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

Yang Zu[†]

School of Economics

University of Nottingham

H. Peter Boswijk[‡]

Amsterdam School of Economics

University of Amsterdam

December 31, 2016

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 proposed tests are applied in a number of empirical examples.

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. [1]

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. The extension of the methods developed in this paper to the case where the parametric model (1) is augmented with jumps and leverage effects will be discussed in Section 5.

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

$$
\mathcal{H}_0 : q(x) = q_0(x) \in \{q(x; \theta), \theta \in \Theta\},
\tag{4}
$$

where $\Theta \subseteq \mathbb{R}^k$ is the parameter space, and θ_0 is the true parameter under the null

---

[1] Jensen and Maheu (2010) consider a Bayesian semiparametric stochastic volatility model, where the distribution of return innovations $\mathrm{d}B_t$ is assumed unknown and modeled nonparametrically. Their model is different from ours, because we test a fully parametric model against a model with a non-parametric volatility process but parametric (Gaussian) innovation distribution.

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 modelling. The 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 $\{b(.; \theta), a(.; \theta)\}^2$; however, tests constructed this way would still be an important "first check" because of its empirical significance in any structural modelling. More detailed information could be obtained by defining test statistics based on the transition distribution of the observed returns. We discuss this issue in Section 8.

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); intraday variation of volatility (the so-called diurnal effect) and prominent microstructure noise effects in prices observed at ultra high frequencies (see e.g. Andersen and Bollerslev (1998)) make the model unsuitable for such data. Throughout, we would need $(X_i)_{i=1}^n$ to be a stationary and ergodic sequence, and that it is β-mixing with exponentially decaying coefficients. We do not impose these properties as high level assumptions; instead, checkable sufficient conditions for these properties to hold in the parametric model are given in Appendix A.

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 on nonparametric tests 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 (1) makes the aforementioned research not directly applicable. Corradi and Swanson (2011), henceforth CS11, consider a conditional distribution based nonparametric test for stochastic volatility models, where the authors assume the observed series to be strictly stationary. In our model we assume the observed return series

---

[2]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 $q(x)$.

(first difference of the observed series) to be stationary and we allow the observed series to exhibit unit-root type dynamics. Although CS11's method can be applied to a range of interest rate stochastic volatility models, where mean-reversion is often observed, it does not cover the well known Heston model and GARCH diffusion model, which are used widely in the option pricing literature to describe the evolution of equity prices. 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 contributions of the paper are threefold. First, although it appears that CS11's work can be adapted to the equity price stochastic volatility models easily, no formal work has been done so far (according to our knowledge); besides, our the marginal distribution based tests are by no means a direct adaptation of CS11's test and the asymptotic theory is different: CS11 use conditional empirical process techniques while we use a central limit theorem of U statistics for our density based tests and use empirical process techniques for our distribution function based test. Second, instead of imposing high level assumptions that the observed series is stationary and strong mixing as in CS11, we give checkable conditions in our paper for the necessary probabilistic properties of the model to hold. Third, it is well-known in the literature that density based tests are more sensitive to local deviations of the model while distribution function based tests are more sensitive to global type of deviations to the null model; to account for different types of deviations to the null model, we consider both the density function based test and the distribution function based test, and study their finite sample power in Monte Carlo experiments.

An alternative way of formulating a test would be to base the test statistic on the conditional distribution of the observed returns. Although one might expect that conditional distribution based tests are superior to marginal distribution based tests as considered in this paper, there is no theoretical result nor empirical evidence to support this claim. Theoretically, we know that there is no uniformly most powerful test in nonparametric testing problems (see e.g. Aït-Sahalia et al. (2009) p. 1105). Empirically, as we argued earlier in this section, the marginal distribution of the observed returns by itself is an important object of empirical modelling and its specification would need to be checked usually in the first place. Therefore we view the tests based on the marginal distribution and the conditional distribution as complementary to each other — both types of tests contain information that will shed light on our understanding of the source of (possible) misspecification of a model. Moreover, we also discuss the possible extension of our tests to exploit the dependence structure of the model in Section 8.[3]

---

[3]We focus on the model specification test under the real world measure in this paper. As suggested by an anonymous referee, we could consider extending our tests to the risk neutral measure, where we would have the advantage of an enlarged information set. In relation to this, Song and Xiu (2016) have

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. Section 4 defines the test statistics, derives their asymptotic null distributions and consistency, and discusses using the bootstrap to approximate the null distribution. Section 5 discusses the issue of testing general models and testing models estimated with Bayesian methods. In Section 6 Monte Carlo evidence for the size and power properties of the tests are given. In Section 7 we study empirical applications. Section 8 discusses possible extensions and concludes. Technical assumptions are collected in Appendix A. The proofs of the theorems are collected in Appendix B.

2   Nonparametric estimation

In this section we 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.

Let h_n be a bandwidth, and $K(.)$ 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 $(X_i)_{i=1}^n$ by $Q(x)$. 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 by application of the Ergodic Law of Large Numbers and the Central

---

made a nice graphical comparison of the nonparametric and parametric estimates of the pricing kernel under the risk neutral measure. A formal statistical analysis under the risk neutral measure would be a useful direction of future research.

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 of the parameter vector, denoted as $\hat{\theta}$, and then evaluate the two functions given $\hat{\theta}$.

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. Furthermore, 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 ourselves 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 $\hat{\theta}_n$ to be $\sqrt{n}$-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 $L(x)$ 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 the 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 $X_i = \frac{1}{\sqrt{\Delta}} \int_{t_{i-1}}^{t_i} \sigma_s dB_s, i = 1, \ldots, n$, 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 $Q^*(x;\hat{\theta})$ can be approximated analogously using the empirical distribution function with the simulated data. For this reason, and for notational convenience, in the following we treat the approximations $q^*(x;\theta)$ and $Q^*(x;\theta)$ as

equal to the corresponding $q(x; \theta)$ and $Q(x; \theta)$.

Remark 1 Euler's scheme is widely used in the literature to simulate stochastic volatility models with leverage effects. This includes Andersen and Lund (1997), Bollerslev and Zhou (2002), Aït-Sahalia and Kimmel (2007) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). When used with a stochastic volatility model, usually the Euler 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), who advocate the use of a generalized Milstein scheme to avoid this problem.

4 Test statistics and asymptotic properties

4.1 Asymptotic null distribution and consistency

Define
$$T_0 = nh^{1/2} \int_{\mathbb{R}} (\hat{q}(x) - K_h * q(x; \hat{\theta}))^2 \mathrm{d}x,$$

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 delivers better asymptotic properties of the test statistic, we refer to Fan (1994) for a discussion of this issue in the general density testing problem with i.i.d. data.

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

$$\left(T_0 - h^{-1/2} \int_{\mathbb{R}} K^2(u)\mathrm{d}u\right) \xrightarrow{d} N(0, \sigma^2),$$

where the variance of the asymptotic distribution $\sigma^2 := 2 \int_{\mathbb{R}} q_0^2(x)\mathrm{d}x \int_{\mathbb{R}} \left(K^{(2)}(v)\right)^2 \mathrm{d}v$ and $K^{(2)}(v)$ denotes 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

$$T_1 = \frac{T_0 - h^{-1/2} \int_{\mathbb{R}} K^2(u)\mathrm{d}u}{\hat{\sigma}} \xrightarrow{d} N(0, 1).$$

8

The test statistic T_0 is not asymptotically pivotal as its asymptotic null distribution depends on the unknown density $q_0(x)$, which makes an asymptotic test that rejects for large values of T_0 infeasible. This motivates the use of the corresponding studentized test T_1, which is pivotal. However, both T_0 or T_1 may be used for a bootstrap test, as analyzed in Section 4.2.

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

$$T_2 = n \int_{\mathbb{R}} (\widehat{Q}(x) - Q(x; \hat{\theta}_n))^2 \mathrm{d}Q(x; \hat{\theta}).$$

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

$$T_2 \xrightarrow{d} \int_{\mathbb{R}} \left(\mathbb{G}_Q I(\cdot \leq x) - \mathbb{G}_Q \psi_{\theta_0}^T(\cdot) \frac{\partial Q(x, \theta)}{\partial \theta} |_{\theta = \theta_0} \right)^2 \mathrm{d}Q(x), \tag{5}$$

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_2 is a functional of a Brownian bridge process, and it depends on the model structure (thus is not model-free) as well as the unknown parameter values. For this reason, this limit theorem cannot be used directly to define critical values of the test. We discuss an approximation method to obtain test critical values in 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 $\mathcal{H}_1$,

$$P\left(\frac{T_0 - h^{-1/2}\int_{\mathbb{R}} K^2(u)\mathrm{d}u}{\sigma} > Z_{1-\alpha}\right) \to 1,$$

and

$$P\left(T_1 > Z_{1-\alpha}\right) \to 1.$$

From the proof of Theorem 3, it is clear that $T_1 \to +\infty$ asymptotically. This implies that although T_1 has an asymptotic standard normal distribution, the test (still) works by rejecting large positive values of T_1 and is thus a one-sided test.

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

(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 alternative hypothesis, and if (SV0)–(SV5) and (N2) in Appendix A 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 (5), then as $n \to \infty$,*

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

As with most nonparametric tests, all the three tests are consistent. That is, they can detect any fixed deviation to the true model as long as the sample size is sufficiently large. We have already noted from above that the test T_0 and T_2 are not feasible as their asymptotic null distribution are not known. We will discuss how to approximate the asymptotic null distributions for these two tests in the next section.

Remark 2 We have thus far considered the validity of the parametric specification of the return density as in (4). This kind of hypothesis is the so-called "composite hypothesis" in nonparametric goodness of fit tests (see e.g. Chapter 14 of Lehmann and Romano (2005)). The tests developed can be used to test the validity of a specific density function (corresponding to a particular value θ_0), the so-called "simple hypothesis" in nonparametric testing (see also Chapter 14 of Lehmann and Romano (2005)). Considering a

simple hypothesis is useful because sometimes we may wish to test the validity of a specific estimated model. In the second empirical application in Section 7, we apply our methodology to test three estimated models in Eraker et al. (2003), an important paper in empirical stochastic volatility modelling.

4.2   Bootstrap null distribution

In the literature of nonparametric goodness-of-fit tests, a usual problem is that the asymptotic distribution of the density based nonparametric test statistic will provide a poor approximation of the null distribution in finite samples. For example, Fan (1994) has found "large" differences between the finite sample null distribution and the asymptotic approximation. In a related context, Aït-Sahalia et al. (2009) Section 5.2 write that " ... in practical applications, the convergence is slow ... This kind of problem arises in virtually all nonparametric tests in which function estimation is used; thus using the asymptotic distribution directly is naive."

On the other hand, the bootstrap could be used by some infeasible nonparametric tests to approximate the null distribution. In relation to our density based test T_0, Neumann and Paparoditis (2000) have proposed an (i.i.d.) parametric-type bootstrap method to approximate the asymptotic null distribution of their test statistic. While in relation to our distribution function based test T_2, nonparametric-type block bootstrap has been considered in Bhardwaj, Corradi, and Swanson (2008), Corradi and Swanson (2011) to approximate the asymptotic null distribution.

These problems lead us to develop a bootstrap approximation to the null distribution, although it is more computationally intensive[4] than the corresponding asymptotic test, if available. We use a parametric or model-based bootstrap procedure to approximate the distributions of the test statistics under the null hypothesis. A 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. In contrast to the classical bootstrap, where one generates bootstrap samples by resampling the available dataset, the 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

---

[4]The practical implementation of the parametric bootstrap proposed in this paper depends on the computational burden of simulating the parametric model and evaluating the test statistics a large number of times.

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, which is the 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 work.

The parametric bootstrap procedure is as follows (use T_0 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_0^* analogous to T_0. Again, this involves application of the simulation method of Section 3.2 for each bootstrap replication.

Step 3 Repeat step 1 and 2 B times to get a bootstrap sample $T_0^{*1}, \ldots, T_0^{*B}$ for the statistic T_0.

When B is large, the empirical distribution of $T_0^{*1}, \ldots, T_0^{*B}$ approximates the finite sample null distribution.

A theoretical justification of the proposed parametric procedure is not given in this paper. This is a highly non-trivial problem, although it may 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 Testing more general models

5.1 Models with drift, jumps and leverage effects

Many extensions of the benchmark model (1) have been considered in the literature. For example, the Heston model (Heston (1993)) allows for correlation between the Brownian motion processes driving the price and the volatility, the so-called leverage effect. Bakshi et al. (1997), Bates (2000), Andersen et al. (2002) and Pan (2002), among others, have included jump terms in the price process. Duffie et al. (2000), Eraker et al. (2003) and

Eraker (2004) have considered the model where both the price process and the volatility process have jumps. A general parametric stochastic volatility models can be written as,

$$
\begin{aligned}
\mathrm{d}Y_t &= \mu \mathrm{d}t + \sigma_t \mathrm{d}B_t + \xi^y \mathrm{d}N_t^y, \\
\mathrm{d}\sigma_t^2 &= b(\sigma_t^2; \theta)\mathrm{d}t + a(\sigma_t^2; \theta)\mathrm{d}W_t + \xi^v \mathrm{d}N_t^v,
\end{aligned}
\tag{6}
$$

where $(B_t)_{t \geq 0}$ and $(W_t)_{t \geq 0}$ are two Brownian motion processes with correlation coefficient ρ, μ is the drift parameter of the price process, N_t^y and N_t^v are Poisson jumps in the price and volatility process respectively, with jump intensity λ^y and λ^v respectively. ξ^y and ξ^v are random jumps sizes usually assumed to follow certain parametric distributions with unknown parameters. b and a are drift and diffusion functions of the volatility process respectively, and θ is an unknown parameter vector in b and a.

The model validation methodology developed in this paper can be used to test these more general continuous-time stochastic volatility models, as long as the observed returns satisfy the same probabilistic properties as in this paper, namely strict stationarity and β-mixing with exponentially decaying coefficients. As discussed in Appendix A, in the benchmark model (1) these properties are implied by our imposed assumptions (SV0)-(SV5). In the general model (6), checkable conditions for these probabilistic properties would be difficult to establish. However, our tests and the asymptotic results would still be valid if these probabilistic properties are imposed as high-level assumptions directly. Actually, Corradi and Swanson (2005) impose high level assumptions to develop their tests. To be specific, the three test statistics T_0, T_1 and T_2 can be applied in the same way to test the hypothesis

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

where $q_0(x)$ now is implied by the general model (6) against the alternative hypothesis that the distribution of the observed returns does not satisfy this restriction. That is, the developed tests can be used to check the validity of the stationary return distribution implied by model (6).

5.2 Models estimated with the Bayesian approach

The Bayesian approach has been popular in the literature of estimating of stochastic volatility models (see e.g. Jacquier et al. (2002), Jacquier et al. (2004), Yu and Meyer (2006), Chib et al. (2002), Eraker (2001), Eraker et al. (2003) and Eraker (2004), among others) because of its ability to deal with the latent volatility process. It is thus natural to ask if the tests developed in this paper could be used to test models estimated with

Bayesian methods.

In principle, the Bayesian paradigm does not allow for model specification tests, as they involve sharp hypotheses and asymptotic distributions. However, most empirical researchers only use the Bayesian approach as an estimation method for the parameters, so the misspecification issue can be analyzed (see e.g. Kleijn and van der Vaart (2012) and Müller (2013)). According to the assumptions used in Theorem 1-3, the models estimated with Bayesian approaches can be tested with our methodology if the Bayesian point estimator is $\sqrt{n}$-consistent (for T_0 and T_1) and if it satisfies the first-order expansion (P1b) (for T_2), both in a frequentist asymptotic sense. It is known that under rather general conditions (see e.g. Chapter 10 of Van der Vaart (2000)), a sensible Bayesian point estimator (for example the posterior mode estimator) will be equivalent to a maximum likelihood estimator asymptotically. In this sense, testing models estimated with a Bayesian approach using our tests is not a problem.

6 Monte Carlo simulations

In this section, we study the finite sample performance of the density-based tests T_0 and T_1 and the distribution-based test T_2. We use the bootstrap method described in the previous section to determine the null distribution of the test statistic.

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 likelihood based methods, but it achieves a good compromise between estimation efficiency and computation time. To save space, for all the simulated size and power results, we only report the results at the 5% significance level.[5]

6.1 Size of the tests

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

$$
\begin{aligned}
\mathrm{d}Y_t &= \sigma_t \mathrm{d}B_t, \\
\mathrm{d}\sigma_t^2 &= 5(0.1 - \sigma_t^2)\mathrm{d}t + 0.75\sqrt{\sigma_t^2}\mathrm{d}W_t,
\end{aligned}
\tag{7}
$$

---

[5]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.

where the two Brownian motions W and B are independent.[6] Within one day, 10 steps are simulated to reduce the discretization error. We consider the sample sizes 1000, 2000 and 3000, roughly corresponding to 4 years, 8 years and 12 years of daily observations.

With the parameters and sample sizes, the test statistics T_0, T_1 and T_2 are simulated 1000 times. The distribution of these realized test statistics are taken as the true distribution (except for the Monte Carlo simulation error). For each of the realized 1000 sample paths, we obtain 5 bootstrap samples and compute their resulting test statistics T_0^*, T_1^* and T_2^*. Aggregating across 1000 samples yields 5000 bootstrap statistics. Their sampling distribution is taken as the distribution of the bootstrap method. Note that this aggregation involves an approximation, based on the assumption that the parameter estimates in each of the Monte Carlo replications are close enough to treat the corresponding bootstrap distributions as identical. This approximation is made to keep the computation time of the Monte Carlo experiments manageable; a similar approach was taken by Aït-Sahalia et al. (2009).

Table 1 summarizes the simulated 5% actual size of all the tests for the three sample sizes. The bootstrap tests seem to have reasonable size properties, especially when the sample size is large. For T_0 and T_2, the actual rejection frequencies approach the nominal size as the sample size increases, though this is not the case for T_1.

	T_0	T_1	T_2
$n = 1000$	0.0277	0.0321	0.0377
$n = 2000$	0.0372	0.0386	0.0397
$n = 3000$	0.0406	0.0375	0.0411

Table 1: Size of the bootstrap tests.

6.2 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 Monte Carlo replications. We take the Heston model (7) as the null hypothesis. We evaluate the power functions of the three test statistics under the three families of alternative models. Each family of models is indexed by τ, with $\tau = 0$ corresponding to the null model (7).

---

[6]Under the so-called in-fill sampling scheme, where $\Delta \to 0$, an explicit formula of the return density can be obtained as the variance mixture of the stationary volatility density with a standard normal variable, see e.g. Genon-Catalot et al. (1999). However with the fixed Δ sampling scheme used in this paper, the return is a variance mixture of the integrated volatility $1/\Delta\sqrt{\int_{t_{i-1}}^{t_i} \sigma_t^2 dt}$ with a standard normal variable. Since the density of $1/\Delta\sqrt{\int_{t_{i-1}}^{t_i} \sigma_t^2 dt}$ is not the stationary volatility density and not known, a closed-form formula for the return is not available.

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

6.2.1 Misspecification in the drift function

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

$$\mathrm{d}\sigma_t^2 = \{(1-\tau)(\alpha(\beta - \sigma_t^2) + \tau\mu(\sigma_t^2)\}\mathrm{d}t + \gamma\sqrt{\sigma_t^2}\mathrm{d}W_t, \tag{8}$$

for $\tau = 0, 0.1, \ldots, 1$, 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 grows 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.

6.2.2 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 (7). In the alternative model, the drift function remains the same, but the diffusion function corresponds to a GARCH diffusion process.

$$\mathrm{d}\sigma_t^2 = \alpha(\beta - \sigma_t^2)\mathrm{d}t + \{(1-\tau)\gamma\sqrt{\sigma_t^2} + \tau\rho(\sigma_t^2)\}\mathrm{d}W_t, \tag{9}$$

16

(a) Drift functions

(b) Power of T_0

(c) Power of T_1
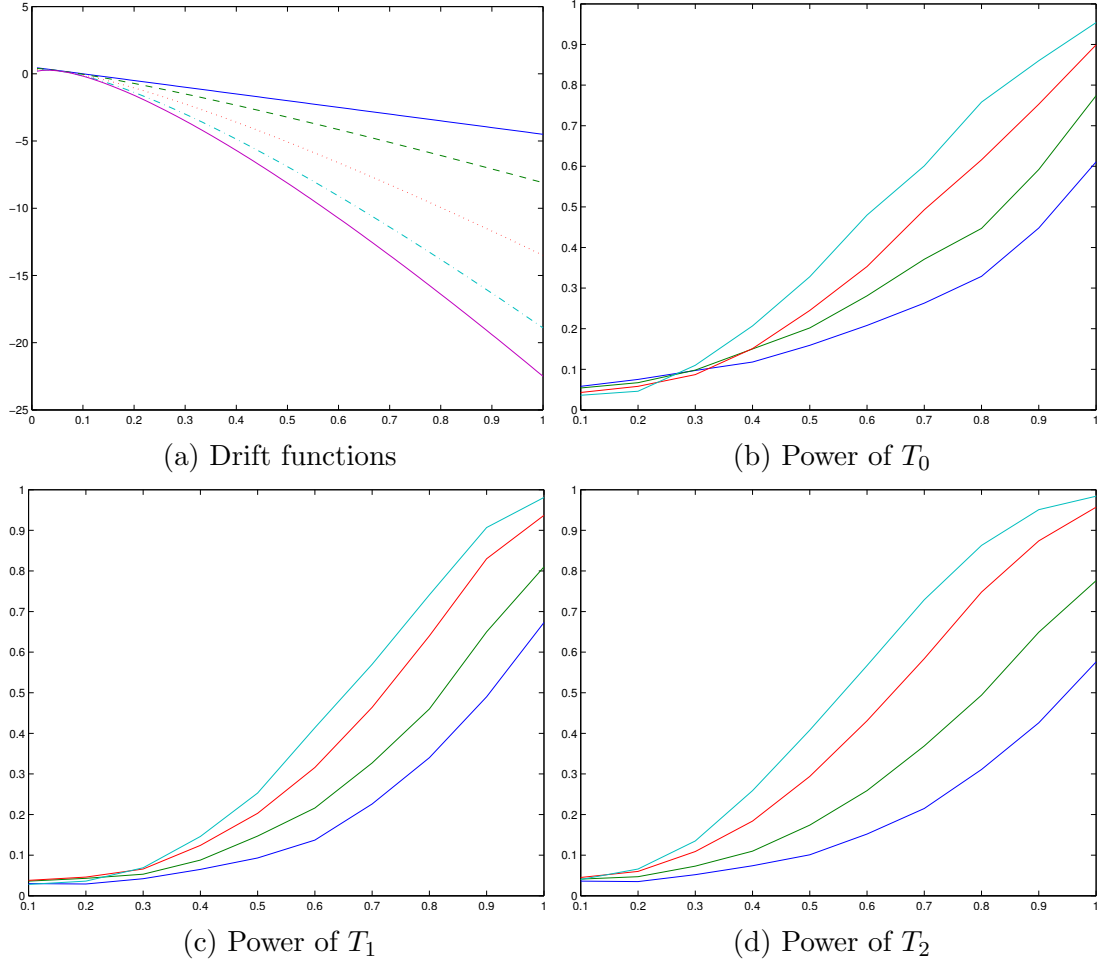
(d) Power of T_2

Figure 1: Power under 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).

for $\tau = 0, 0.1, \ldots, 1$, 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 grows large. The power of T_0 seems to be slightly better than T_2 when the deviation is small, while when the deviation is large, the power seems to be similar.

6.2.3 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

$$d\sigma_t^2 = \alpha(\beta - \sigma_{t-}^2)dt + \gamma\sqrt{\sigma_{t-}^2}\,dW_t + \xi^v dN_t^y, \tag{10}$$

(a) Diffusion functions

(b) Power of T_0 under different sample sizes

(c) Power of T_1 under different sample sizes
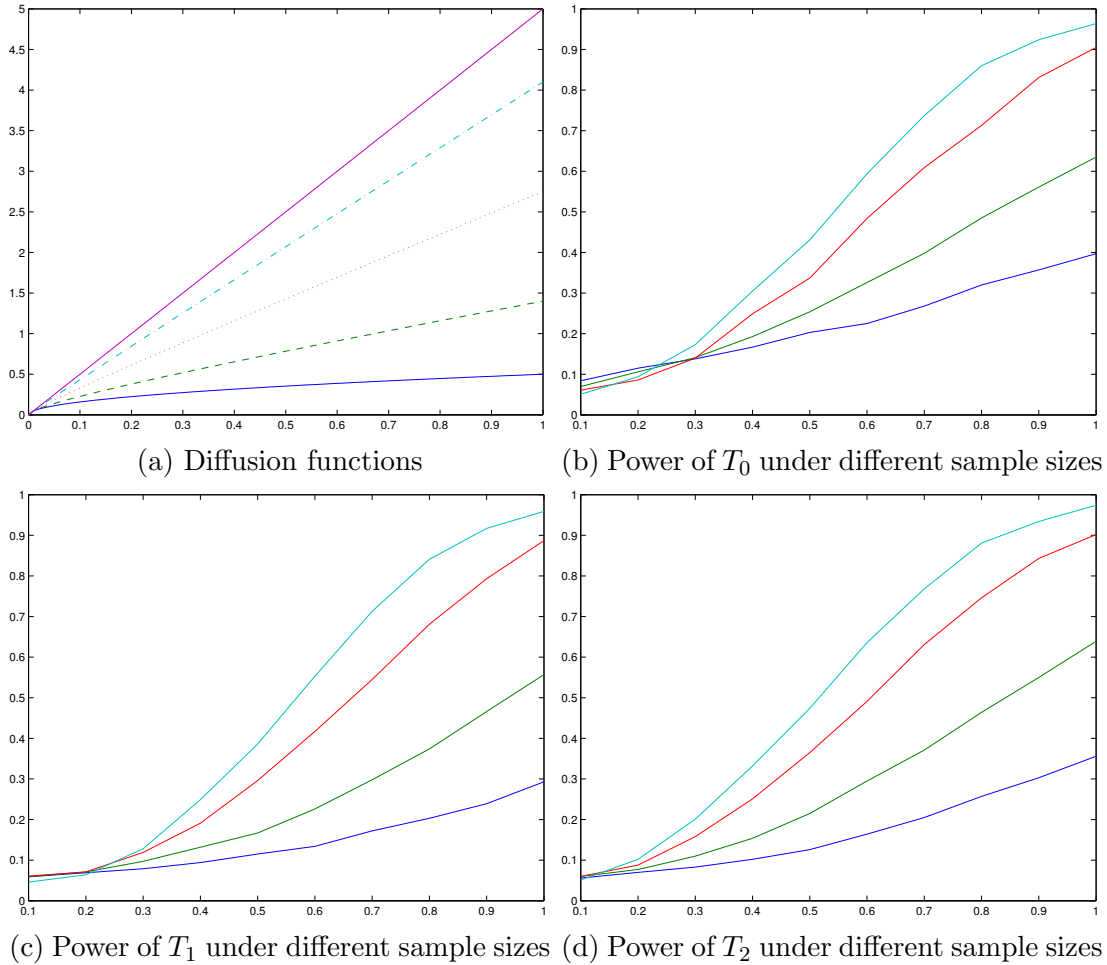
(d) Power of T_2 under different sample sizes

Figure 2: Power under 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).

where $\sigma_{t_-}^2 := \lim_{s \uparrow t} \sigma_s^2$, N_t^v is a Poisson process with intensity λ^v, ξ^v is the jump size following an exponential distribution, independent of (W_t) and (B_t). We use the Compound Poisson type, large and infrequent, jump specification for the volatility jumps, which is a stylized fact observed for daily level equity or equity index data. We consider 5 jump intensities: 0.5, 1, 1.5, 2, 2.5 and 3. These numbers can be understood as the average number of jumps in a year. The jump intensity values are based on the estimation result in Eraker et al. (2003) for the S&P 500 data, where the estimated jump intensity is 1.5 jump per year with a standard deviation of 0.5 jump per year. The jump size ξ^y is exponentially distributed with mean 0.2, corresponding to a 20% average sized jump. The jump size distribution and parameter are also chosen based on the model and estimation results in Eraker et al. (2003). 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.

(a) Power of T_0

(b) Power of T_1
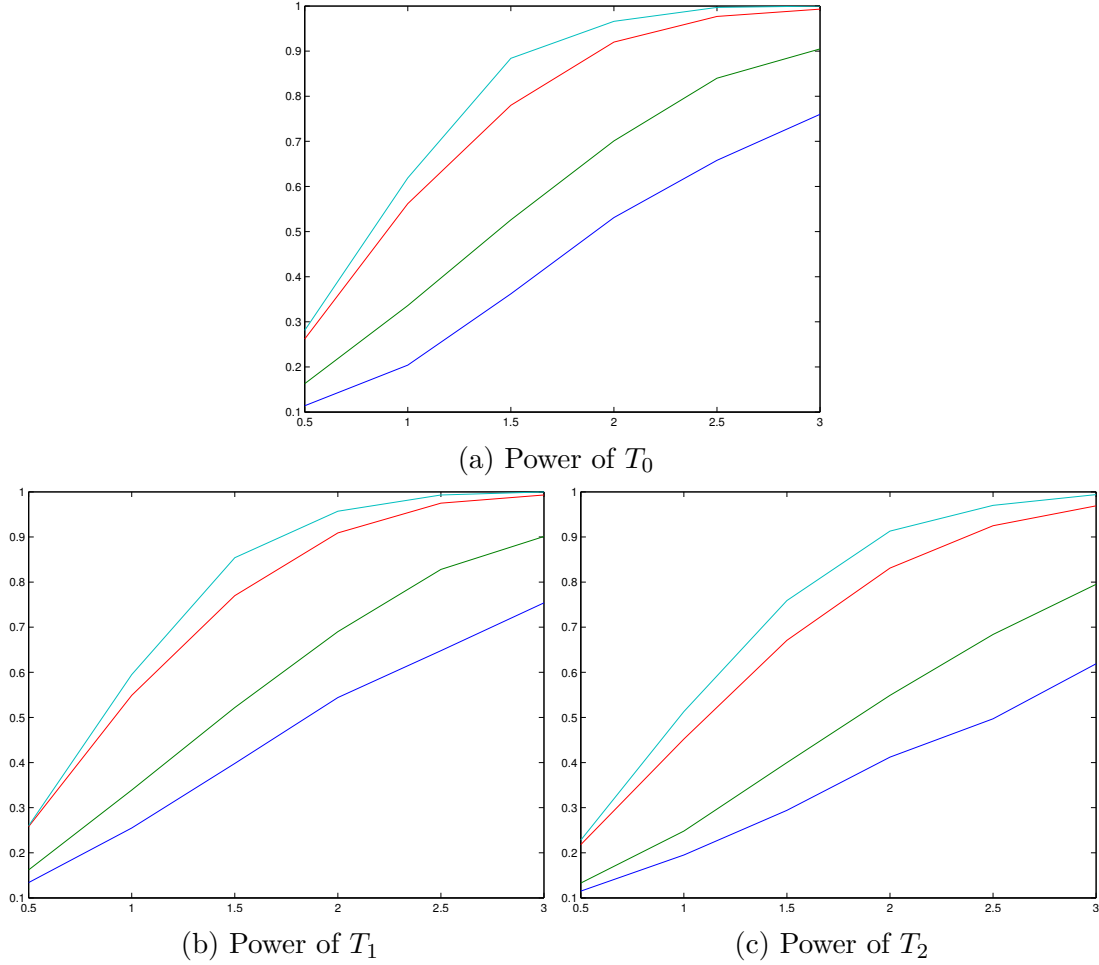
(c) Power of T_2

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

6.2.4 Jumps in the price process

We now consider the power of the three tests when the alternative model have jumps in the price process.[7] To be specific, we consider a sequence of alternative models having a Poisson-type jump in the price process with jump size ξ_t^y following a normal distribution, while the volatility model is correctly specified:

$$
\begin{aligned}
\mathrm{d}Y_t &= \sigma_t \mathrm{d}B_t + \xi^y \mathrm{d}N_t^y, \\
\mathrm{d}\sigma_t^2 &= 5(0.1 - \sigma_t^2)\mathrm{d}t + 0.75\sqrt{\sigma_t^2}\mathrm{d}W_t,
\end{aligned}
\tag{11}
$$

---

[7]We are grateful for the AE for suggesting this simulation.

where N_t^y is a Poisson process with jump intensity λ^y, ξ^y follows i.i.d. normal distribution with mean μ^y and standard deviation σ^y. We set $\rho = -0.4$, $\mu^y = -0.1$ (an average price jump is -10% on an annual basis) and $\sigma^y = 0.1$. The price jump size distribution and its parameters are based on the specification used in Eraker et al. (2003) and the estimation results therein. A negative μ^y is in line with empirical observation as most of the price jumps are downwards. The sequence of models we consider have different jump intensities: $\lambda^y = 5, 10, 15, 20, 25, 30$. Figure 4 gives the power functions of the three tests under different sample sizes. We observe again the monotonic increase in power of the three tests for this type of deviations to the null hypothesis. Again as the jump intensity of the price process increases, the deviations are easier to identify. The distribution function based test seems more powerful than the density function based test for this type of deviations.

By comparing the power results for price jump deviations and the volatility jump deviations in Section 6.2.3, we have also found some interesting results on how the return distribution is affected differently by these two types of jumps. In Section 6.2.3 we have seen that the power of the three tests are close to 1 when the volatility jump intensity is close to 3 (on average 3 jumps a year); while in this section, the power of all the tests get close to 1 when the jump intensity is close to 30 (on average 30 jumps a year). This clearly means that the return distribution based tests are more sensitive to volatility jumps, hinting that jumps in volatility have a large effect on the return distribution.

Considering the modelling implication of the two types of jumps we find that the obtained results are reasonable: volatility jumps, even if they only happen a few times a year, will bring persistent change to the volatility process as it needs to revert to its long-run mean for an extended period of time, which will cause substantial distortion to the return distribution. While price jumps, if they only happen a few times a year, even when the jump size is large, only add a couple of "outliers" to the tail of the return distribution and will not have a big influence on the shape of the return distribution. Only when the jump intensity is high enough (30 here), the tail part will receive enough mass and the change in the return distribution will become obvious. The testing results echo the comments made in Eraker et al. (2003) that "the volatility jumps bring persistent change to the returns and price jumps bring transient changes...", but they also provide a concrete and transparent illustration of the difference.

6.3   Power performance of the tests in a more general model

Section 5 has discussed that the tests developed in this paper can be used to test more general models. To study the robustness of the power performance of the three tests in more general models, in this section we redo the Monte Carlo experiment in the previous
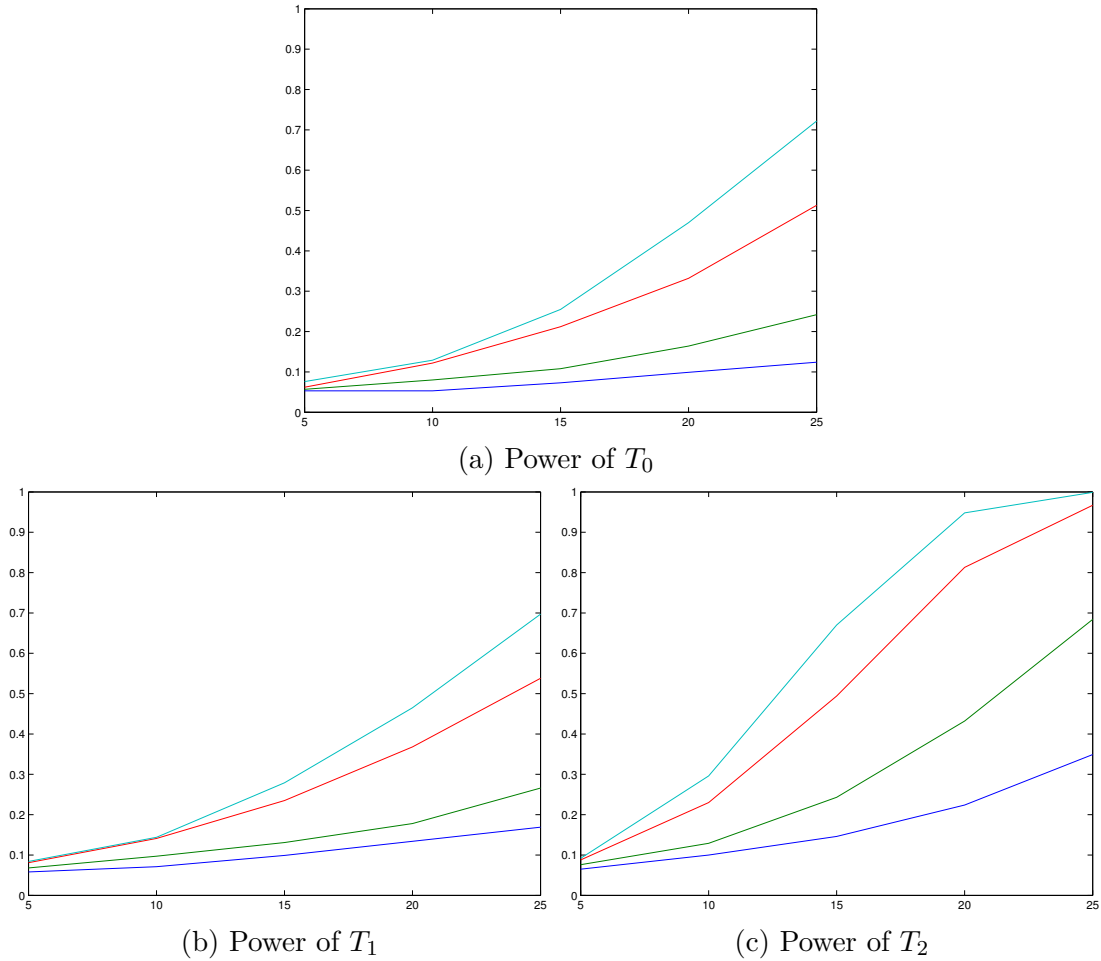
Figure 4: Power under different price jump intensities. (a), (b), (c): $n = 500$ (blue), $n = 1000$ (green), $n = 2000$ (red), $n = 3000$ (cyan).

subsection in a model with leverage effect.

To be specific, in the null model (1) the Brownian motions $\{W_t\}_{t\geq 0}$ and $\{B_t\}_{t\geq 0}$ now correlated with a coefficient $\rho = -0.4$. We consider the same 4 types of deviations as in (8), (9), (10) and (11) in Section 6.2, the corresponding power plots are given in Figure 1-4 of the online appendix to this paper to save space. The simulation results show that, when the leverage effect exists, the power of the tests is not much affected.

7 Empirical applications

7.1 Apple stock, 2000–2014

In this section, we apply our tests to a daily Apple stock price dataset, with the purpose of illustrating the use of the tests proposed in this paper. The dataset contains the adjusted closing prices from January 3rd, 2000 to February 3rd, 2014, making 3543 observations in total. Figure 5 gives the plot of the log-price levels and log-returns.
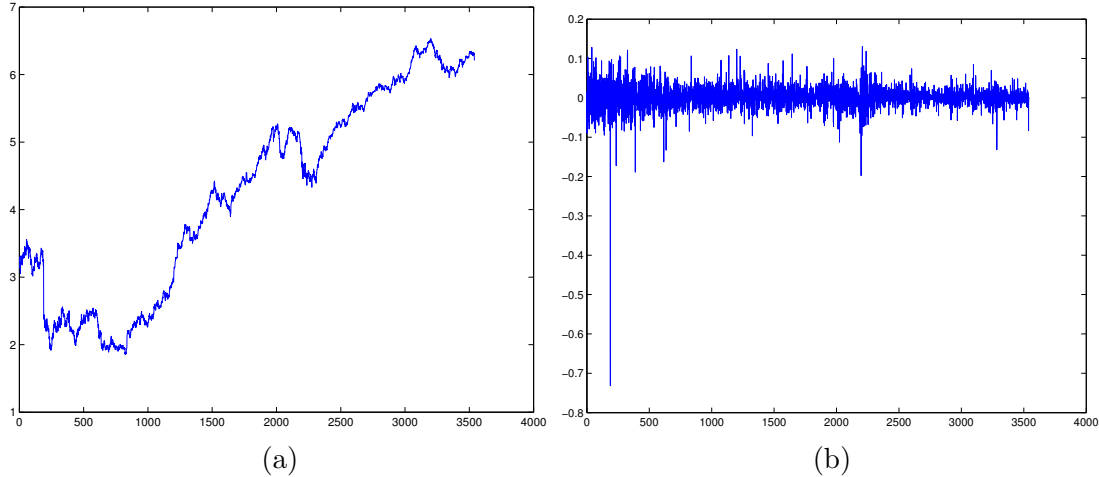

Figure 5: (a): Log prices of Apple stock, 2000–2014. (b): Log returns of Apple stock, 2000–2014.

The Heston model (7) is fitted to the data, using the GMM approach of Meddahi (2002). The estimated parameters are $\hat{\alpha} = 17.5119$, $\hat{\beta} = 0.1793$, $\hat{\gamma} = 2.4715$. We then apply the nonparametric specification tests proposed in this paper to study the validity of the Heston model. 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_0, T_1 and T_2 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.

Although it is well-known that the Heston model is misspecified for many return series, the results of this empirical application provides evidence that the model misspecification (partly) comes from the misspecification in the marginal distribution of the model. We also give a plot of the nonparametrically estimated density of Apple returns and the corresponding model implied density in Figure 6. Although the parametric density provides a reasonable fit to the empirical density, our tests results show that the discrepancy is large enough to reject the model. Also, the parametric density seems to exhibit a slight location shift from the empirical density, which is also in line with the observed lower

p-value for the distribution function based test and the known results in nonparametric tests that the distribution function based tests are in general more sensitive to global shifts of the whole distribution. The location shift could be caused by a combination of price jumps and leverage effects, although the skewness and excess kurtosis in the empirical density (relative to the parametric density) are not very large.
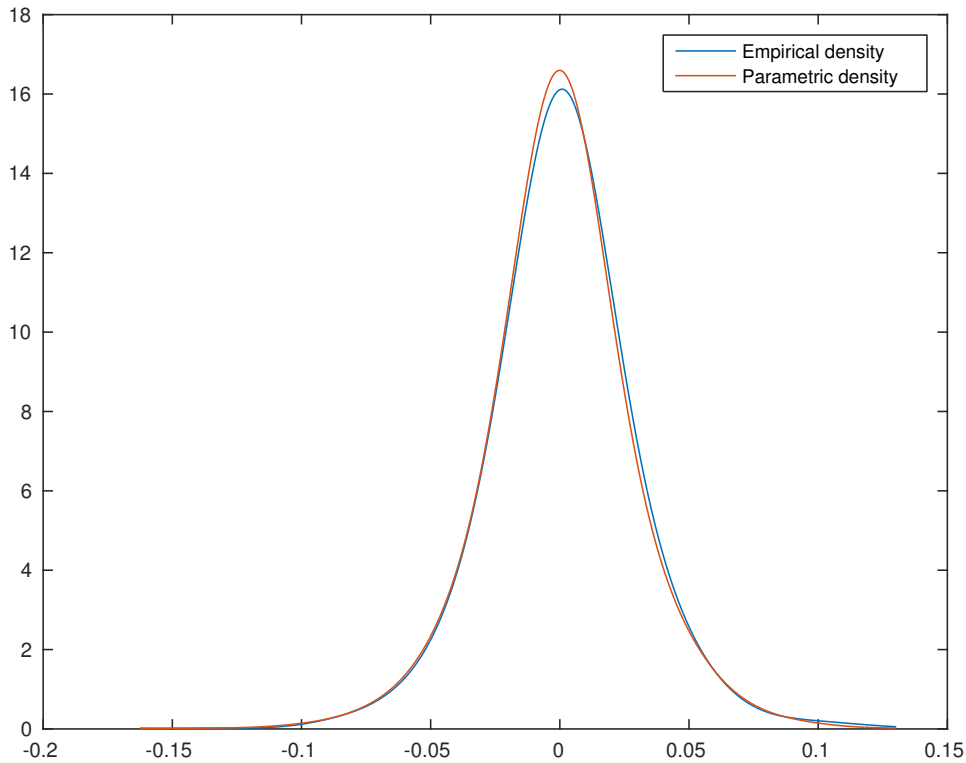

Figure 6: Nonparametric and parametric density estimates of Apple returns.

7.2 Eraker et al. (2003)'s models of S&P 500 returns

The Heston model is popular in empirical finance and has been used widely in option pricing. However, the statistical fit of the Heston model has long been questioned; one reason is that it has limited capacity in modelling the large movements in empirical returns. Towards a more realistic model for the stock price, Bakshi et al. (1997), Bates (2000), Andersen et al. (2002) and Pan (2002), among others, have included jump terms in the price process. The empirical results in Bakshi et al. (1997) and Bates (2000) have also pointed to the possible existence of jump terms in the volatility process. Eraker et al. (2003) have advanced the empirical literature of stochastic volatility in this direction by considering stochastic volatility models with jumps in both the price process and the volatility process. They estimate the model with a likelihood based method implemented by a Markov Chain Monte Carlo sampling scheme and find evidence for the existence

of volatility jumps. To be specific, Eraker et al. (2003) have fitted 20 years of S&P 500 equity index data to four models, namely a stochastic volatility model (SV), which is essentially the Heston model; a stochastic volatility model with price jumps (SVJ); a stochastic volatility model with correlated price and volatility jumps (SVCJ); and a stochastic volatility model with independent price and volatility jumps (SVIJ). Eraker et al. (2003) have found that the models with jumps in price *and* volatility (both the SVIJ and the SVCJ model) have the best fit.

Although Eraker et al. (2003) have used several diagnostics to test the validity of their models, none of these has considered the validity of the model implied return distribution. Treating the estimated models as specific models, we apply our tests to study the validity of the model implied return distributions using the methodology discussed in Section 5.

Since the SVIJ model and the SVCJ model exhibit very similar properties in all the diagnostics in Eraker et al. (2003), with the SVIJ model performing slightly better, we thus only study the SVIJ model, together with the SV model and the SVJ model. We acquire the same S&P 500 index data from January 2, 1980 to December 31, 1999, and test the estimated model with these data.

Since our tests are based on the density function and the distribution function of the return data, we first give a plot of the empirical density function estimated and all the model implied return densities in Figure 7. The empirical density is estimated using a kernel density estimator with the bandwidth selected by the cross-validation method. It is seen that there are some discrepancies between the empirical density and the model implied densities, especially at the left tail and around the peak. The SV model implies a return density with a lower peak than the empirical density, its distance to the empirical density is also the largest in the left tail. The SVJ model improves upon the SV model both in the peak of the density and in the left tail. One surprising observation is that the SVIJ model does not seem to improve the distributional fit of the returns upon the SVJ model: the peak of the SVIJ model implied density is even slightly further away from the empirical density than the SV model, but the SVIJ model does seem to improve the fit of the left tail over the SV model, and has a left tail indistinguishable from the SVJ model. Visual inspection suggests that the SVJ model probably provides the best fit to the empirical density. We next apply the statistical tests developed in this paper to see if this is indeed the case.

We apply the three tests to the three estimated models and calculate the p-values of the tests based on the parametric bootstrap method. Cross-validation is used to determine the bandwidth and 1000 parametric bootstrap replications are used. The p-values for the three tests for the SV model, the SVJ model and the SVIJ model are given in Table 2. It seems that the statistical test results are consistent with our visual inspection of the
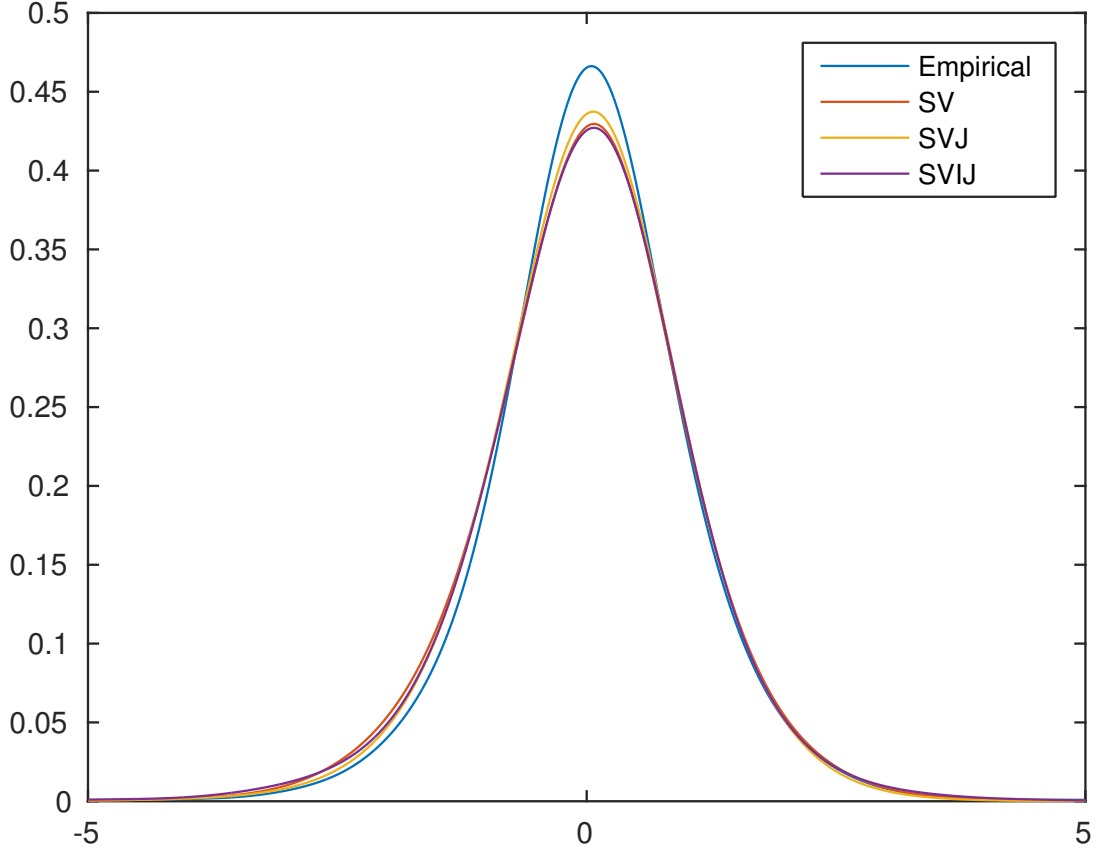
24

Figure 7: Empirical density of S&P 500 daily returns and model implied return densities.

	T_0	T_1	T_2
SV	0.029	0.033	0.017
SVJ	0.189	0.194	0.129
SVIJ	0.039	0.040	0.022

Table 2: p-values of all the tests for the models in Eraker et al. (2003).

densities. The SV model implied density is too far from the empirical density, thus is rejected significantly at the 5% level. The SVJ model improves over the SV model, and we don't find enough statistical evidence to reject the density implied by it, even using a 10% significance level. Our tests confirm that the SVIJ model implied return density is worse than that implied by the SVJ model, it is not favoured by all three tests and is rejected significantly at the 5% level.

With these test results, we find some new evidence on the statistical fit of the models considered in Eraker et al. (2003). As discussed in Eraker et al. (2003), page 1274, "models with only jumps in returns and diffusive stochastic volatility can generate realistic patterns of both unconditional and conditional nonnormalities ...". Our test results confirm this claim that the SVJ model is sufficient in modelling the unconditional return

distribution. They also write, "... but they (return jumps) have difficulty capturing the dynamics of the conditional volatility of returns," for which jumps in the volatility would probably be needed. Here, Eraker et al. (2003) have discussed the potential improvement offered by the SVIJ model. From our evidence, however, we remark that even if the SVIJ model could potentially improve the fit of the conditional return distribution (dynamics of conditional volatility), this improvement does not come without any costs: introducing the volatility jump factor leads to a worse fit of the unconditional distribution relative to the SVJ model.

8 Conclusion and extensions

We propose three tests for stochastic volatility model specification by comparing the parametrically and nonparametrically estimated stationary marginal density functions and distribution functions of the observed returns. The consistency of the three tests are derived and we have studied their finite sample power property under various alternative models. In empirical applications, our tests provide some new insights in the model misspecification issue of some popular stochastic volatility models.

To consider extensions, our approach can be adapted to discrete-time stochastic volatility models easily, as long as the volatility process is stationary and β-mixing as assumed in this paper. One can then compare the estimated density functions and distribution functions as discussed for continuous-time models analogously. Another possible extension to consider is to formulate the test statistics based on the conditional distribution of returns: as discussed in the introductory section, the stationary marginal return distribution does not contain information on the dynamics in the data. To exploit the dependence structure in the model, we could extend our approach 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. These are left for future research.

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 two dimensional joint densities of $(X_1, \dots, X_n)$ 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 $x \in \mathbb{R}$.

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

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

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*

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

and the *speed measure*

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

Then the assumption

(SV2)

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

where the $\cdot$ in the integral means an 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

The 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 an 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_0. Notice that

$$
\begin{aligned}
T_0 \\
= \; & nh^{1/2} \int (\hat{q}(x) - K_h * q(x; \hat{\theta}_n))^2 \mathrm{d}x \\
= \; & nh^{1/2} \int (\hat{q}(x) - K_h * q(x))^2 \mathrm{d}x + nh^{1/2} \int (K_h * q(x) - K_h * q(x; \hat{\theta}_n))^2 \mathrm{d}x \\
& + nh^{1/2} \int (\hat{q}(x) - K_h * q(x))(K_h * q(x) - K_h * q(x; \hat{\theta}_n)) \mathrm{d}x.
\end{aligned}
$$

Define

$$T_0' = nh^{1/2} \int (\hat{q}(x) - K_h * q(x))^2 \mathrm{d}x.$$

It will be shown later that $T_0' = O_p(1)$; the second term $nh^{1/2} \int (K_h * q(x) - K_h * q(x; \hat{\theta}_n))^2 dx = O_p(h^{1/2})$ because $\hat{\theta}_n$ is a $\sqrt{n}$-consistent estimator of θ_0 and by the square integrable assumption on the Lipschitz constant $L(x)$ as in Assumption (P1a), so this term is dominated by T_0'; the crossproduct term is clearly dominated by T_0' by the Cauchy-Schwarz inequality. Thus we have that

$$T_0 = T_0'(1 + o_p(1)),$$

and we derive the asymptotic distribution of T_0' in the following.

First notice

$$
\int \left(\frac{1}{n} \sum_{i=1}^{n} (K_h(x - X_i) - K_h * q(x)) \right)^2 dx
$$
$$
= \frac{1}{n^2} \sum_{i=1}^{n} \int (K_h(x - X_i) - K_h * q(x))^2 dx
$$
$$
+ \frac{2}{n^2} \sum_{i<j} \int (K_h(x - X_i) - K_h * q(x))(K_h(x - X_j) - K_h * q(x)) dx
$$
$$
:= \frac{1}{n^2} \sum_{i=1}^{n} \varphi_n(X_i, X_i) + \frac{2}{n^2} \sum_{i<j} \varphi_n(X_i, X_j),
$$

where
$$
\varphi_n(u, v) := \int (K_h(x - u) - K_h * q(x))(K_h(x - v) - K_h * q(x)) dx.
$$

Next, we show that

1. The sum of the diagonal terms

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

2. The sum of the off-diagonal terms

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

3. Then we show that

$$
T_0' \xrightarrow{d} N\left(0, 2 \int q_0^2(u) du \int (K^{(2)}(u))^2 du \right),
$$

and the asymptotic distribution of T_0 follows easily.

Step 1   We first compute the order of the mean,

$$E\left(\frac{1}{n^2}\sum_{i=1}^{n}\varphi_n(X_i, X_i)\right)$$

$$= \frac{1}{n}E\varphi_n(X_1, X_1)$$

$$= \frac{1}{n}\int\int(K_h(x - X_1) - K_h * q(x))^2\mathrm{d}xq(X_1)\mathrm{d}X_1$$

$$= \frac{1}{nh^2}\int\int\left(K\left(\frac{x - X_1}{h}\right)\right)^2 q(X_1)\mathrm{d}x\mathrm{d}X_1(1 + o(1))$$

$$= \frac{1}{nh}\int\int(K(u))^2q(X_1)\mathrm{d}u\mathrm{d}X_1(1 + o(1))$$

$$= (nh)^{-1}\int K^2(u)\mathrm{d}u(1 + o(1)). \tag{12}$$

Then we compute the order of the variance

$$\mathrm{Var}\left(\frac{1}{n^2}\sum_{i=1}^{n}\varphi_n(X_i, X_i)\right)$$

$$\leqslant E\left(\frac{1}{n^2}\sum_{i=1}^{n}\varphi_n(X_i, X_i)\right)^2$$

$$= \frac{1}{n^3}E\varphi_n^2(X_1, X_1) + \frac{2}{n^4}\sum_{i<j}E\varphi_n(X_i, X_i)\varphi_n(X_j, X_j). \tag{13}$$

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

$$E\varphi_n^2(X_1, X_1)$$

$$= \int\left(\int K_h^2(x - X_1)\mathrm{d}x\right)^2 q(X_1)\mathrm{d}X_1(1 + o(1))$$

$$= \frac{1}{h^4}\int\left(\int K^2\left(\frac{x - X_1}{h}\right)\mathrm{d}x\right)^2 q(X_1)\mathrm{d}X_1(1 + o(1))$$

$$= \frac{1}{h^2}\int\left(\int K^2(u)\mathrm{d}u\right)^2 q(X_1)\mathrm{d}X_1(1 + o(1))$$

$$= O\left(\frac{1}{h^2}\right). \tag{14}$$

For the second term,

$$
\begin{aligned}
& E\varphi_n(X_i, X_i)\varphi_n(X_j, X_j) \\
&= \int\int\int K_h^2(x - X_i)\mathrm{d}x \int K_h^2(x - X_j)\mathrm{d}x q(X_i, X_j)\mathrm{d}X_i\mathrm{d}X_j(1 + o(1)) \\
&= \frac{1}{h^2}\int\int\left(\int K^2(u)\mathrm{d}u\right)^2 q(X_i, X_j)\mathrm{d}X_i\mathrm{d}X_j(1 + o(1)) \\
&= O\left(\frac{1}{h^2}\right).
\end{aligned}
\tag{15}
$$

Use the result in (14) and (15) in (13), we get

$$
\begin{aligned}
& \mathrm{Var}\left(\frac{1}{n^2}\sum_{i=1}^{n}\varphi_n(X_i, X_i)\right) \\
&\leqslant \frac{1}{n^3}O\left(\frac{1}{h^2}\right) + \frac{2}{n^4}n^2 O\left(\frac{1}{h^2}\right) \\
&= O\left(\frac{1}{n^2 h^2}\right).
\end{aligned}
\tag{16}
$$

Use the results in (12) and (16) and apply Markov's inequality we have

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

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

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

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

$$
E\varphi_n^2(\tilde{X}_i, \tilde{X}_j)
$$

$$
= \int\int\left(\int K_h(x - X_i)K_h(x - X_j)\mathrm{d}x\right)^2 q(X_i)q(X_j)\mathrm{d}X_i\mathrm{d}X_j(1 + o(1))
$$

$$
= \frac{1}{h^2}\int\int\left(\int K(u)K\left(u + \frac{X_i - X_j}{h}\right)\mathrm{d}x\right)^2 q(X_i)q(X_j)\mathrm{d}X_i\mathrm{d}X_j(1 + o(1))
$$

$$
= \frac{1}{h^2}\int\int\left(K^{(2)}\left(\frac{X_i - X_j}{h}\right)\right)^2 q(X_i)q(X_j)\mathrm{d}X_i\mathrm{d}X_j(1 + o(1))
$$

$$
= \frac{1}{h}\int\int(K^{(2)}(u))^2 q(X_j + uh)q(X_j)\mathrm{d}u\mathrm{d}X_j(1 + o(1))
$$

$$
= \frac{1}{h}\int(K^{(2)}(u))^2\mathrm{d}u\int q^2(x)\mathrm{d}x(1 + o(1)).
$$

then the asymptotic variance is

$$
\sigma_n^2 = \frac{n^2}{2}E\varphi_n^2(\tilde{X}_i, \tilde{X}_j) = \frac{n^2}{2h}\int(K^{(2)}(u))^2\mathrm{d}u\int q^2(x)\mathrm{d}x(1 + o(1)). \qquad (17)
$$

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 $1 > \delta > 0$,

$$
E|\varphi_n(X_1, X_j)\varphi_n(X_i, X_j)|^{1+\delta}
$$

$$
= \int\int\int\left|\int K_h(x - X_1)K_h(x - X_j)\mathrm{d}x\int K_h(x - X_i)K_h(x - X_j)\mathrm{d}x\right|^{1+\delta}
$$
$$
q(X_1, X_i, X_j)\mathrm{d}X_1\mathrm{d}X_i\mathrm{d}X_j(1 + o(1))
$$

$$
= \int\int\int\left|\frac{1}{h^4}\int K\left(\frac{x - X_1}{h}\right)K\left(\frac{x - X_j}{h}\right)\mathrm{d}x\int K\left(\frac{x - X_i}{h}\right)K\left(\frac{x - X_j}{h}\right)\mathrm{d}x\right|^{1+\delta}
$$
$$
q(X_1, X_i, X_j)\mathrm{d}X_1\mathrm{d}X_i\mathrm{d}X_j(1 + o(1))
$$

$$
= \int\int\int\left|\frac{1}{h^2}K^{(2)}\left(\frac{X_1 - X_j}{h}\right)K^{(2)}\left(\frac{X_i - X_i}{h}\right)\right|^{1+\delta} q(X_1, X_i, X_j)\mathrm{d}X_1\mathrm{d}X_i\mathrm{d}X_j(1 + o(1))
$$

$$
= h^2\int\int\int\left|\frac{1}{h^2}K^{(2)}(u)K^{(2)}(v)\right|^{1+\delta} q(X_j + uh, X_j + vh, X_j)\mathrm{d}u\mathrm{d}v\mathrm{d}X_j(1 + o(1))
$$

$$
= \frac{1}{h^{2\delta}}\left(\int|K^{(2)}(u)|^{1+\delta}\mathrm{d}u\right)^2\int q(X_j + uh, X_j + vh, X_j)\mathrm{d}X_j(1 + o(1))
$$

$$
= O\left(\frac{1}{h^{2\delta}}\right).
$$

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

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

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

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

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

$$E\left|\int \varphi_n(X_1,X_i)\varphi_n(X_1,X_j)\mathrm{d}P(X_1)\right|^{2(1+\delta)}$$

$$=\iint\left|\int\left(\int K_h(x-X_1)K_h(x-X_i)\mathrm{d}x\int K_h(x-X_1)K_h(x-X_j)\mathrm{d}x\right)q(X_1)\mathrm{d}X_1\right|^{2(1+\delta)}$$
$$q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1))$$

$$=\iint\left|\int\frac{1}{h^4}\left(\int K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_i}{h}\right)\mathrm{d}x\int K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_j}{h}\right)\mathrm{d}x\right)q(X_1)\mathrm{d}X_1\right|^{2(1+\delta)}$$
$$q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1))$$

$$=\iint\left|\int\frac{1}{h^2}K^{(2)}\left(\frac{X_1-X_i}{h}\right)K^{(2)}\left(\frac{X_1-X_j}{h}\right)q(X_1)\mathrm{d}X_1\right|^{2(1+\delta)}q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1))$$

$$=\iint\left|\int\frac{1}{h}K^{(2)}(u)K^{(2)}\left(u+\frac{X_i-X_j}{h}\right)q(X_i+uh)\mathrm{d}u\right|^{2(1+\delta)}q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1))$$

$$\leqslant\frac{1}{h^{2(1+\delta)}}\iiint\left|K^{(2)}(u)K^{(2)}\left(u+\frac{X_i-X_j}{h}\right)\right|^{2(1+\delta)}q(X_i+uh)\mathrm{d}u\,q(X_i,X_j)\mathrm{d}X_i\mathrm{d}X_j(1+o(1))$$

$$=\frac{h}{h^{2(1+\delta)}}\iiint|K^{(2)}(u)K^{(2)}(u+v)|^{2(1+\delta)}q(X_i+uh)\mathrm{d}u\,q(X_i,X_i-vh)\mathrm{d}X_i\mathrm{d}v(1+o(1))$$

$$\leqslant C\times\frac{h}{h^{2(1+\delta)}}\left(\int|K^{(2)}(u)|^{2(1+\delta)}\mathrm{d}u\right)^2$$

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

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

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

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

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

We have then checked the condition

$$\frac{1}{\sigma_n^2}\left\{n^2\left\{M_{n1}^{\frac{1}{1+\delta}}+M_{n5}^{\frac{1}{2(1+\delta)}}+M_{n6}^{\frac{1}{2}}\right\},n^{\frac{3}{2}}\left\{M_{n2}^{\frac{1}{2(1+\delta)}}+M_{n3}^{\frac{1}{2}}+M_{n4}^{\frac{1}{2(1+\delta)}}\right\}\right\}\to0,$$

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

Step 3 The asymptotic distribution of T_0' is an easy consequence of Step 1 and 2. The asymptotic distribution of T_0 is obtained by noticing $T_0 = T_0'(1 + o_p(1))$.

For the asymptotic distribution of T_1, first, using 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_1 follows easily from Slutsky's lemma. $\qquad\square$

In proving Theorem 2, we will need the following weak convergence results for empirical processes 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

$$J_{[\,]}(\infty, \mathcal{F}, L_p(P)) < \infty. \tag{18}$$

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

$$\sqrt{n}(\widehat{Q}(x) - Q(x; \hat{\theta}_n)) \rightsquigarrow^{l^\infty(\mathcal{F})} x \mapsto \mathbb{G}_Q I(u \leqslant x) - \mathbb{G}_Q \psi_{\theta_0}^T \frac{\partial Q(x, \theta)}{\partial \theta}\Big|_{\theta=\theta_0}, \tag{19}$$

using the strategy discussed in Van der Vaart (2000), pp. 278–279. Here the notation $\leadsto^{l^\infty}$ 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 (18) 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

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

and using the continuous mapping theorem we get

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

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, (19) 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. $\square$

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

$$
\begin{aligned}
&T_0 \\
={}& nh^{1/2} \int (\hat{q}(x) - K_h * q(x; \hat{\theta}_n))^2 \mathrm{d}x \\
={}& nh^{1/2} \int (\hat{q}(x) - K_h * q_1(x))^2 \mathrm{d}x + nh^{1/2} \int (K_h * q_1(x) - K_h * q(x; \hat{\theta}_n))^2 \mathrm{d}x \\
&+ 2nh^{1/2} \int (\hat{q}(x) - K_h * q_1(x))(K_h * q_1(x) - K_h * q(x; \hat{\theta}_n))\mathrm{d}x
\end{aligned}
$$

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

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

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 pseudotrue model. The limit exists because we are considering functions in 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-Schwarz 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 $(T_0 - \int_K^2(u)\mathrm{d}u)/\sigma \to +\infty$ under $\mathcal{H}_1$ and the the claimed results for T_0 is true. The consistency of the test T_1 follows easily. $\square$

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 $\hat{F}(x)$ to be the empirical distribution function of the sample $\{X_i\}_{i=1}^n$. Then under the alternative hypothesis, we can make the following decomposition

$$
\begin{aligned}
&T_2 \\
&= n \int (\hat{Q}(x) - Q(x; \hat{\theta}_n))^2 \mathrm{d}Q(x; \hat{\theta}_n) \\
&= n \int (\hat{Q}(x) - \hat{F}(x))^2 \mathrm{d}Q(x; \hat{\theta}_n) + n \int (\hat{F}(x) - Q(x; \hat{\theta}_n))^2 \mathrm{d}Q(x; \hat{\theta}_n) \\
&\quad + 2n \int (\hat{Q}(x) - \hat{F}(x))(\hat{F}(x) - Q(x; \hat{\theta}_n)) \mathrm{d}\mathrm{d}Q(x; \hat{\theta}_n).
\end{aligned}
$$

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

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

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-Schwarz inequality the cross product term is $O_p(n^{1/2})$. Then it is obvious that $T_2 \to \infty$ under $\mathcal{H}_1$ when $n \to \infty$ and the test is consistent. $\qquad \square$

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.

Andersen, T. G., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time equity return models. *The Journal of Finance 57*(3), 1239–1284.

Andersen, T. G. and T. Bollerslev (1998). Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance 53*(1), 219–265.

Andrews, D. W. K. (1997). A conditional Kolmogorov test. *Econometrica 65*, 1097–1128.

Andrews, D. W. K. (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.

Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *The Journal of Finance 52*(5), 2003–2049.

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.

Bates, D. S. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics 94*(1), 181–238.

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.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics 108*(2), 281–316.

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.

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica 68*(6), 1343–1376.

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

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics 19*(2), 177–191.

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of Finance 59*(3), 1367–1403.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *The Journal of Finance 58*(3), 1269–1300.

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 (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli 5*, 855–872.

Genon-Catalot, V., T. Jeantheau, and C. Laredo (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli 6*, 1051–1079.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies 6*, 327–43.

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.

Jacquier, E., N. G. Polson, and P. E. Rossi (2002). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics 20*(1), 69–87.

Jacquier, E., N. G. Polson, and P. E. Rossi (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics 122*(1), 185–212.

Jensen, M. J. and J. M. Maheu (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics 157*(2), 306–316.

Kleijn, B. and A. W. van der Vaart (2012). The Bernstein-von-Mises theorem under misspecification. *Electronic Journal of Statistics 6*, 354–381.

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

Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses*. Berlin: Springer.

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*.

Müller, U. K. (2013). Risk of Bayesian inference in misspecified models, and the sandwich

covariance matrix. *Econometrica 81*(5), 1805–1849.

Neumann, M. H. and E. Paparoditis (2000). On bootstrapping L2-type statistics in density testing. *Statistics & Probability Letters 50*(2), 137–147.

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

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics 63*(1), 3–50.

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.

Song, Z. and D. Xiu (2016). A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics 190*(1), 176–196.

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

Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference.* Springer.

Yu, J. and R. Meyer (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews 25*(2-3), 361–384.

Zu, Y. (2015). Nonparametric specification tests for stochastic volatility models based on volatility density. *Journal of Econometrics 187*(1), 323–344.