of a long path can be understood as multiple paths of shorter length and thus the two approaches can both simulate the intended observations. One difference is that our paths have different starting values while they have a common one. They impose this restriction in order to evaluate the conditional distribution given a certain value. However, we remark that for the purpose of conditional distribution function approximation given a certain value, simulating paths conditional on that value is not necessary.

Remark 2 Bhardwaj, Corradi, and Swanson (2008) notice that the Milestein scheme simulation for the stochastic volatility is not convergent if the commutative condition is not satisfied, which is the case for most of the stochastic volatility models with leverage effects. They use a generalized Milestein scheme from Kloeden and Platen (1992) to deal with this problem. We emphasize that the Euler's scheme is valid both in univariate and multivariate diffusions, and it is convergent both in the weak and in the strong sense when the simulating interval goes to 0. When used with stochastic volatility model, usually the Eulers Scheme is applied to a finer grid *within* the needed sampling interval, as in the method used in this paper. This will not cause the "stochastic integral" problem as discussed in Bhardwaj, Corradi, and Swanson (2008). Actually the Euler's scheme is widely used in the literature to simulate stochastic volatility models with leverage effects. This include Andersen and Lund (1997), Bollerslev and Zhou (2002), Aït-Sahalia and Kimmel (2007) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), to name just a few.

4 Test statistics and asymptotic properties

4.1 Asymptotic null distribution and the consistency

Define

$$T_1 = nh^{1/2} \int_{\mathbb{R}} (\hat{q}(x) - K_h * q(x;\hat{\theta}))^2 \mathrm{d}x,$$

where $K_h * q(x;\hat{\theta}) = \int_{\mathbb{R}} K_h(x-y) q(y;\hat{\theta}) \mathrm{d}y$ is the convolution of $K_h(x) = K(x/h)/h$ with $q(x;\hat{\theta})$, the function $K(.)$ and bandwidth h are the same as used in the definition of $\hat{q}(x)$. Using the convoluted return density in the formulation of the test statistic corrects the bias of the test statistic and deliver better asymptotic properties of the test statistic, we refer to Fan (1994) for a discussion of this issue in the general density test problem with *i.i.d.* data.

Theorem 1 *Under the null hypothesis, and if (SV0)–(SV5), (N1)–(N4), and (P1a) are satisfied, then*

$$\begin{aligned} &\left(T_1 - h^{-1/2} \int_{\mathbb{R}} K^2(u) \mathrm{d}u\right) \\ &\xrightarrow{d} \; N\left(0, 2\int_{\mathbb{R}} q_0^2(x) \mathrm{d}x \int_{\mathbb{R}} \left(K^{(2)}(v)\right)^2 \mathrm{d}v\right), \end{aligned}$$

where $K^{(2)}(v)$ denote the convolution of the kernel function K with itself. Let

$$\hat{\sigma}^2 = \frac{2}{n}\sum_{i=1}^{n}\hat{q}(X_i)\int_{\mathbb{R}}\left(K^{(2)}(v)\right)^2 \mathrm{d}v,$$

which is a consistent estimator of the variance of the asymptotic distribution, then

$$T_2 = \frac{T_1 - h^{-1/2}\int_{\mathbb{R}} K^2(u)\mathrm{d}u}{\hat{\sigma}} \xrightarrow{d} N(0,1).$$

The test T_1 is not pivotal as it depends on the unknown density $q_0(x)$. However the corresponding studentized test T_2 is pivotal.

A Cramer-von Mises type statistic can be formulated by comparing distribution estimates:

$$T_3 = n\int_{\mathbb{R}}(\widehat{Q}(x) - Q(x;\hat{\theta}_n))^2\mathrm{d}Q(x;\hat{\theta}).$$

Theorem 2 *Under the null hypothesis, and if (SV0)–(SV5), (N2) and (P1b) are satisfied, then as $n \to \infty$*

$$T_3 \xrightarrow{d} \int_{\mathbb{R}}\left(\mathbb{G}_Q I(\cdot \leq x) - \mathbb{G}_Q\psi_{\theta_0}^T(\cdot)\frac{\partial Q(x,\theta)}{\partial\theta}|_{\theta=\theta_0}\right)^2 \mathrm{d}Q(x), \tag{4}$$

where $\mathbb{G}_Q$ is a Q-Brownian bridge indexed by $\mathcal{F} = \{I(\cdot \leq x), x \in \mathbb{R}\} \cup \{\psi_\theta(\cdot)\}$, with zero mean and the covariance function $\Gamma(f,g) = \lim_{k\to\infty}\sum_{i=1}^{\infty}\mathrm{Cov}(f(X_k), g(X_i))$ with $f, g \in \mathcal{F}$.

The above limiting distribution of T_3 is a functional of a Brownian bridge process, and it depends on the model structure (thus not model-free) as well as the unknown parameter values. For this reasons, this limiting theorem cannot be used directly to define critical values of the test. We discuss the approximation method to obtain test critical values in the next Section 4.2.

We then look at the asymptotic power of these tests under fixed alternatives. To be specific, we consider

$$\mathcal{H}_1 : \{q(x) = q_1(x) \neq q(x;\theta), \forall\theta \in \Theta\}.$$

We will need assumptions on the parametric estimator under the alternative model.

(P1a1) Under the alternative hypothesis $\mathcal{H}_1$,

$$|\hat{\theta} - \theta^*| = O_p(n^{-1/2}),$$

where θ^* is the pseudo true value of the model corresponding to $q_1(x)$.

Theorem 3 *Assume Conditions (SV0)–(SV5) and Assumptions (N1)–(N4), and (P1a1) in Appendix A; 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* $\mathcal{H}_1$*,*

$$P\left(T_1 - h^{-1/2}\int_{\mathbb{R}} K^2(u)\mathrm{d}u > \hat{\sigma} Z_{1-\alpha}\right) \to 1,$$

and

$$P\left(T_2 > Z_{1-\alpha}\right) \to 1.$$

For the distribution based test, we assume that under the alternative hypothesis the parametric estimator satisfy

(P1b1) Under the alternative hypothesis $\mathcal{H}_1$,

$$\sqrt{n}(\hat{\theta} - \theta^*) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_{\theta^*}(X_i) + o_p(1),$$

with $P_{\theta^*}\psi_{\theta^*} = 0$ and $P_{\theta^*}\|\psi_{\theta^*}\|^2 < \infty$, and $Q(x;\theta)$ is differentiable with respect to θ.

Theorem 4 *Under the null hypothesis, and if (SV0)–(SV5), (N2) and (P1b) are satisfied; let* $\alpha \in (0,1)$ *be a level of significance, and* $c_{1-\alpha}$ *be the* $1-\alpha$ *quantile of the limiting distribution in (4), then as* $n \to \infty$*,*

$$P\left(T_3 > c_{1-\alpha}\right) \to 1.$$

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.

4.2 Bootstrap null distribution

In the literature of nonparametric goodness of fit test, it is a usual problem that the asymptotic distribution of the density based nonparametric test statistic will provide a poor approximation for the null distribution in finite sample. For the distribution function based test, as we noted before that the asymptotic distribution of the test statistic is not feasible. These problems lead us to develop a bootstrap approximation to the null distribution, which is a common practice in nonparametric tests. Different bootstrap methods can be considered in nonparametric test. Parametric-type of bootstrap has been considered in e.g. Fan (1995), Andrews (1997), Franke, Kreiss, and Mammen (2002), Andrews (2005), Gao and Gijbels (2008) and Aït-Sahalia, Fan, and Peng (2009)among others. A nonparametric-type block bootstrap has been considered in Bhardwaj, Corradi, and Swanson (2008), Corradi and Swanson (2011) and several of their other works.

We use a parametric bootstrap procedure to approximate the distributions of the tests under the null hypothesis. Parametric bootstrap is also called model based bootstrap. As contrary to classical bootstrap, where one generate bootstrap samples by resampling the available dataset, parametric bootstrap involves generating bootstrap samples by first estimating a parametric model and then simulating data from the estimated parametric model (see Section 6.5 of Efron and Tibshirani (1994)). The dependence of the bootstrap sample on the original data is only through the estimated parameters. The parametric bootstrap is in particular useful in approximating the null distribution in a testing context because it always simulates data based on the null model: it will mimic the null model both under the null hypothesis and under the alternative hypothesis. In contrast, bootstrap procedures that do not exploit the model structures will usually mimic the data generating process (alternative model) under the alternative hypothesis. For example, in testing diffusion models, Corradi and Swanson (2011) use a block bootstrap procedure. Since the block bootstrap procedure mimics the data generating process under the alternative hypothesis, the bootstrapped statistic cannot reproduce the null distribution under a misspecified model, and they further define a re-centered test statistic to make the block bootstrap to work.

The parametric bootstrap procedure is as follows (use T_1 as an example):

Step 1 Given a parametric estimate $\hat{\theta}$, and step size Δ, simulate n (original sample size) discretely observed returns $\{X_i^*\}_{i=1}^n$, which is called one bootstrap sample. Notice this step has to be done using the method in Section 3.2 over a finer grid.

Step 2 With this bootstrap sample, compute the nonparametric estimator $\hat{q}^*(x)$ and the parametric estimator $\hat{\theta}^*$, then compute the test statistic T_1^* analogous to T_1.

Step 3 Repeat step 1 and 2 B times to get a bootstrap sample $T_1^{*1}, \ldots, T_1^{*B}$ for the statistic T_1.

When B is large, the empirical distribution of $T_1^{*1}, \ldots, T_1^{*B}$ approximates the finite sample null distribution.

Theoretical justification of the proposed parametric procedure is missing in this paper. This is a highly non-trivial problem, although it can probably be solved using the methodology developed in Fan (1994) and Andrews (1997). In absence of such results, we use extensive Monte Carlo simulations to study the power properties of the tests under various realistic scenarios and across different sample sizes in the next section.

5 Monte Carlo simulations

In this section, we study the finite sample performance of the density-based tests T_1 and T_2 and the distribution-based test T_3. We use the bootstrap method described in the previous section to determine the null distribution of the test statistic. The empirical quantiles of the null distribution are used to determine the critical values. The cross-validation method (e. g. Wasserman (2004), Section 20.3) is used to determine the bandwidth, and we use the Gaussian kernel in all the nonparametric kernel density estimators. The GMM method of Meddahi (2002) is used to estimate the parametric model. The GMM method is less efficient than the likelihood based method, but it achieves a good compromise between the estimation efficiency and computation time. To save space, for all the simulated size and power, we only report the results at 5% significance level.[2]

5.1 Size of the tests

We simulate 1000 sample paths of daily observations ($\Delta = 1/252$) from the Heston model,

$$\begin{aligned} \mathrm{d}Y_t &= \sigma_t \mathrm{d}W_t, \\ \mathrm{d}\sigma_t^2 &= 5(0.1 - \sigma_t^2)\mathrm{d}t + 0.75\sqrt{\sigma_t^2}\mathrm{d}B_t, \end{aligned} \tag{5}$$

where the two Brownian motions W and B are independent. Within one day, 10 steps are simulated to reduce discretization error. We consider the sample sizes 1000, 2000 and 3000, roughly corresponds to 4 years, 8 years and 12 years of daily observations.

With the parameters and sample sizes, the test statistics T_1, T_2 and T_3 are simulated 1000 times. The distribution of these realized test statistics are taken as the true distribution (except for the Monte Carlo errors). For each of the realized 1000 sample path, we obtain 5 bootstrap samples and compute their resulting test statistics T_1^*, T_2^* and T_3^*. Aggregating them together across 1000 sample yields 5000 bootstrap statistics. Their sampling distribution is taken as the distribution of the bootstrap method.

Table 1 summarizes the simulated 5% level size of all the tests for the three sample sizes. The bootstrap test statistics seem to have a reasonable size property, especially when the sample size is large. For T_1 and T_3, the size property becomes better as the sample size increases, though this is not the case for T_2.

[Table 1 about here.]

5.2 Power of the tests

We study the power performance of the tests under four different sample sizes 500, 1000, 2000 and 3000, and we still use 1000 simulations. We take the Heston model (5) as the null hypothesis. We evaluate the power functions of the three test statistics under the

[2]The computations in this section are conducted with Matlab® 2012b on the Lisa computing cluster at the University of Amsterdam. The default random number generator in Matlab (The Mersenne twister algorithm) is used with seed 12345.

three families of alternative models. Each family of models is indexed by τ, with $\tau = 0$ corresponding to the null model (5).

In the first family of alternative models, the drift functions of the volatility processes are deviated from the Heston model. In the second family of models, the diffusion functions are deviated from the Heston model. In the third family of models, jumps are included in the volatility process.

5.2.1 Misspecification in the drift function

We evaluate the power function of the three test statistics under the following sequence of alternative models,

$$\mathrm{d}\sigma_t^2 = \{(1-\tau)(\alpha(\beta - \sigma_t^2) + \tau\mu(\sigma_t^2)\}\mathrm{d}t + \gamma\sqrt{\sigma_t^2}\mathrm{d}W_t, \tag{6}$$

for $\tau = 0, 0.1, ..., 1$, where $\mu(\sigma_t^2) = \sigma_t^2[a(b - \ln\sigma_t^2)]$ with $a = 9$, $b = 3.5$. The functional form of the drift part is motivated by the log SARV model

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

By Ito's lemma, the volatility process of the log SARV model is

$$\mathrm{d}\sigma_t^2 = \sigma_t^2\left[\kappa(\theta - \ln\sigma_t^2) + \frac{1}{2}\gamma^2\right]\mathrm{d}t + \gamma\sigma_t^2\mathrm{d}W_t,$$

where the drift function is a highly nonlinear function of σ_t^2 and we use this to determine the specification of $\mu(\sigma_t^2)$.

Figure 1 shows the differences of the drift functions between the null model and the alternative models. It also gives the 5% level power functions of the three tests at the 4 different sample sizes. All the three tests show increasing power as the sample size goes large, confirming the consistency result of the tests. The performance of the three tests seems to be similar for this type of deviations in the drift function.

[Figure 1 about here.]

5.2.2 Misspecification in the diffusion function

Here we consider the misspecification in the diffusion function of the volatility process. The null model is still the Heston model (5). In the alternative model, the drift function remains the same, but the diffusion function is deviating away to a GARCH diffusion process.

$$
\begin{aligned}
\mathrm{d}Y_t &= \sigma_t \mathrm{d}B_t, \\
\mathrm{d}\sigma_t^2 &= \alpha(\beta - \sigma_t^2)\mathrm{d}t + \{(1-\tau)\gamma\sqrt{\sigma_t^2} + \tau\rho(\sigma_t^2)\}\mathrm{d}W_t
\end{aligned}
\tag{7}
$$

for $\tau = 0, 0.1, ..., 1$, where $\rho(\sigma_t^2) = c\sigma_t^2$ with $c = 5$. When $\tau = 1$, the alternative model is a GARCH diffusion process.

Figure 2 shows the differences of the diffusion functions between the null model and the alternative models. It also gives the 5% level power functions of the three tests at the 4 different sample sizes. Again all the three tests show increasing power (to 1) as the sample size goes large. The power of T_1 seems to be slightly better than T_3 when the deviation is small, while when the deviation is large, the power seems to be similar. The power of T_2 seems to be lower than the other two tests; when the sample size is small ($n = 500$), the test T_2 seems to have no power at all.

[Figure 2 about here.]

5.2.3 Jumps in the volatility process

We now consider the power of the three tests against a sequence of models where the volatility process contains jumps

$$
\mathrm{d}\sigma_t^2 = \alpha(\beta - \sigma_t^2)\mathrm{d}t + \gamma\sqrt{\sigma_t^2}\mathrm{d}B_t + J_{t-}\mathrm{d}N_t, \tag{8}
$$

where N_t is a Poisson process with intensity λ. J_t is the jump size that is independent of (W_t) and (B_t). We consider the following 5 jump intensities: 52, 104, 252, 252×1.5, 252×2, these can be understood as the average number of jumps in a year. We consider a jump size that is normally distributed with mean 0 and standard deviation equal to 1.5%.

Figure 3 gives the power functions of the three tests under different sample sizes. We observe again the consistency of the three tests for this type of deviations to the null hypothesis. Also we see that as the jump intensity of the volatility process increases, the tests are more powerful in detecting the deviations. The relative performance of the three tests seems to be similar to this type of deviations.

[Figure 3 about here.]

6 Empirical application

In this section, we apply our tests to a daily Apple stock price dataset. The dataset contains the adjusted close prices from Jan 3rd 2000 to 3rd Feb 2014, making 3543 observations in total. Figure 4 gives the plot of the series, the log returns, the nonparametrically estimated density function, as well as the empirical distribution function of the dataset.

[Figure 4 about here.]

The dataset is fitted to the Heston model (5). The estimated parameter is $\hat{\alpha} = 17.5119$, $\hat{\beta} = 0.1793$, $\hat{\gamma} = 2.4715$. We then apply the nonparametric specification tests proposed in this paper to study the validity of the Heston model estimated. We still use the cross-validation method to determine the bandwidth and use the Gaussian kernel. Based on 1000 bootstrap samples, the estimated p-values for T_1, T_2 and T_3 are 0.009, 0.014 and 0.000 respectively. The p-values of all the tests provide strong evidence of rejection of the Heston model. The p-value of the distribution-based test is smaller than the p-values of the other two tests. From the Monte Carlo evidence, under the misspecification of the diffusion function, the distribution-based test seems to be slightly more powerful than the density-based tests. This give us the hint that the misspecification of the current model may come from the diffusion function. The results of this empirical application provides evidences that the model may not provide a good approximation of the marginal distribution of empirical data, although it is widely used in pricing options in real world applications.

7 Discussion and conclusion

We propose three tests for stochastic volatility model specification by comparing the parametrically and nonparametrically estimated stationary marginal density functions and distribution functions. Our approach can be adapted to discrete-time stochastic volatility models easily, as long as the volatility process is stationary. For example ,consider the classical discrete-time stochastic volatility model

$$\begin{aligned} y_{t_i} &= \sigma_{t_i}\varepsilon_{t_i}, \\ \log\sigma_{t_i}^2 &= \omega + \gamma\log\sigma_{t_{i-1}}^2 + \sigma_\eta\eta_t, \\ (\varepsilon_t, \eta_t) &\sim \mathrm{i.\,i.\,d.}\, N(0, I_2). \end{aligned}$$

When $|\gamma| < 1$, and the volatility process $\log\sigma_t^2$ is initiated from the stationary distribution $N(\omega/(1-\gamma), \sigma_\eta^2/(1-\gamma^2))$, the volatility process is strictly stationary. It is also β-mixing with exponentially decaying coefficients, see Pham and Tran (1985), and thus ergodic. Then the conditions for nonparametric estimation and parametric estimation are satisfied, and one can compare the estimated density functions and distribution functions as discussed for continuous-time model analogously.

As discussed in the introduction section, the stationary marginal return distribution does not contain the information of dynamics in the data, such that tests defined on the marginal distribution will not be able to detect the misspecification in the dynamic structure of a model. To exploit the dependence structure in the model, we could consider to the one-step conditional distribution function and density function of $X_i|X_{i-1}, i = 2, \ldots, n$, to formulate the test statistics.

Denote the density function of $X_i|X_{i-1}$ by $p(y, x)$ and the corresponding conditional distribution function by $P(y, x)$.

For the nonparametric estimation of the two functions, we can proceed as follows. A simple kernel type density estimator for $p(y, x)$ is

$$\hat{p}(y, x) = \frac{\frac{1}{nh_n^2} \sum_{i=1}^{n-1} K\left(\frac{x-X_i}{h_n}\right) K\left(\frac{y-X_{i+1}}{h_n}\right)}{\frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)},$$

and the estimator for the conditional distribution function $P(y, x)$ is

$$\hat{P}(y, x) = \frac{\sum_{i=1}^{n-1} I(X_i \leqslant x) I(X_{i+1} \leqslant y)}{\sum_{i=1}^{n} I(X_i \leqslant x)}.$$

For parametric estimation of the above functions, we again need approximations and we can again resort to simulation based approximation.

Analogously to the univariate density based test, conditional density based test statistics can be formulated as:

$$T_4 = \int_{\mathbb{R}^2} \left(\hat{p}(x, y) - \mathbb{K}_h * p(x, y; \hat{\theta}_n)\right)^2 \mathrm{d}x\mathrm{d}y,$$

where $\mathbb{K}_h(x, y) = K(x/h) \times K(y/h)/h^2$ is the two dimensional kernel used in the definition of $\hat{p}(x, y)$. And the conditional distribution function based test is

$$T_5 = n \int_{\mathbb{R}^2} \left(P(x, y; \hat{\theta}_n) - \hat{P}(x, y)\right)^2 \mathrm{d}P(x, y; \hat{\theta}_n).$$

A similar parametric bootstrap is used to obtained the null distribution of the tests.

Appendix A: Basic setup and probabilistic properties

(N1) (kernel function) The kernel function K is a bounded, symmetric, nonnegative function on $\mathbb{R}$, satisfying

$$\int_{-\infty}^{\infty} K(x)\mathrm{d}x = 1, \quad \int_{-\infty}^{\infty} xK(x)\mathrm{d}x = 0, \quad \int_{-\infty}^{\infty} x^2 K(x)\mathrm{d}x = 2k < \infty,$$

where $k > 0$ is a constant, and

$$\int_{-\infty}^{\infty} K^2(x)\mathrm{d}x < \infty.$$

(N2) (density function) $q(x)$ and its second order derivative are bounded and uniformly continuous on $\mathbb{R}$.

The above assumptions on the kernel function, and the smoothness assumption on the density function are not the weakest possible. However, Assumptions (N1) and (N2) are sufficient for the present purpose and simplify the argument in the proof.

(N3) For the process $(X_i)_{i=1}^n$, all four dimensional joint densities of $(y_{i_1}, \ldots, y_{i_1})$ exist, are bounded and Lipschitz continuous. This implies that the corresponding distribution functions satisfy the same conditions.

(N4) As $n \to \infty$, $h_n \to 0$ and $nh_n \to \infty$.

These set of conditions will be used to derive the asymptotic properties of the test statistics, but together with the mixing conditions we assume throughout, they are also sufficient to make $\hat{q}(x)$ a (pointwise) consistent estimator of $q(x)$ for all $x \in \mathbb{R}$. (N3) is stronger than necessary for consistency, but will be required for the asymptotic distribution of the test based on the bivariate distribution of $(X_i, X_{i+1})_{i=1}^{n-1}$.

The tests developed in this paper require the observed return sequence to be stationary, ergodic and β-mixing with exponentially decaying coefficients. In the nonparametric model, it is sufficient to assume the observed return sequence $(X_i)_{i=1}^n$ to satisfy the above conditions directly. However, in the parametric model, it is non-trivial to check that these conditions are satisfied for particular choices of the functions $b(x;\theta)$ and $a(x;\theta)$. In the parametric stochastic volatility model (1), we first assume

(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).

The following are standard assumptions from Genon-Catalot et al. (1998).

(SV1) For all $\theta \in \Theta$, the function $b(x) = b(x;\theta)$ is continuous on $(0, +\infty)$, and the function $a(x) = a(x;\theta)$ is continuously differentiable on $(0, +\infty)$, such that

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

and

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

This assumption ensures the existence and uniqueness of an almost surely positive strong solution to the stochastic differential equation (1) generating the volatility process.

Define, for $v_0 > 0$, the *scale measure*

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

and the *speed measure*

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

Then the assumption

(SV2)

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

where the $\cdot$ in the integral means a arbitrary point in the domain of $s(x;\theta)$, ensures a unique and positive recurrent solution on $(0,\infty)$, see Genon-Catalot et al. (1998).

The last condition in (SV2) guarantees the existence of a stationary distribution (for the volatility process), with density defined as

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

If the process is initiated from this stationary distribution, i.e., under assumption

(SV3) The initial random variable σ_0^2 has density $\pi(x;\theta)$,

the solution is strictly stationary and ergodic.

Now we give sufficient conditions to ensure that the volatility process is β-mixing with exponentially decaying coefficients. From Theorem 3.6 of Chen et al. (2010), a sufficient condition (together with (SV1) and (SV2)) for exponential decay of the β-mixing coefficients is that the process is ρ-mixing, so in the following we give the conditions for the process to be ρ-mixing. Also, we note the result that if a diffusion process is ρ-mixing, its ρ-mixing coefficients decay at an exponential rate (Bradley (2005), Theorem 3.3, or Genon-Catalot, Jeantheau, and Laredo (2000), Proposition 2.5). Furthermore, β-mixing and ρ-mixing with exponential decay are almost equivalent concepts for scalar diffusions, as discussed in Chen et al. (2010).

(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) Let

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

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

Appendix B: Lemmas and proofs

Conditions in Appendix A ensure strict stationarity, ergodicity and β-mixing of the volatility process. Notice that the return sequence $(X_i)_{i=1}^n$ is a sequence of stochastic integrals of the volatility process with respect to an independent Brownian motion B over small fixed intervals. By the following lemma from Zu (2015), the return series inherit the stationarity, ergodicity and the β-mixing properties from the volatility process.

Lemma 1 *In the model (1), if the volatility process $(\sigma_t^2)_{t\geq 0}$ is stationary, ergodic and β-mixing with a certain decay rate, then the normalized return sequence $(X_i)_{i=1}^n$ is also stationary, ergodic and β-mixing, with a mixing decay rate at least as fast as that of $(\sigma_t^2)_{t\geq 0}$.*

In all the proofs in this appendix, we take the above mentioned probabilistic properties for the return series as given to avoid repetition. When we use a integral without the range of integration, the integration is over the full real axis $\mathbb{R}$.

Proof (of Theorem 1) We first derive the asymptotic distribution of T_1. Notice that

$$\begin{aligned}
& T_1 \\
= \; & nh^{1/2}\int (\hat{q}(x) - K_h * q(x;\hat{\theta}_n))^2 \mathrm{d}x \\
= \; & nh^{1/2}\int (\hat{q}(x) - K_h * q(x))^2 \mathrm{d}x + nh^{1/2}\int (K_h * q(x) - K_h * q(x;\hat{\theta}_n))^2 \mathrm{d}x \\
& + nh^{1/2}\int (\hat{q}(x) - K_h * q(x))(K_h * q(x) - K_h * q(x;\hat{\theta}_n)) \mathrm{d}x.
\end{aligned}$$

Define

$$T_1' = nh^{1/2}\int (\hat{q}(x) - K_h * q(x))^2 \mathrm{d}x.$$

It will be shown later that $T_1' = O_p(1)$; the second term $nh^{1/2}\int (K_h * q(x) - K_h * q(x;\hat{\theta}_n))^2 \mathrm{d}x = O_p(h^{1/2})$ because $\hat{\theta}_n$ is a $\sqrt{n}$-consistent estimator for θ_0 and the square integrable assumption on the Lipschitz constant $L(x)$ as in Assumption (P1a), so this term is dominated by T_1'; the crossproduct term is clearly dominated by T_1' by the Cauchy-Schuwarz inequality. Thus we have that

$$T_1 = T_1'(1 + o_p(1)),$$

and we derive the asymptotic distribution of T_1' in the following.

First notice

$$
\begin{aligned}
& \int \left(\frac{1}{n} \sum_{i=1}^{n} (K_h(x - X_i) - K_h * q(x)) \right)^2 \mathrm{d}x \\
= \; & \frac{1}{n^2} \sum_{i=1}^{n} \int (K_h(x - X_i) - K_h * q(x))^2 \mathrm{d}x \\
& + \frac{2}{n^2} \sum_{i<j} \int (K_h(x - X_i) - K_h * q(x))(K_h(x - X_j) - K_h * q(x)) \mathrm{d}x \\
:= \; & \frac{1}{n^2} \sum_{i=1}^{n} \varphi_n(X_i, X_i) + \frac{2}{n^2} \sum_{i<j} \varphi_n(X_i, X_j),
\end{aligned}
$$

where

$$\varphi_n(u, v) := \int (K_h(x - u) - K_h * q(x))(K_h(x - v) - K_h * q(x)) \mathrm{d}x.$$

Next, we show that

1. The sum of the diagonal terms

$$\frac{1}{n^2} \sum_{i=1}^{n} \varphi_n(X_i, X_i) \xrightarrow{p} (nh)^{-1} \int K^2(u) \mathrm{d}u.$$

2. The sum of the off-diagonal terms

$$nh^{1/2} \left(\frac{2}{n^2} \sum_{i<j} \varphi_n(X_i, X_j) \right) \xrightarrow{d} N \left(0, 2 \int q_0^2(u) \mathrm{d}u \int (K^{(2)}(u))^2 \mathrm{d}u \right).$$

3. Then we show that

$$T_1' \xrightarrow{d} N \left(0, 2 \int q_0^2(u) \mathrm{d}u \int (K^{(2)}(u))^2 \mathrm{d}u \right),$$

and the asymptotic distribution of T_1 follows easily.

Step 1 We first compute the order of the mean,

$$
\begin{aligned}
& E\left(\frac{1}{n^2}\sum_{i=1}^{n}\varphi_n(X_i,X_i)\right) \\
= \; & \frac{1}{n}E\varphi_n(X_1,X_1) \\
= \; & \frac{1}{n}\int\int (K_h(x-X_1)-K_h*q(x))^2 \mathrm{d}x q(X_1)\mathrm{d}X_1 \\
= \; & \frac{1}{nh^2}\int\int\left(K\left(\frac{x-X_1}{h}\right)\right)^2 q(X_1)\mathrm{d}x\mathrm{d}X_1(1+o(1)) \\
= \; & \frac{1}{nh}\int\int (K(u))^2 q(X_1)\mathrm{d}u\mathrm{d}X_1(1+o(1)) \\
= \; & (nh)^{-1}\int K^2(u)\mathrm{d}u(1+o(1)). \qquad (9)
\end{aligned}
$$

Then we compute the order of the variance

$$
\begin{aligned}
& \mathrm{Var}\left(\frac{1}{n^2}\sum_{i=1}^{n}\varphi_n(X_i,X_i)\right) \\
\leqslant \; & E\left(\frac{1}{n^2}\sum_{i=1}^{n}\varphi_n(X_i,X_i)\right)^2 \\
= \; & \frac{1}{n^3}E\varphi_n^2(X_1,X_1)+\frac{2}{n^4}\sum_{i<j}E\varphi_n(X_i,X_i)\varphi_n(X_j,X_j). \qquad (10)
\end{aligned}
$$

We look at the two terms separately. For the first term

$$
\begin{aligned}
& E\varphi_n^2(X_1,X_1) \\
= \; & \int\left(\int K_h^2(x-X_1)\mathrm{d}x\right)^2 q(X_1)\mathrm{d}X_1(1+o(1)) \\
= \; & \frac{1}{h^4}\int\left(\int K^2\left(\frac{x-X_1}{h}\right)\mathrm{d}x\right)^2 q(X_1)\mathrm{d}X_1(1+o(1)) \\
= \; & \frac{1}{h^2}\int\left(\int K^2(u)\mathrm{d}u\right)^2 q(X_1)\mathrm{d}X_1(1+o(1)) \\
= \; & O\left(\frac{1}{h^2}\right). \qquad (11)
\end{aligned}
$$

For the second term,

$$
\begin{aligned}
& E\varphi_n(X_i,X_i)\varphi_n(X_j,X_j) \\
= & \int\int\int K_h^2(x-X_i)\mathrm{d}x \int K_h^2(x-X_j)\mathrm{d}x q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\
= & \frac{1}{h^2}\int\int \left(\int K^2(u)\mathrm{d}u\right)^2 q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\
= & O\left(\frac{1}{h^2}\right). \qquad (12)
\end{aligned}
$$

Use the result in (11) and (12) in (10), we get

$$
\begin{aligned}
& \mathrm{Var}\left(\frac{1}{n^2}\sum_{i=1}^n \varphi_n(X_i,X_i)\right) \\
\leqslant & \frac{1}{n^3}O\left(\frac{1}{h^2}\right) + \frac{2}{n^4}n^2 O\left(\frac{1}{h^2}\right) \\
= & O\left(\frac{1}{n^2h^2}\right). \qquad (13)
\end{aligned}
$$

Use the results in (9) and (13) and apply Markov's inequality we have

$$
\begin{aligned}
& P\left(\left|\frac{1}{n^2}\sum_{i=1}^n \varphi_n(X_i,X_i) - (nh)^{-1}\int K^2(u)\mathrm{d}u\right| > \varepsilon\right) \\
\leqslant & \frac{E\left|\frac{1}{n^2}\sum_{i=1}^n \varphi_n(X_i,X_i) - (nh)^{-1}\int K^2(u)\mathrm{d}u\right|^2}{\varepsilon^2} \\
= & O\left(\frac{1}{n^2h^2}\right) = o(1),
\end{aligned}
$$

because $nh \to \infty$. Thus we have proved that

$$
\frac{1}{n^2}\sum_{i=1}^n \varphi_n(X_i,X_i) - (nh)^{-1}\int K^2(u)\mathrm{d}u \quad = \quad o_p(1).
$$

Step 2 Now we use Theorem A, Appendix 1 in Hjellvik, Yao, and Tjøstheim (1998) to show

$$
nh^{1/2}\left(\frac{2}{n^2}\sum_{i<j}\varphi_n(X_i,X_j)\right) \xrightarrow{d} N\left(0, 2\int q_0^2(u)\mathrm{d}u \int (K^{(2)}(u))^2\mathrm{d}u\right).
$$

Notice that $\varphi_n(x,y)$ is a degenerate symmetric kernel, and the mixing condition is satisfied.

First we calculate the asymptotic variance. Let $\tilde{X}_i$, $\tilde{X}_j$ be independent variables with the same distribution as X_i. First we compute

$$\begin{aligned}
& E\varphi_n^2(\tilde{X}_i, \tilde{X}_j) \\
= & \int\int \left(\int K_h(x-X_i)K_h(x-X_j)\mathrm{d}x\right)^2 q(X_i)q(X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\
= & \frac{1}{h^2}\int\int \left(\int K(u)K\left(u+\frac{X_i-X_j}{h}\right)\mathrm{d}x\right)^2 q(X_i)q(X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\
= & \frac{1}{h^2}\int\int \left(K^{(2)}\left(\frac{X_i-X_j}{h}\right)\right)^2 q(X_i)q(X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\
= & \frac{1}{h}\int\int (K^{(2)}(u))^2 q(X_j+uh)q(X_j)\mathrm{d}u\mathrm{d}X_j(1+o(1)) \\
= & \frac{1}{h}\int (K^{(2)}(u))^2\mathrm{d}u \int q^2(x)\mathrm{d}x(1+o(1)).
\end{aligned}$$

then the asymptotic variance

$$\sigma_n^2 = \frac{n^2}{2}E\varphi_n^2(\tilde{X}_i, \tilde{X}_j) = \frac{n^2}{2h}\int (K^{(2)}(u))^2\mathrm{d}u \int q^2(x)\mathrm{d}x(1+o(1)). \qquad (14)$$

Then we check the conditions related to the 6 quantities M_{ni}, $i = 1, \ldots, 6$, as defined in Theorem A, Appendix 1 in Hjellvik, Yao, and Tjøstheim (1998). For M_{n1}, notice that for $1 > \delta > 0$,

$$\begin{aligned}
& E|\varphi_n(X_1, X_j)\varphi_n(X_i, X_j)|^{1+\delta} \\
= & \int\int\int \left|\int K_h(x-X_1)K_h(x-X_j)\mathrm{d}x \int K_h(x-X_i)K_h(x-X_j)\mathrm{d}x\right|^{1+\delta} \\
& q(X_1, X_i, X_j)\mathrm{d}X_1\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\
= & \int\int\int \left|\frac{1}{h^4}\int K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_j}{h}\right)\mathrm{d}x \int K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_j}{h}\right)\mathrm{d}x\right|^{1+\delta} \\
& q(X_1, X_i, X_j)\mathrm{d}X_1\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\
= & \int\int\int \left|\frac{1}{h^2}K^{(2)}\left(\frac{X_1-X_j}{h}\right)K^{(2)}\left(\frac{X_i-X_i}{h}\right)\right|^{1+\delta} q(X_1, X_i, X_j)\mathrm{d}X_1\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\
= & h^2\int\int\int \left|\frac{1}{h^2}K^{(2)}(u)K^{(2)}(v)\right|^{1+\delta} q(X_j+uh, X_j+vh, X_j)\mathrm{d}u\mathrm{d}v\mathrm{d}X_j(1+o(1)) \\
= & \frac{1}{h^{2\delta}}\left(\int |K^{(2)}(u)|^{1+\delta}\mathrm{d}u\right)^2 \int q(X_j+uh, X_j+vh, X_j)\mathrm{d}X_j(1+o(1)) \\
= & O\left(\frac{1}{h^{2\delta}}\right).
\end{aligned}$$

Using the same strategy it can be shown that M_{n1} also has this upper bound and we thus have

$$n^2 M_{n1}^{\frac{1}{1+\delta}}/\sigma_n^2 = O\left(\frac{h}{h^{2\delta/(1+\delta)}}\right) = O\left(h^{\frac{1-\delta}{1+\delta}}\right) = o(1).$$

Similarly, we can show that

$$\begin{aligned} E|\varphi_n(X_1,X_j)\varphi_n(X_i,X_j)|^{2(1+\delta)} &= O\left(\frac{h^2}{h^{4(1+\delta)}}\right) = O\left(\frac{1}{h^{4\delta+2}}\right), \\ E|\varphi_n(X_1,X_j)\varphi_n(X_i,X_j)|^{2} &= O\left(\frac{1}{h^2}\right), \\ E|\varphi_n(X_1,X_i)\varphi_n(X_j,Y_k)|^{2(1+\delta)} &= O\left(\frac{1}{h^{4\delta+2}}\right), \end{aligned}$$

which further imply that

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

$$n^{\frac{3}{2}} M_{n3}^{\frac{1}{2}}/\sigma_n^2 = O\left(\frac{1}{n^{1/2}}\right) = o(1).$$

$$n^{\frac{3}{2}} M_{n4}^{\frac{1}{2(1+\delta)}}/\sigma_n^2 = O\left(\frac{1}{n^{1/2}h^{\delta/(1+\delta)}}\right) = o(1),$$

by noticing again that $\delta/(1+\delta) < 1/2$ and $nh \to \infty$.

For M_{n5}, we first calculate

$$\begin{aligned} & E\left|\int \varphi_n(X_1,X_i)\varphi_n(X_1,X_j)\mathrm{d}P(X_1)\right|^{2(1+\delta)} \\ = & \int\int\left|\int\left(\int K_h(x-X_1)K_h(x-X_i)\mathrm{d}x\int K_h(x-X_1)K_h(x-X_j)\mathrm{d}x\right)q(X_1)\mathrm{d}X_1\right|^{2(1+\delta)} \\ & q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\ = & \int\int\left|\int\frac{1}{h^4}\left(\int K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_i}{h}\right)\mathrm{d}x\int K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_j}{h}\right)\mathrm{d}x\right)q(X_1)\mathrm{d}X_1\right|^{2(1+\delta)} \\ & q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\ = & \int\int\left|\int\frac{1}{h^2}K^{(2)}\left(\frac{X_1-X_i}{h}\right)K^{(2)}\left(\frac{X_1-X_j}{h}\right)q(X_1)\mathrm{d}X_1\right|^{2(1+\delta)} q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\ = & \int\int\left|\int\frac{1}{h}K^{(2)}(u)K^{(2)}\left(u+\frac{X_i-X_j}{h}\right)q(X_i+uh)\mathrm{d}u\right|^{2(1+\delta)} q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\ \leqslant & \frac{1}{h^{2(1+\delta)}}\int\int\int\left|K^{(2)}(u)K^{(2)}\left(u+\frac{X_i-X_j}{h}\right)\right|^{2(1+\delta)} q(X_i+uh)\mathrm{d}u\,q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1)) \\ = & \frac{h}{h^{2(1+\delta)}}\int\int\int|K^{(2)}(u)K^{(2)}(u+v)|^{2(1+\delta)}q(X_i+uh)\mathrm{d}u\,q(X_i,X_i-vh)\mathrm{d}X_i\mathrm{d}v(1+o(1)) \\ \leqslant & C\times\frac{h}{h^{2(1+\delta)}}\left(\int|K^{(2)}(u)|^{2(1+\delta)}\mathrm{d}u\right)^2 \\ = & O\left(\frac{1}{h^{2\delta+1}}\right). \end{aligned}$$

Using the same method, we can show that the other quantities in the definition of M_{n5}

also have this upper bound and $M_{n5} = O\left(\frac{1}{h^{2\delta+1}}\right)$. We thus have

$$n^2 M_{n5}^{\frac{1}{2(1+\delta)}}/\sigma_n^2 = O\left(\frac{h}{h^{\frac{2\delta+1}{2(1+\delta)}}}\right) = O\left(h^{\frac{1}{2(1+\delta)}}\right) = o(1).$$

Similarly we have

$$E\left|\int \varphi_n(X_1, X_i)\varphi_n(X_1, X_j)\mathrm{d}P(X_1)\right|^2 = O\left(\frac{1}{h}\right),$$

and

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

We have then checked the condition

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

and the CLT for the U-statistic is proved and we have finished proving Step 2.

Step 3 The asymptotic distribution of T_1' is an easy consequence of Step 1 and 2. The asymptotic distribution of T_1 is obtained by noticing $T_1 = T_1'(1 + o_p(1))$.

For the asymptotic distribution of T_2. First use the result in Fan and Ullah (1999) Theorem 4.1 we have

$$\frac{1}{n}\sum_{i=1}^{n}\hat{q}(X_i) \xrightarrow{p} \int q^2(x)\mathrm{d}x,$$

then the CLT for T_2 follows easily from Slutsky's lemma. □

In proving Theorem 2, we will need the following weak convergence results for empirical process of β-mixing sequences:

Lemma 2 *(Kosorok (2008), Theorem 11.24) Let $X_1, X_2, \ldots$ be stationary with marginal distribution P, and β-mixing with*

$$\sum_{k=1}^{\infty} k^{2/(p-2)}\beta(k) < \infty,$$

for some $2 < p < \infty$. Let $\mathcal{F}$ be a class of functions in $L_2(P)$ satisfying the entropy condition, where the bracketing number satisfies

$$J_{[\,]}(\infty, \mathcal{F}, L_p(P)) < \infty. \tag{15}$$

then

$$\mathbb{G}_n f = n^{1/2}\sum_{i=1}^{n}(f(X_i) - Pf) \xrightarrow{d} \mathbb{G}_P f,$$

in $l^\infty(\mathcal{F})$, where $f \mapsto \mathbb{G}_P f$ is a tight, mean zero Gaussian process with covariance function

$$\Gamma(f,g) = \lim_{k\to\infty} \sum_{i=1}^{\infty} \mathrm{Cov}(f(X_k), g(X_i)),$$

for all $f, g \in \mathcal{F}$.

Proof (of Theorem 2) We prove the theorem using empirical processes techniques. We are working with dependent data, so we need the empirical process result for β-mixing sequences in Lemma 2, which is Theorem 11.24 in Kosorok (2008). The exponential decay of β-mixing coefficients is sufficient for the above lemma to work.

We first prove

$$\sqrt{n}(\widehat{Q}(x) - Q(x;\hat{\theta}_n)) \rightsquigarrow^{l^\infty(\mathcal{F})} x \mapsto \mathbb{G}_Q I(u \leqslant x) - \mathbb{G}_Q \psi_{\theta_0}^T \frac{\partial Q(x,\theta)}{\partial \theta}|_{\theta=\theta_0}, \tag{16}$$

using the strategy discussed in Van der Vaart (2000), pp. 278–279. Here the notation $\rightsquigarrow^{l^\infty}$ denote weak convergence of stochastic process. Notice that

$$\begin{aligned}\sqrt{n}(\widehat{Q}(x) - Q(x;\hat{\theta}_n)) &= \sqrt{n}(\widehat{Q}(x) - Q(x;\theta)) - \sqrt{n}(Q(x;\hat{\theta}_n) - Q(x;\theta)) \\ &= \sqrt{n}(\widehat{Q}(x) - Q(x;\theta)) - \sqrt{n}(\hat{\theta}_n - \theta)\frac{\partial Q(x,\theta)}{\partial\theta}, \\ &= \sqrt{n}(\widehat{Q}(x) - Q(x;\theta)) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\theta(X_i)\frac{\partial Q(x,\theta)}{\partial\theta} + o_p(1),\end{aligned}$$

where we use the differentiability of the parameterization and the assumption (P1b) about the expansion of the parametric estimator. With this, the above limiting distribution is determined by the joint distribution of

$$\left(\sqrt{n}(\widehat{Q}(x) - Q(x;\theta)), \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\theta_0}(X_i)\right).$$

Notice that our $\mathcal{F}$ is the class of indicator functions $\mathcal{F} = \{I(-\infty, x)\}$, which satisfies the entropy condition (15) and thus is a Donsker class. Adding the k components of ψ_θ to $\mathcal{F}$ will make a larger class which we call $\mathcal{G}$, which is again Donsker (a finite class is Donsker); this is because the union of Donsker classes is also Donsker. Therefore

$$\left(\sqrt{n}(\widehat{Q}(x) - Q(x;\theta)), \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\theta_0}(X_i)\right) \rightsquigarrow^{l^\infty(\mathcal{G})} g \mapsto \mathbb{G}_Q g,$$

and using the continuous mapping theorem we get

$$\sqrt{n}(\widehat{Q}(x) - Q(x;\theta)) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\theta_0}(X_i)\frac{\partial Q(x,\theta)}{\partial\theta} \rightsquigarrow^{l^\infty} x \mapsto \mathbb{G}_Q I(u \leqslant x) - \mathbb{G}_Q \psi_\theta^T \frac{\partial Q(x,\theta)}{\partial\theta}.$$

Notice that the process $\mathbb{G}_Q I(u \leqslant x)$ and the variable $\mathbb{G}_Q \psi_\theta^T$ are dependent because they can be viewed as marginals of the process $g \mapsto \mathbb{G}_Q g$. Therefore, under the null hypothesis, (16) is proved.

Then, under the null hypothesis,

$$\begin{aligned} T_{2n} &= n \int (\widehat{Q}(x) - Q(x;\hat{\theta}_n))^2 \mathrm{d}Q(x;\hat{\theta}) \\ &= \int \left(\sqrt{n}(\widehat{Q}(x) - Q(x;\hat{\theta}_n))\right)^2 \mathrm{d}Q(x;\theta_0)(1 + O_p(n^{-1/2})) \\ &= \int \left(\sqrt{n}(\widehat{Q}(x) - Q(x;\hat{\theta}_n))\right)^2 \mathrm{d}Q(x;\theta)(1 + o_p(1)), \end{aligned}$$

and the result of the theorem follows easily from continuous mapping, because the map $z \mapsto \int z^2(t)\mathrm{d}Q(t)$ from $D[-\infty, +\infty]$ into $\mathbb{R}$ is continuous with respect to the supremum norm. □

Proof (of Theorem 3) Under the alternative hypothesis, we can make the following decomposition of the test statistic,

$$\begin{aligned} & T_1 \\ = \; & nh^{1/2} \int (\hat{q}(x) - K_h * q(x;\hat{\theta}_n))^2 \mathrm{d}x \\ = \; & nh^{1/2} \int (\hat{q}(x) - K_h * q_1(x))^2 \mathrm{d}x + nh^{1/2} \int (K_h * q_1(x) - K_h * q(x;\hat{\theta}_n))^2 \mathrm{d}x \\ & +2nh^{1/2} \int (\hat{q}(x) - K_h * q_1(x))(K_h * q_1(x) - K_h * q(x;\hat{\theta}_n)) \mathrm{d}x \end{aligned}$$

Using the same approach as in Theorem 1, it can be shown that the first term satisfies

$$\begin{aligned} & nh^{1/2} \left(\int (\hat{q}(x) - K_h * q_1(x))^2 \mathrm{d}x - (nh)^{-1} \int K^2(u)\mathrm{d}u \right) \\ \xrightarrow{d} \; & N\left(0, 2 \int q_0^2(u)\mathrm{d}u \int (K * K)^2(u)\mathrm{d}u \right). \end{aligned}$$

For the second term, by definition

$$\int (K_h * q_1(x) - K_h * q(x;\hat{\theta}_n))^2 \mathrm{d}x \xrightarrow{p} \int (q_1(x) - q(x;\hat{\theta}_n))^2 \mathrm{d}x = O_p(1),$$

as $h \to 0$, because this is the L_2 distance between the alternative model and the pesudotrue model. The limit exists because we are considering functions in the L_2 space. Thus we have

$$nh^{1/2} \int (K_h * q_1(x) - K_h * q(x;\hat{\theta}_n))^2 \mathrm{d}x \to \infty.$$

For the third term, by Cauchy-Schuwarz inequality

$$\begin{aligned}
&\int (\hat{q}(x) - K_h * q_1(x))(K_h * q_1(x) - K_h * q(x;\hat{\theta}_n))\mathrm{d}x \\
\leqslant\ & \left(\int (\hat{q}(x) - K_h * q_1(x))^2 \mathrm{d}x\right)^{1/2} \left(\int (K_h * q_1(x) - K_h * q(x;\hat{\theta}_n))^2 \mathrm{d}x\right)^{1/2} \\
=\ & O_p(n^{-1/2}h^{-1/4}).
\end{aligned}$$

Then it is obvious that $T_1 \to \infty$ under $\mathcal{H}_1$ and the test is consistent.

The consistency of the test T_2 can be shown analogously. □

Proof (of Theorem 4) Let $F(x) = Q(x, \theta^*)$ be the projection of $Q_1(x)$ onto the space of parametric models. That is, $Q(x, \theta^*)$ is the pseudotrue model. Let $X_1, X_2, \ldots$ be the observations generated from $Q(x, \theta^*)$, and denote $\hat{F}(x)$ be the empirical distribution function of the sample $\{X_i\}_{i=1}^n$. Then under the alternative hypothesis, we can make the following decomposition

$$\begin{aligned}
& T_3 \\
=\ & n\int (\hat{Q}(x) - Q(x;\hat{\theta}_n))^2 \mathrm{d}Q(x;\hat{\theta}_n) \\
=\ & n\int (\hat{Q}(x) - \hat{F}(x))^2 \mathrm{d}Q(x;\hat{\theta}_n) + n\int (\hat{F}(x) - Q(x;\hat{\theta}_n))^2 \mathrm{d}Q(x;\hat{\theta}_n) \\
& +2n\int (\hat{Q}(x) - \hat{F}(x))(\hat{F}(x) - Q(x;\hat{\theta}_n))\mathrm{d}\mathrm{d}Q(x;\hat{\theta}_n).
\end{aligned}$$

When $n \to \infty$, the first term

$$n\int (\hat{Q}(x) - \hat{F}(x))^2 \mathrm{d}Q(x;\hat{\theta}_n) \xrightarrow{p} n\int (Q_1(x) - F(x))^2 \mathrm{d}F(x) = O_p(n).$$

Using the same method as in the proof of Theorem 2, it can be shown that the second term $n\int (\hat{F}(x) - Q(x;\hat{\theta}_n))^2 \mathrm{d}Q(x;\hat{\theta}_n)$ satisfy the same convergence in distribution result as in that theorem, such that

$$n\int (\hat{F}(x) - Q(x;\hat{\theta}_n))^2 \mathrm{d}Q(x;\hat{\theta}_n) = O_p(1).$$

By Cauchy-Schuwarz inequality the cross product term is $O_p(n^{1/2})$. Then it is obvious that $T_3 \to \infty$ under $\mathcal{H}_1$ when $n \to \infty$ and the test is consistent. □

References

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies 9*, 385–426.

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., L. P. Hansen, and J. A. Scheinkman (2010). Operator methods for continuous-time Markov processes. In Y. Aït-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*, Volume 1, pp. 1–66. Amsterdam: North Holland.

Aït-Sahalia, Y. and R. Kimmel (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics 83*(2), 413–452.

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. and J. Lund (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics 77*(2), 343–377.

Andrews, D. W. (1997). A conditional Kolmogorov test. *Econometrica: Journal of the Econometric Society*, 1097–1128.

Andrews, D. W. (2005). Higher-order improvements of the parametric bootstrap for markov processes. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, Cambridge*, 171–215.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica 76*, 1481–1536.

Bhardwaj, G., V. Corradi, and N. R. Swanson (2008). A simulation-based specification test for diffusion processes. *Journal of Business & Economic Statistics 26*(2), 176–193.

Bollerslev, T. and H. Zhou (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics 109*, 33–65.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys 2*, 107–144.

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.

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.

Efron, B. and R. J. Tibshirani (1994). *An introduction to the bootstrap*, Volume 57. CRC press.

Escanciano, J. C. (2009). On the lack of power of omnibus specification tests. *Econometric Theory 25*(01), 162–194.

Eubank, R. and V. LaRiccia (1992). Asymptotic comparison of cramer-von mises and nonparametric function estimation techniques for testing goodness-of-fit. *The Annals of Statistics*, 2071–2086.

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 A. Ullah (1999). On goodness-of-fit tests for weakly dependent processes using kernel method. *Journal of Nonparametric Statistics 11*, 337–360.

Franke, J., J.-P. Kreiss, and E. Mammen (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli 8*(1), 1–37.

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.

Hjellvik, V., Q. Yao, and D. Tjøstheim (1998). Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference 68*, 295–321.

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.

Kloeden, P. E. and E. Platen (1992). *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Berlin: Springer.

Kristensen, D. (2011, October). Semi-nonparametric estimation and misspecification testing of diffusion models. *Journal of Econometrics 164*(2), 382–403.

Li, F. (2007). Testing the parametric specification of the diffusion function in a diffusion process. *Econometric Theory 23*(02), 221–250.

Meddahi, N. (2002). Moments of continuous time stochastic volatility models. *Unpublished paper: University of Montreal*.

Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics.* Cambridge: Cambridge University Press.

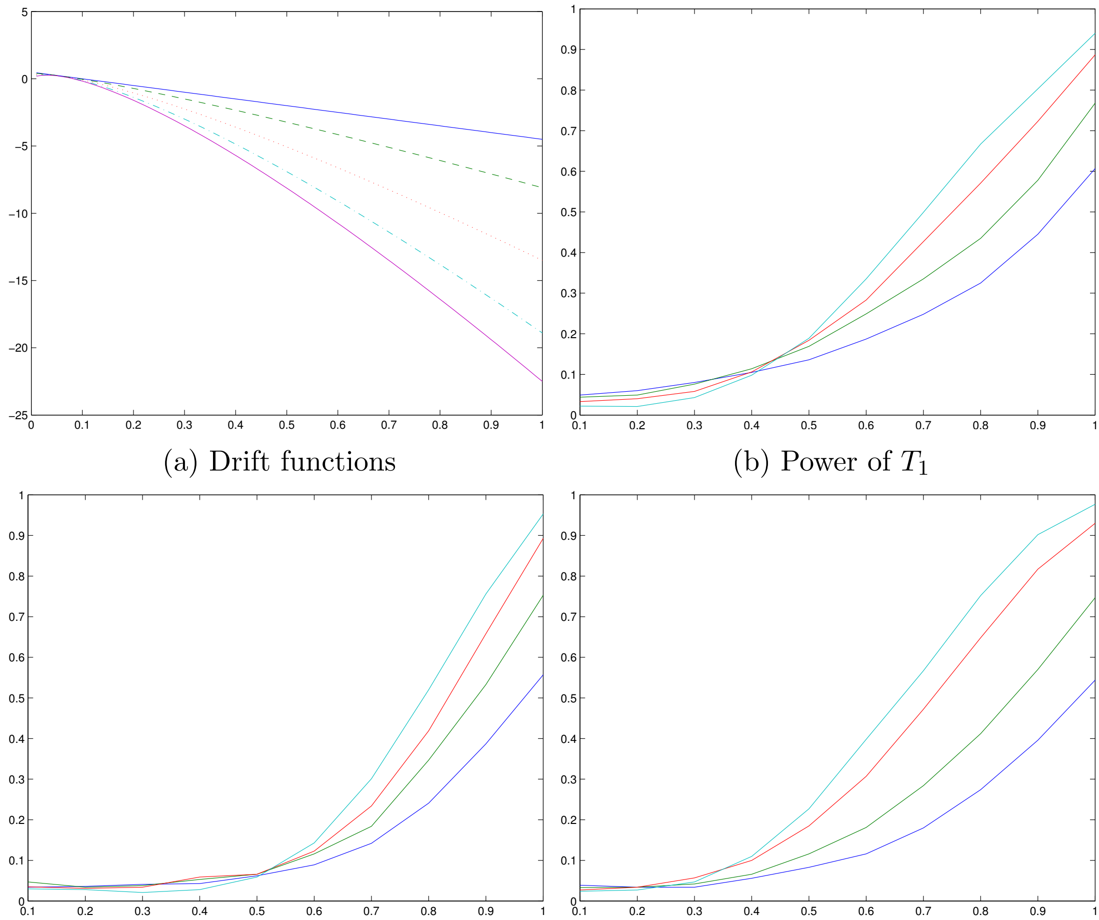
Pham, T. D. and L. T. Tran (1985). Some mixing properties of time series models. *Stochastic Processes and Their Applications 19*, 297–303.

Renault, E. (2009). Moment-based estimation of stochastic volatility models. In T. G. Andersen, R. A. Davis, and J.-P. K. T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 269–311. Berlin: Springer.

Van der Vaart, A. (2000). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

Wasserman, L. (2004). *All of statistics: a concise course in statistical inference.* Springer.

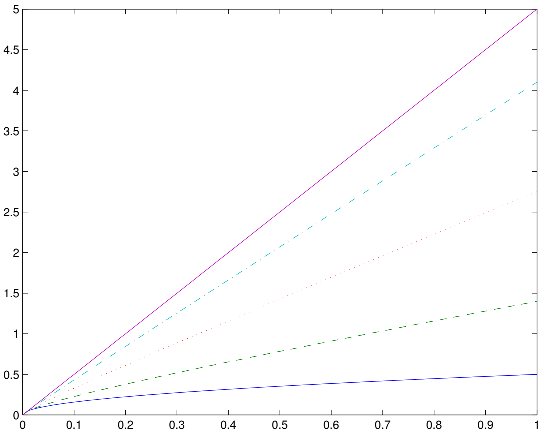
Zu, Y. (2015). Nonparametric specification tests for stochastic volatility models based on volatility density. *Journal of Econometrics, forthcoming*.



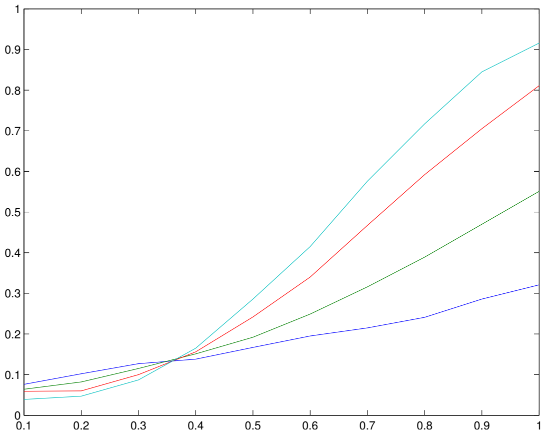
(a) Drift functions (b) Power of T_1

(c) Power of T_2 (d) Power of T_3

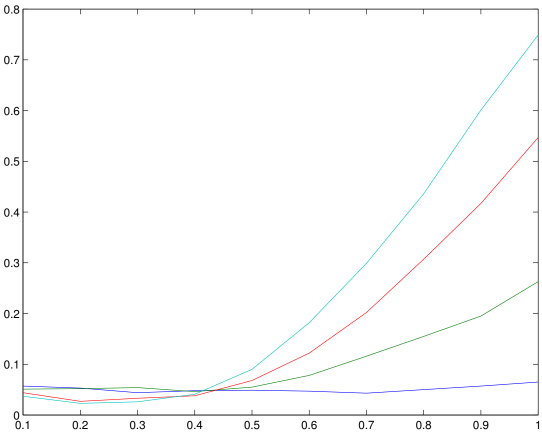
Figure 1: Power under the misspecification of the drift function. (a): drift function with $\tau = 0$ (solid), $\tau = 0.2$ (dashed),$\tau = 0.5$ (dotted),$\tau = 0.8$ (dash-dotted),$\tau = 1$ (purple solid). (b), (c), (d): $n = 500$ (blue), $n = 1000$ (green), $n = 2000$ (red), $n = 3000$ (cyan)).



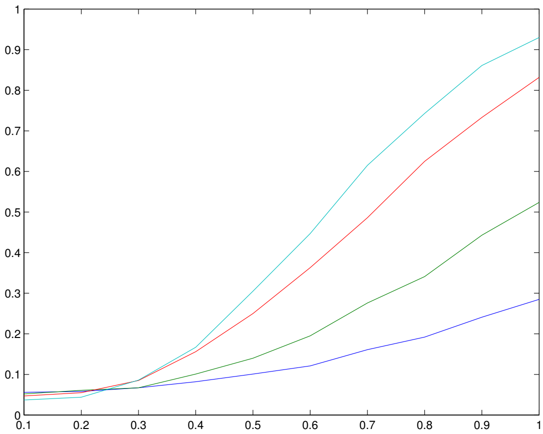
(a) Diffusion functions



(b) Power of T_1 under different sample sizes

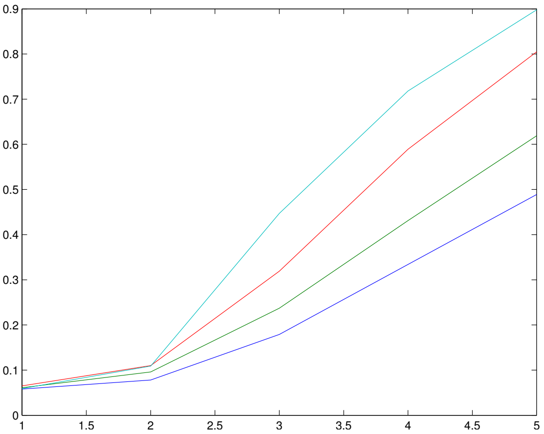


(c) Power of T_2 under different sample sizes

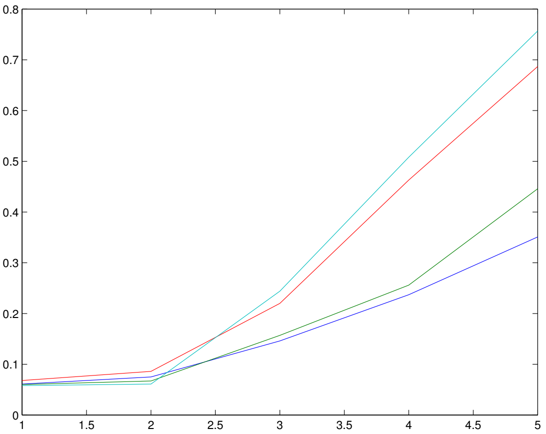


(d) Power of T_3 under different sample sizes

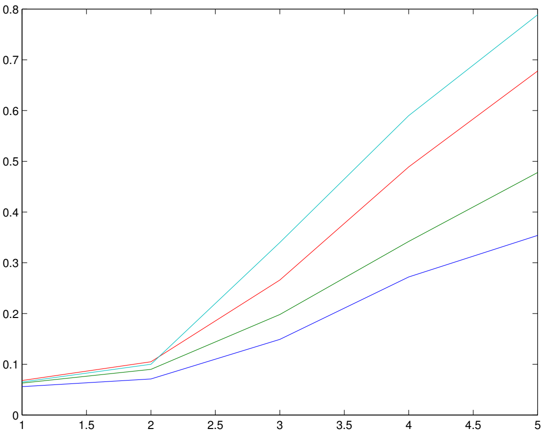
Figure 2: Power under the misspecification of the diffusion function. (a): drift function with $\tau = 0$ (solid), $\tau = 0.2$ (dashed),$\tau = 0.5$ (dotted),$\tau = 0.8$ (dash-dotted),$\tau = 1$ (purple solid). (b), (c), (d): $n = 500$ (blue), $n = 1000$ (green), $n = 2000$ (red), $n = 3000$ (cyan)).



(a) Power of T_1



(b) Power of T_2



(c) Power of T_3

Figure 3: Power under different jump intensities. (a), (b), (c): $n = 500$ (blue), $n = 1000$ (green), $n = 2000$ (red), $n = 3000$ (cyan)).

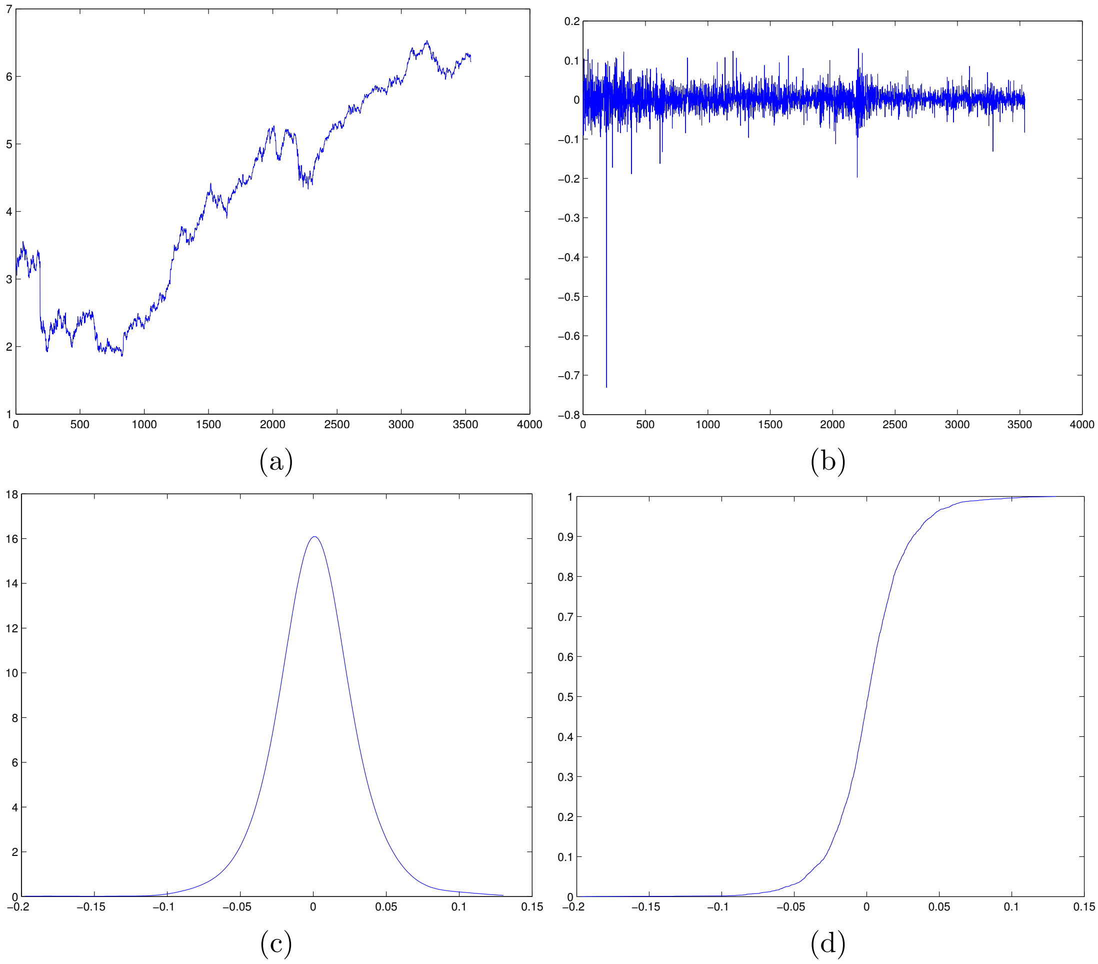


Figure 4: (a): Log prices of Apple stock from 2000 to 2014. (b): Log returns of Apple stock: from 2000 to 2014. (c): Nonparametric density estimate. (d): Empirical distribution function

	T_1	T_2	T_3
$n = 1000$	0.0277	0.0581	0.0377
$n = 2000$	0.0372	0.0363	0.0400
$n = 3000$	0.0406	0.0564	0.0411

Table 1: Size of the bootstrap tests, cross validation