Supplementary materials for this article are available at http://pubs.amstat.org/toc/jasa/104/487.

Nonparametric Transition-Based Tests for Jump Diffusions

Yacine AÏT-SAHALIA, Jianqing FAN, and Heng PENG

We develop a specification test for the transition density of a discretely sampled continuous-time jump-diffusion process, based on a comparison of a nonparametric estimate of the transition density or distribution function with their corresponding parametric counterparts assumed by the null hypothesis. As a special case, our method applies to pure diffusions. We provide a direct comparison of the two densities for an arbitrary specification of the null parametric model using three different discrepancy measures between the null and alternative transition density and distribution functions. We establish the asymptotic null distributions of proposed test statistics and compute their power functions. We investigate the finite-sample properties through simulations and compare them with those of other tests. This article has supplementary material online.

KEY WORDS: Generalized likelihood ratio test; Jump diffusion; Local linear fit; Markovian process; Null distribution; Specification test; Transition density.

1. INTRODUCTION

Consider a given parameterization for a jump diffusion defined on a probability space $(\Omega, \mathcal{F}, P)$

$$dX_t = \mu(X_{t-}, \theta)\, dt + \sigma(X_{t-}, \theta)\, dW_t + J_{t-}\, dN_t, \tag{1.1}$$

where W_t is a Brownian motion, N_t is a Poisson process with stochastic intensity $\lambda(X_{t-}; \theta)$ and jump size 1, and J_{t-}, the jump size, is a random variable with density $\nu(\cdot; X_{t-}, \theta)$. These dynamics are parameterized by a vector $\theta \in \Theta$, where Θ is a compact subset of $\mathbb{R}^K$. We are interested in testing the joint parametric family

$$\mathcal{P} \equiv \{(\mu(\cdot, \theta), \sigma^2(\cdot, \theta), \lambda(\cdot, \theta), \nu(\cdot; \cdot, \theta)) | \theta \in \Theta\}, \tag{1.2}$$

where Θ is a compact subset of $\mathbb{R}^K$.

The parametric family of models (1.1) provides explanatory power for understanding the underlying dynamics. A traditional parametric testing approach involves embedding the model (1.1) into a larger family of parametric models and using this new family as the alternative model. The question is then whether the larger family is wide enough to capture the true underlying dynamics. This leads us naturally to consider a nonparametric model as an alternative:

$$dX_t = \mu(X_{t-})\, dt + \sigma(X_{t-})\, dW_t + J_{t-}\, dN_t, \tag{1.3}$$

where the intensity of N and density of J also are nonparametric, $\lambda(X_{t-})$ and $\nu(\cdot; X_{t-})$.

If we believe that the true process is a jump diffusion with local characteristics $(\mu, \sigma^2, \lambda, \nu)$, then a specification test asks whether there are values of the parameters in Θ for which the parametric model $\mathcal{P}$ is an acceptable representation of the true process—that is, do the functions $(\mu, \sigma^2, \lambda, \nu)$ belong to the parametric family $\mathcal{P}$? Direct estimation of the local characteristics of the process with discrete data is problematic and can sometimes lead to inconsistent estimates. In contrast, every parameterization $\mathcal{P}$ corresponds to a parameterization of the marginal π and transition densities p of the process X:

$$\{(\pi(\cdot, \theta), p(\cdot|\cdot, \theta))| \\ (\mu(\cdot, \theta), \sigma^2(\cdot, \theta), \lambda(\cdot, \theta), \nu(\cdot; \cdot, \theta)) \in \mathcal{P}, \theta \in \Theta\}. \tag{1.4}$$

Whereas estimation of the densities explicitly takes into account the discreteness of the data, the main problem with testing the model (1.1) through (1.4) is that parametric expressions for the transition densities under the null model generally are unknown in closed form.

In this article we exploit recent developments in the field to develop a specification test for the transition density of the process, based on a direct comparison of a nonparametric estimate of the transition function and a closed-form expansion of the parametric transition density. The null and alternative hypotheses are then of the form

$$H_0: p(y|x) = p(y|x, \theta) \quad \text{vs.} \\ H_1: p(y|x) \neq p(y|x, \theta). \tag{1.5}$$

Here we directly compare the parametrically and nonparametrically estimated transition densities and distributions. By focusing directly on the transition density p, we concentrate on an object that plays a central role in financial statistics in applications as diverse as prediction, derivative pricing or pricing of kernels (after possibly a change of measure), risk management (through the magnitude of the tails of p), and portfolio choice (through the definition of the investment opportunity set).

Testing the specification of continuous-time models has been an active area of research in recent years, although the literature has focused on the purely diffusive case, thereby excluding jumps, and on univariate models. Aït-Sahalia (1996) proposed two tests for pure diffusions, one based on the marginal density

Yacine Aït-Sahalia is Otto A. Hack 1903 Professor of Finance and Economics, Department of Economics, Princeton University, Princeton, NJ 08544 (E-mail: *yacine@princeton.edu*). Jianqing Fan is Frederick L. Moore Professor of Finance and Statistics, Department of Operation Research and Financial Engineering, Princeton University, Princeton, NJ 08544 and Honorary Professor, Shanghai University of Finance and Economics, Shanghai, China (E-mail: *jqfan@princeton.edu*). Heng Peng is Assistant Professor, Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong (E-mail: *hpeng@math.hkbu.edu.hk*). Aït-Sahalia's research is supported by National Science Foundation (NSF) grants DMS-0532370 and SES-0350772. Fan's research is supported by NSF grants DMS-0532370, DMS-0355179, and DMS-0354223. Peng's research is supported by a CERG grant from the Hong Kong Research Grant Council (HKBU 201707), a grant from Hong Kong Baptist University (FRG/06-07/II-14), and a grant from the National Nature Science Foundation of China (NNSF 10871054).

Journal of the American Statistical Association
September 2009, Vol. 104, No. 487, Theory and Methods
DOI: 10.1198/jasa.2009.tm08198

π and the other based on the transition density p, and derived their asymptotic distributions. The basic idea of these tests is to use the mapping between the drift and diffusion on the one hand, and the marginal and transitional densities on the other hand, to test the model's specification using densities at the observed discrete frequency instead of the infinitesimal characteristics of the process (μ, σ^2). Chen and Gao (2007) and Thompson (2008) proposed tests based on the empirical likelihood method. Chen, Gao, and Tang (2008) applied an integrated empirical likelihood method, addressed the issue of bandwidth selection, and derived the asymptotic null distribution. Fan and Zhang (2003) restricted the alternative models to the class of nonparametric univariate pure diffusion models (1.3) and tested the drift and diffusion functions separately. Other contributions in closely related topics include those of Andrews (1997), Wang (2002), Altissimo and Mele (2009), and Li and Tkacz (2006).

Related to the present article is the approach proposed by Hong and Li (2005) in the case of univariate diffusions. Those authors made use of the fact that under the null hypothesis, the random variables $\{P(X_i|X_{i-\Delta}, \theta)\}$ are a sequence of i.i.d. uniform random variables; they then detected departures from the null hypothesis by comparing the kernel-estimated bivariate density of $\{(Z_i, Z_{i+\Delta})\}$ with that of the uniform distribution on the unit square, where $Z_i = P(X_i|X_{i-\Delta}, \theta)$ and $P(\cdot|\cdot)$ is the cumulative transition distribution.

The article is organized as follows. In Sections 2 and 3 we describe the construction of the two estimators of p, nonparametric and parametric. In Section 4 we introduce the three test statistics that we use to compare these two estimators. In Section 5 we derive the asymptotic properties of these statistics, including their distributions under the null hypothesis and their power properties. In particular, we show that the asymptotic null distributions of our test statistics are asymptotically distribution-free. In Section 6 we examine the corresponding finite-sample properties of the statistics, including their respective power, under various models, and compare the power of transition density–based tests with that of other tests. In Section 7 we test various models on two classical time series in finance. We conclude in Section 8.

2. NONPARAMETRIC ESTIMATION OF THE TRANSITION DENSITY

Suppose that the observed process $\{X_t\}$ is sampled at the regular time points $\{i\Delta, i = 1, \ldots, n+1\}$. We make the dependence on the transition function and related quantities on Δ implicit by redefining

$$X_i = X_{i\Delta}, \qquad i = 1, \ldots, n+1,$$

which is assumed to be a stationary and β-mixing process. Let $p(y|x)$ be the transition density of the series $\{X_i, i = 1, \ldots, n+1\}$; that is, it is the conditional density of X_{i+1} given $X_i = x$ evaluated at $X_{i+1} = y$. This conditional density can be estimated by the local linear method of Fan, Yao, and Tong (1996). We briefly summarize this method and discuss the issues related to bandwidth selection.

Let h_1 and h_2 be two bandwidths and let K and W be two kernel functions. The conditional density $p(y|x)$ is approximately the regression function of $K_{h_2}(X_{i+1} - y)$ given $X_i = x$ for small h_2, where $K_h(z) = K(z/h)/h$. Using the local linear fit, for each given x, we minimize

$$\sum_{i=1}^{n} \{K_{h_2}(X_{i+1} - y) - \alpha - \beta(X_i - x)\}^2 W_{h_1}(X_i - x) \tag{2.1}$$

with respect to the the local parameters α and β. The resulting estimate of the conditional density is simply $\hat{\alpha}$. The estimator can be expressed explicitly as

$$\hat{p}(y|x) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} W_n\left(\frac{X_i - x}{h_1}; x\right) K\left(\frac{X_{i+1} - y}{h_2}\right), \tag{2.2}$$

where W_n is the effective kernel induced by the local linear fit. Explicitly, this is given by

$$W_n(z; x) = W(z) \frac{s_{n,2}(x) - z s_{n,1}(x)}{s_{n,0}(x) s_{n,2}(x) - s_{n,1}(x)^2},$$

where

$$s_{n,j}(x) = \frac{1}{nh_1} \sum_{i=1}^{n} \left(\frac{X_i - x}{h_1}\right)^j W\left(\frac{X_i - x}{h_1}\right).$$

Note that the effective kernel W_n depends on the sampling data points and the location x. This is the key to the design adaptation and location adaptation property of the local linear fit (Fan 1992).

A possible estimate of the transition distribution $P(y|x) = P(X_{i+1} < y | X_i = x)$ is given by

$$\hat{P}(y|x) = \frac{1}{nh_1} \sum_{i=1}^{n} W_n\left(\frac{X_i - x}{h_1}; x\right) I(X_{i+1} < y), \tag{2.3}$$

which is the local linear estimator of the regression function

$$P(y|x) = E\{I(X_{i+1} < y) | X_i = x\}.$$

More complicated estimation schemes of the conditional distribution based on a local logistic regression and an adjusted form of the kernel regression that guarantee the estimated value in the interval [0, 1] have been given by Hall, Wolff, and Yao (1999). Note that the estimator (2.3) is also the cumulative distribution of the conditional density $\hat{p}(y|x)$ with $h_2 \to 0$.

As with most nonparametric estimation procedures, issues of bandwidth selection arise in practice. Fan and Yim (2004) and Hall, Racine, and Li (2004) suggested using CV to select bandwidths for estimating the conditional density. We did so in our numerical studies. For estimation of the cumulative transition distribution, the problem is equivalent to the nonparametric regression method, and thus a wealth of bandwidth selectors are available. In our numerical studies, we used the plug-in method of Ruppert, Sheather, and Wand (1995).

3. THE PARAMETRIC TRANSITION DENSITY

We have defined a consistent nonparametric estimator of the transition density $p(y|x)$. Under the null hypothesis, however, it also is possible to specify and estimate separately the parametric transition density corresponding to the assumed parametric model. We need a parametric form for the transition density not only to compare with the nonparametric estimator, but also to deliver a root-n–consistent estimator of θ (in the form of the maximum likelihood estimator [MLE]).

For this purpose, we rely on the new closed-form expansions developed in Aït-Sahalia (2002) and Aït-Sahalia (2008) for pure diffusions and extended to jump diffusions by Yu (2007). By the Bayes rule, we have

$$p(y|x;\theta) = \sum_{n=0}^{+\infty} p(y|x, N_\Delta = n;\theta)\Pr(N_\Delta = n|x;\theta). \tag{3.1}$$

With $\Pr(N_\Delta = 0|x;\theta) = O(1)$, $\Pr(N_\Delta = 1|x;\theta) = O(\Delta)$, and $\Pr(N_\Delta > 1|x;\theta) = o(\Delta)$, and given the fact that when at least one jump occurs the dominant effect is due to the jump (vs. the increment due to the Brownian motion), an expansion at order K in Δ of p obtained by extending the pure diffusive result to jump diffusions is given by

$$\begin{aligned}\tilde{p}^{(K)}&(y|x;\theta)\\ &= \exp\left(-\frac{1}{2}\ln(2\pi\Delta\sigma^2(y;\theta)) + \frac{c_{-1}(y|x;\theta)}{\Delta}\right)\\ &\quad\times \sum_{k=0}^{K} c_k(y|x;\theta)\frac{\Delta^k}{k!} + \sum_{k=1}^{K} d_k(y|x;\theta)\frac{\Delta^k}{k!}.\end{aligned} \tag{3.2}$$

The unknowns are the coefficients c_k and d_k of the series. Relative to the pure diffusive case, the coefficients d_k are the new terms needed to capture the presence of the jumps in the transition function and will capture the different behavior of the tails of the transition density when jumps are present.

The coefficients c_k and d_k can be computed analogously to the pure diffusive case, resulting in a system of equations that can be solved in closed form, starting with

$$c_{-1}(y|x;\theta) = \frac{1}{2}\left(\int_x^y \frac{du}{\sigma(u;\theta)}\right)^2, \tag{3.3}$$

$$\begin{aligned}c_0(y|x;\theta) = \frac{1}{\sqrt{2\pi}\sigma(y;\theta)}\exp\biggl(&\int_x^y \frac{\mu(u;\theta)}{\sigma^2(u;\theta)}\,du\\ &- \int_x^y \frac{\partial\sigma(u;\theta)/\partial u}{2\sigma(u;\theta)}\,du\biggr).\end{aligned} \tag{3.4}$$

Coefficients of higher order of the diffusive part of the expansion (i.e., c_k, $k \geq 1$) are no longer functions of the diffusive characteristics of the process only; instead, they also involve the characteristics of the jump part. In particular, for $k = 1$,

$$\begin{aligned}c_1(y|x;\theta) = -\left(\int_x^{y_t} \frac{du}{\sigma(u;\theta)}\right)^{-1} &\int_x^y \biggl\{\frac{du}{\sigma(u;\theta)}\\ &\times \exp\left(\int_x^s \frac{\mu(u;\theta)}{\sigma^2(u;\theta)}\,du - \int_x^{y_t}\frac{\partial\sigma(u;\theta)/\partial u}{2\sigma(u;\theta)}\,du\right)\\ &\times \left(\int_s^y \frac{du}{\sigma(u;\theta)}\right)(\lambda(s;\theta) - \mathcal{A}\cdot c_0(y|s;\theta))\biggr\}\\ &\times \frac{ds}{\sigma(s;\theta)},\end{aligned} \tag{3.5}$$

where the operator $\mathcal{A}$ is the generator of the diffusive part of the process only, defined by its action,

$$\mathcal{A}\bullet f = \mu\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}, \tag{3.6}$$

on functions in its domain.

The leading term in the jump part of the expansion (i.e., d_1) is given by

$$d_1(y|x;\theta) = \lambda(x;\theta)\nu(y-x;\theta). \tag{3.7}$$

As in the pure diffusive case, higher-order terms d_k and $k \geq 2$ are obtained recursively from the preceding ones (see Yu 2007).

The approximation error created by replacing the exact (but unknown) $p(y|x;\theta)$ with $\tilde{p}^{(K)}(y|x;\theta)$ is of order $O(\Delta^K)$. Our asymptotic scheme assumptions are such that this error is always smaller than the sampling error introduced by estimation of p. As a practical matter, Jensen and Poulsen (2002), Stramer and Yan (2007), and Hurn, Jeisman, and Lindsay (2007) conducted extensive comparisons of different techniques for approximating the transition function p and demonstrated that the method described here is both the most accurate and the most rapidly implemented for the types of problems and sampling frequencies typically encountered in finance (monthly, weekly, daily, or higher). The relative approximation error is often <0.001, a level that in practice is negligible compared with the n-dependent sampling error in the parametric estimation, and even more so the nonparametric estimation, of p.

4. TESTING AGAINST A JUMP–DIFFUSION MODEL

We now consider testing against the parametric family corresponding to model (1.1). Let $p(y|x;\theta)$ be the transition density of the sequence $\{X_i, i = 1,\dots,n+1\}$ induced by the model (1.1). When the sampling interval Δ is a constant, the maximum likelihood estimator $\hat{\theta}$ obtained by maximizing the parametric log-likelihood given in Section 3 converges at rate $O(n^{-1/2})$ in the stationary case under mild conditions (see Aït-Sahalia 2008 and Yu 2007). (The rate can be different if the data-generating process is nonstationary.) This yields a parametric estimate of the transition density $p(y|x;\hat{\theta})$, computed using the techniques described in Section 3. The effect of the random and discrete sampling when estimating continuous-time diffusions has been studied by Aït-Sahalia and Mykland (2003) and Aït-Sahalia and Mykland (2004).

Our approach to testing against the null hypothesis (1.1) is to compare the differences between parametric and nonparametric estimates of p as spelled out by (1.5). The log-likelihood function of the observed data $\{X_1,\dots,X_{n+1}\}$ is

$$\ell(p) = \sum_{i=1}^n \log p(X_{i+1}|X_i),$$

after ignoring the stationary density $\pi(X_1)$. A natural test statistic or metric is a comparison of the likelihood ratio under the null hypothesis and the alternative hypothesis. This leads to the test statistic

$$\sum_{i=1}^n \log\hat{p}(X_{i+1}|X_i)/p(X_{i+1}|X_i,\hat{\theta}).$$

Note that this is not the maximum likelihood ratio test, because the nonparametric estimate $\hat{p}$ is not derived from the maximum likelihood estimate (Fan, Zhang, and Zhang 2001). This type of test may be called a generalized likelihood ratio test. It is well known that the nonparametric regression function cannot be well estimated when X_i is in the boundary region. Thus we introduce a weight function, w, to reduce the influences of the unreliable estimates, leading to the test statistic

$$T_0 = \sum_{i=1}^n \log\{\hat{p}(X_{i+1}|X_i)/p(X_{i+1}|X_i,\hat{\theta})\}w(X_i, X_{i+1}). \tag{4.1}$$

We start by explaining some intuition to justify our construction of the test statistics. Under the null hypothesis of (1.5), the parametric and nonparametric estimators are approximately the same. Heuristically, the null distribution of T_0 is obtained by a Taylor expansion,

$$T_0 \approx \sum_{i=1}^{n} \frac{\hat{p}(X_{i+1}|X_i) - p(X_{i+1}|X_i, \hat{\theta})}{p(X_{i+1}|X_i, \hat{\theta})} w(X_i, X_{i+1}) - \frac{1}{2}\sum_{i=1}^{n} \left\{ \frac{\hat{p}(X_{i+1}|X_i) - p(X_{i+1}|X_i, \hat{\theta})}{p(X_{i+1}|X_i, \hat{\theta})} \right\}^2 w(X_i, X_{i+1}).$$

Because the nonparametric locally linear estimator $\hat{p}$ is not an MLE, whether or not the first term is asymptotically negligible is not obvious. To avoid unnecessary technicalities, we consider the following χ^2-test statistic:

$$T_1 = \sum_{i=1}^{n} \left\{ \frac{\hat{p}(X_{i+1}|X_i) - p(X_{i+1}|X_i, \hat{\theta})}{p(X_{i+1}|X_i, \hat{\theta})} \right\}^2 w(X_i, X_{i+1}). \quad (4.2)$$

A natural alternative method to T_1 is

$$T_2 = \sum_{i=1}^{n} \{\hat{p}(X_{i+1}|X_i) - p(X_{i+1}|X_i, \hat{\theta})\}^2 w(X_i, X_{i+1}). \quad (4.3)$$

The transition density–based test statistics depend on two smoothing parameters h_1 and h_2 and thus are somewhat more cumbersome to implement. Like the Cramer–von Mises test, a viable alternative method is to compare the discrepancies between transition distributions. This leads to the test statistic

$$T_3 = \sum_{i=1}^{n} \{\hat{P}(X_{i+1}|X_i) - P(X_{i+1}|X_i, \hat{\theta})\}^2 w(X_i, X_{i+1}). \quad (4.4)$$

The biases in the asymptotic null distribution can be reduced by incorporating a bias-reduction technique (Fan and Zhang 2004).

For nonparametric testing problems, there generally is no uniformly most powerful test. Each test can be powerful for detecting certain classes of alternatives. The density-based tests T_1 and T_2 are more powerful in detecting fine features, such as sharp and short aberrants, that deviate from the density under the null hypothesis, whereas T_3 is more powerful for detecting global departures from the null hypothesis (e.g., when the density has been shifted or rescaled).

5. ASYMPTOTIC PROPERTIES OF THE TEST STATISTICS

5.1 Asymptotic Null Distributions

Throughout this article, we let $\|f\|^2 = \int f^2(x)\,dx$, $\|f\|_w^2 = \int f^2(x)w(x)\,dx$, and "$*$" denote the convolution operation. Thus $\|\mathbf{1}\|_w = \int w(x,y)\,dx\,dy$. Furthermore, the notation $rT_n \overset{a}{\sim} \chi^2_{a_n}$ means that

$$(2a_n)^{-1/2}\{rT_n - a_n(1 + o(1))\} \xrightarrow{\mathcal{D}} N(0, 1).$$

Theorem 1. Under Conditions 1–8 in Appendix A, if the transition density of the observed data follows from $p(y|x;\theta)$, then we have

$$\frac{1}{\sigma_1}\{T_1 - \mu_1\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where $\mu_1 = \|\mathbf{1}\|_w \|W\|^2 \|K\|^2/(h_1 h_2) - \Omega_x \|W\|^2/h_1$, $\sigma_1^2 = 2\|w\|^2 \|W * W\|^2 \|K * K\|^2/(h_1 h_2)$, and $\Omega_x = \int \mathrm{E}\{w(X,Y)|X = x\}\,dx$. In other words, $r_1 T_1 \overset{a}{\sim} \chi^2_{a_n}$, where $a_n = r_1\mu_1$ and $r_1 = \|\mathbf{1}\|_w \|W\|^2 \|K\|^2/(\|w\|^2 \|W * W\|^2 \|K * K\|^2)$.

Note that the effective number of parameters under the alternative is of $O(1/(h_1 h_2))$, corresponding to partitioning of a rectangle using subrectangles of length h_1 and width h_2, whereas the number of parameters under the null hypothesis is finite. Thus the order of the degrees of freedom is given by $O(1/(h_1 h_2))$, the same order as a_n.

The test statistic T_2 can be considered a special case of T_1 with $w(x,y)$ as $p(y|x;\theta)w(x,y)$ in Theorem 1. Thus an application of Theorem 1 readily yields the asymptotic null distribution of T_2. Unlike the distribution of T_1, the distribution of T_2 depends on nuisance parameters under H_0; that is, the Wilks phenomenon does not hold in this case.

Next, we consider the asymptotic null distribution of T_3. Because $\hat{P}(y|x)$ is a nonparametric estimate of the conditional distribution function, we need only weigh down the contribution from the sparse regions in the x-coordinate. For this reason, we consider only the weight function $w(x,y) = w(x)$ for T_3. This allows us to evaluate the asymptotic mean and variance more explicitly.

Theorem 2. When the observed data are realizations from a stationary Markovian process with transition density $p(y|x;\theta)$,

$$\frac{1}{\sigma_3}\{T_3 - \mu_3\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

provided that Conditions 1–7 and 9 in the Appendix A hold. Here

$$\mu_3 = \|W\|^2 \int w(t)\,dt/(6h_1) \qquad \text{and}$$

$$\sigma_3^2 = \|W * W\|^2 \|w\|^2/(45h_1).$$

Namely, $r_3 T_3 \overset{a}{\sim} \chi^2_{b_n}$, where $b_n = r_3\mu_3$ and $r_3 = 15\|W\|^2\|\mathbf{1}\|_w^2/(\|W * W\|^2\|w\|^2)$.

The statistic T_3 is similar to the Cramér–von Mises test T_{CVM}. It is well known that

$$\mathrm{E}(T_{\mathrm{CVM}}) = \frac{1}{6}, \qquad \mathrm{Var}(T_{\mathrm{CVM}}) = \frac{4n-3}{180n} \approx \frac{1}{45}.$$

These results are compatible with the asymptotic mean and variance of T_3 (comparing in particular the numerical coefficients 1/6 and 1/45). Anderson and Darling (1952) derived the limit distribution of T_{CVM}, which is not normally distributed. This differs from Theorem 2, because the aggregation of many local T_{CVM} results in an asymptotic normal distribution.

Comparing Theorems 1 and 2 shows that the asymptotic variance of T_1 is an order of magnitude larger than that of T_3. Because the transition density–based tests need to localize in two directions, much fewer local data points are available for estimating the transition density than for estimating the transition distribution. Thus the null distribution of T_3 should be more stably approximated compared with the null distributions of T_1 and T_2. On the other hand, because T_1 and T_2 have greater degrees of freedom, the transition density–based tests are more omnibus, designed to detect a wider range of alternative models.

5.2 Bootstrap

The asymptotic null distributions depend on many smaller-order terms and are negligible when nh_1h_2 is large. For example, any constant can be added to the degree of freedom a_n or any other numbers of order $o(1/h_1)$, but in practical applications, the convergence is slow, because the local sample size nh_1h_2 cannot be large enough. 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 asymptotic normality in Theorems 1 and 2 justifies using the bootstrap.

Let θ_0 be the parameter that the data-generating process projects on the family of models under the null hypothesis. If θ_0 is given, then the null distribution is known and can be simulated. The bootstrap estimate of the null distribution mimics this process, except that θ_0 is replaced by its estimate $\hat{\theta}_n$ from the sample. Thanks to Theorems 1 and 2, the asymptotic null distribution does not depend on θ_0. This implies that the null distributions under θ_0 and $\hat{\theta}_n$ should be very close. The bootstrap distribution should give a better approximation than the asymptotic null distribution, because it uses the null structure. Our asymptotic theory gives the justification of this process. As we show in Section 6.2, the bootstrapped distributions are much closer than the asymptotic ones.

5.3 Power Under Contiguous Alternatives

To compute the power function, we consider the contiguous alternatives H_{1n}

$$p_n(y|x) = p(y|x;\theta) + g_n(x,y)$$

for the test statistic T_1 and

$$P_n(y|x) = P(y|x;\theta) + G_n(x,y)$$

for the test statistic T_3. To ease the technical arguments and presentation, we assume that $g_n(x,y) = \delta_n g(x,y) + o(\delta_n)$ and $G_n(x,y) = \delta_n G(x,y) + o(\delta_n)$, where $\delta_n^{-1} g_n(x,y)$ and $\delta_n^{-1} G_n(x,y)$ have bounded continuous second-order derivatives. These assumptions can be relaxed to allow a more general form of the alternatives.

Theorem 3. Under Conditions 1–6 in Appendix A, if $nh_1 \times h_2\delta_n^2 = O(1)$, then, under the alternative hypothesis H_{1n}, we have

$$(T_1 - \mu_1 - d_{1n})/\sigma_1^* \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

where

$$d_{1n} = n\mathrm{E}\frac{g_n^2(X,Y)w(X,Y)}{p^2(Y|X;\theta)} + O(nh_1^2\delta_n + nh_2^2\delta_n + \delta_n h_1^{-1}h_2^{-1})$$

and $\sigma_1^* = \sqrt{\sigma_1^2 + 4\sigma_{1A}^2}$ with

$$\sigma_{1A}^2 = n\mathrm{E}\frac{g_n^2(X,Y)w^2(X,Y)}{p^2(Y|X)} - n\mathrm{E}\left\{\mathrm{E}\left(\frac{g_n(X,Y)w(X,Y)}{p(Y|X)}\Big|X\right)\right\}^2.$$

The condition $nh_1h_2\delta_n^2 = O(1)$ is imposed so that the test T_1 has a nontrivial power. If $nh_1h_2\delta_n^2 = o(1)$, then it is easy to see that the asymptotic mean and variance of T_1 are dominated by μ_1 and σ_1^2, respectively, and the test statistic behaves the same under the null alternative hypothesis. Thus the test has no power to detect the alternative hypothesis. On the other hand, if $nh_1h_2\delta_n^2 \to \infty$, then the asymptotic mean and variance of T_1 are dominated by d_{n1} and σ_{1A}^2, respectively. It is then easy to see that the asymptotic power is 1.

Because T_2 is a specific example of T_1, we do not discuss its asymptotic distribution under the alternative hypothesis. We now consider the rate at which the alternative can be detected by the test statistic T_1. This amounts to determining that rate δ_n so that T_1 is stochastically larger under the alternative hypothesis than under the null hypothesis.

Theorem 4. Under Conditions 1–8 in Appendix A, T_1 can detect alternatives with rate $\delta_n = O(n^{-2/5})$ when $h_1 = c_1^* n^{-1/5}$ and $h_2 = c_2^* n^{-1/5}$. More precisely, if $\delta_n = dn^{-2/5}$ for some constant d, then we have the following properties for the power function:

$$\limsup_{d\to 0}\limsup_{n\to\infty} P\{(T_1 - \mu_1)/\sigma_1 > c_\alpha | H_{1n}\} \le \alpha$$

and

$$\liminf_{d\to\infty}\liminf_{n\to\infty} P\{(T_1 - \mu_1)/\sigma_1 > c_\alpha | H_{1n}\} = 1.$$

We next consider the asymptotic alternative distribution of T_3 and the rate at which it can be detected by the test statistic T_3.

Theorem 5. Under Conditions 1–6 and 9 in Appendix A, if $nh_1\delta_n^2 = O(1)$, then, under H_{1n} for the test problem, we have

$$(T_3 - \mu_3 - d_{3n})/\sigma_3^* \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

where $d_{3n} = n\mathrm{E}G_n^2(X,Y)w(X) + O\{nh_1^2\delta_n + \delta_n h_1^{-1}\}$, $\sigma_3^* = \sqrt{\sigma_3^2 + 4\sigma_{3A}^2}$, and

$$\sigma_{3A}^2 = n\mathrm{E}\left\{\int G_n(X_i,Y_j)w(X_i)I(Y_i < Y_j)\,dP(Y_j|X_i)\right\}^2 - n\mathrm{E}\left\{\int G_n(X_i,Y_j)w(X_i)P(Y_j|X_i)\,dP(Y_j|X_i)\right\}^2.$$

Using exactly the same argument as that of Theorem 4, we can obtain the following theorem. We omit the details of the proof.

Theorem 6. Under Conditions 1–6 in Appendix A, the test statistic T_3 can detect alternatives with rate $\delta_n = O(n^{-4/9})$ when $h_1 = c_* n^{-2/9}$.

From Theorem 4, T_1 and T_2 can detect the alternative with the rate $\delta_n = O(n^{-2/5})$, and they depend on both h_1 and h_2. For the transition distribution–based test, the rate that can be detected by T_3 is $O(n^{-4/9})$, which is optimal according to Ingster (1993) and Lepski and Spokoiny (1999). This is due to the fact that the alternative under consideration is global; namely, the density under the alternative is basically globally shifted away from the null hypothesis. In contrast, the conditional density–based tests are more powerful for detecting local features than those based on the conditional distribution. To see this, assume that the deviation in the conditional density is $g_n(x,y) = \delta_n g_1(x) g_2(|y|/a_n) + o(\delta_n a_n)$ for a sequence of $a_n \to 0$, where g_2 is a symmetric density function. The deviation $g_n(x,y)$ is rather local in the y-axis. Then the deviation on the conditional distribution is given by $G_n(x, y) = \delta_n a_n g_1(x)\,\text{ind}(y)$, an order smaller than that in g_n, where $\text{ind}(y) = 0$, $1/2$, or 1 depending on whether $y < 0$, $y = 0$, or $y > 0$. The main term d_{1n} induced by the local alternative is of order $n\delta_n^2 a_n$, whereas the main term d_{3n} is of order $n\delta_n^2 a_n^2$, an order of magnitude smaller. Thus T_1 and T_2 are more powerful in detecting local features in the alternative hypothesis.

The test statistic T_3 is one step closer to the spirit of the conditional Kolmogorov (or Cramer–von Mises) test, which uses the cumulative conditional distribution and has parametric rates of convergence. But our method is more nonparametric in the sense that we localize in the x-direction and thus is more powerful in detecting local features in that direction.

6. MONTE CARLO SIMULATIONS

6.1 Specification of the Tests

We focus on the finite-sample performance of the transition density–based test T_1 and the transition distribution–based test T_3. We use the bootstrap method to determine the null distribution of the test statistic. To compare the power of various test statistics, we use the empirical quantiles under the null models to determine the critical values. This highlights the relative powers of the five test statistics under consideration.

By Theorem 4, the optimal rate of h_1 and h_2 is $O(n^{-1/5})$. This differs from the optimal rate $O(n^{-1/6})$ used to estimate the nonparametric conditional density by a factor of obly $n^{1/30}$. Thus the bandwidth selected by the cross-validation method or other methods for density estimation provides a reasonable proxy for the testing problems. This also allows us to examine whether the optimally estimated density is significantly different from the null hypothesis. In our numerical work, we use the cross-validation method of Fan and Yim (2004) to select h_1 and h_2. We use the plug-in method of Ruppert, Sheather, and Wand (1995) to select h_1 for the test statistic T_3.

For the test statistic T_3, we take $w(x) = I\{|x - \mu| < 1.5\sigma\}$, where μ is sample mean and σ is the sample standard deviation for the data simulated under the null hypothesis. The interval in this indicator function covers approximately 80% of the data. For the test T_1, we set $w(x, y) = I\{(|x - \mu| < 1.5\sigma) \cap |y - \mu(y|x)| < 1.5\sigma^*\}$, where $\mu(y|x)$ is the conditional mean and σ^* is the standard deviation of the differences $\{X_{i+1} - X_i\}$, computed for the data under the null hypothesis.

6.2 Accuracies of the Null Distributions

We simulate 1,000 sample paths of weekly observations ($\Delta = 1/52$) from the Vasicek model

$$dX_t = \kappa(\mu - X_t)\,dt + \sigma\,dW_t. \tag{6.1}$$

The parameter values $\kappa = 0.12$, $\mu = 0.06$, $\sigma = 0.013$ are taken from the weighted least squares estimates using the 3-month Treasury Bill data, comprising 2,400 weekly observations between January 8, 1954 and December 31, 1999.

With given parameters and sample sizes, the test statistics T_1 and T_3 are simulated 1,000 times. The distributions of these realized test statistics T_1 and T_3 can be obtained using the kernel density estimate and are considered the true distribution (except the Monte Carlo errors). To limit computation and facilitate the presentation, for each of the realized 1,000 sample paths, we obtain 5 bootstrap samples and compute their resulting test statistics T_1^* and T_3^*. Aggregating them together across 1,000 samples yields 5,000 bootstrap statistics. Their sampling distributions, computed via the kernel density estimate, is depicted as the distribution of the bootstrap method. We also present the normal distribution with mean and variance estimated from these 5,000 bootstrap statistics. This can be viewed as using the asymptotic null distributions with the first two moments calibrated using the bootstrap method.

Figure 1 summarizes the simulation results with $n = 600$, 1,200, and 2,400. It clearly shows that the bootstrapped distributions are much closer to the true ones than the asymptotic ones. In addition, the accuracies are generally acceptable for practical applications. On the other hand, as discussed in Section 5.2, the asymptotic distributions are quite far from the true ones. This is due to the ignorance of lower-order terms, which are not negligible at finite sample. The bootstrap methods can be viewed as an effort to account for these smaller-order terms, as evidenced by the two bootstrap methods in Figure 1. We also could empirically add the lower order terms to calibrate the asymptotic null distributions; for example, replacing μ_1 and σ_1^2 by

$$\mu_1' = \mu_1 + 20, \qquad \sigma_1'^2 = \sigma_1^2 + 40$$

and

$$\mu_3' = \mu_3 + 1.5, \qquad \sigma_3'^2 = \sigma_3^2 + 0.18,$$

the asymptotic distribution becomes much closer to the true one. These are shown as solid curves in Figure 1.

6.3 Power of the Tests

We compare the performance of our tests using the test of Hong and Li (2005) and its two variants. The test proposed by Hong and Li (2005) is based on the idea that $Z_t = P(X_t|X_{t-1}, \hat{\theta})$ is approximately a sequence of independent random samples from the uniform distribution. To test this, Hong and Li (2005) proposed the test statistic $\hat{M}(j)$, which compares a kernel estimator for the joint density of Z_t, Z_{t-j} with the density of the uniform distribution on the unit square, namely

$$\hat{M}(j) = \int_0^1 \int_0^1 [\hat{g}_j(z_1, z_2) - 1]^2\,dz_1\,dz_2,$$

where

$$\hat{g}_j(z_1, z_2) = (n - j)^{-1} \sum_{t=j+1}^{n} K_h(z_1, Z_t) K_h(z_2, Z_{t-j}),$$

where $K_h(z_1, z_2)$ is a boundary-modified kernel. The centered and scaled test statistic,

$$T_4 = [(n - j)h\hat{M}(j) - A_h^0]/V_0^{1/2}$$

and follows asymptotically a standard normal distribution for constants A_h and V_0. We take $j = 1$ in the implementation and follow the prescriptions of Hong and Li (2005) to select the bandwidths.

Useful alternative methods are the distribution-based tests for uniformity. We consider in particular the following two versions of the tests:

$$T_5 = \int_0^1 (\hat{F}_Z(x) - x)^2\,dx,$$

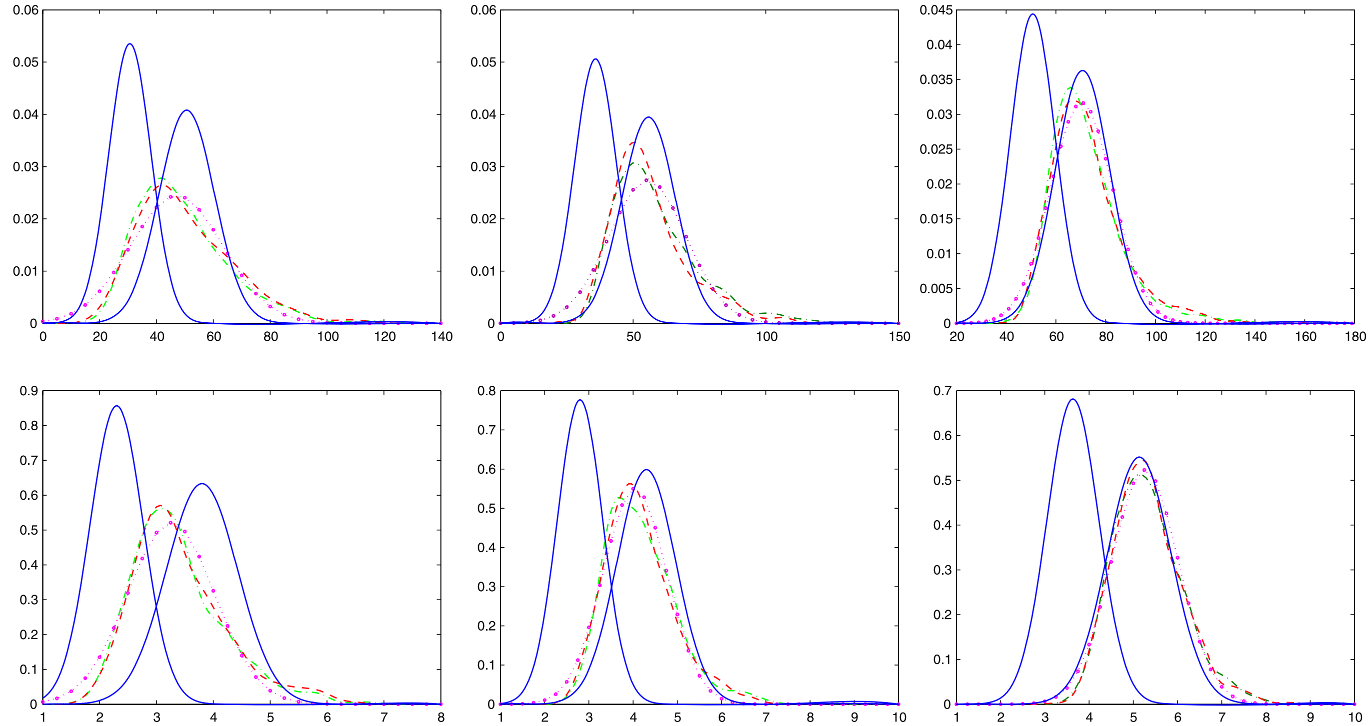

Figure 1. Null distributions of test statistics T_1 (top panel) and T_3 (bottom panel) for $n = 600$ (left), $n = 1{,}200$ (middle), and $n = 2{,}400$ (right). Solid line—asymptotic distributions without calibration (right) and with calibration (left); dash–dot (green) line—empirical distribution of the test statistic (true); dash (red) line—distribution by bootstrap; doted (magenta) line—normal distribution with the mean and variance obtained by bootstrap.

in which $\hat{F}_Z$ is the empirical cumulative distribution of $\{Z_t\}$, and

$$T_6 = \int_0^1 \int_0^1 \left(\hat{F}_{Z_{t-1},Z_t} - x_1 x_2\right)^2 dx_1\, dx_2,$$

in which $\hat{F}_{Z_{t-1},Z_t}$ is the empirical cumulative distribution based on the data $\{(Z_{t-1}, Z_t)\}$. These extensions of the Cramer–von Mises test have several advantages; they do not need to select bandwidth or deal with boundary effect. The asymptotic null distributions of these test statistics have been studied and are easily computed.

The test statistics T_0, T_1, T_2, and T_3 also can be applied to the transformed data $\{Z_t\}$. For example, an application of T_1 yields

$$T_7 = \sum_{i=1}^{n} \{\hat{p}(Z_{i+1}|Z_i) - 1\}^2 w(Z_{i+1}, Z_i)$$

where $\hat{p}(Z_{i+1}|Z_i)$ the nonparametric estimate of the conditional density based on $\{Z_t\}$.

Here we use a sample size of 2,400 (the same as the Treasury Bill data used to obtain the parameters) and 1,000 simulationhs. Alternatives include all stationary Markovian processes. This reduces the likelihood of model misspecification. To facilitate computation, we take the Vasicek model (6.1) as the null hypothesis. We evaluate the power functions of five competing test statistics under the three families of alternative models, progressively deviating away from the null. Each family of models is indexed by τ, with $\tau = 0$ corresponding to the null model (6.1).

In the first family of alternative models, the volatility functions are deviated from the Vasicek model. In the second example, the drift function is deviated from the Vasicek model. In the third example, the jump-diffusion processes are included. We now describe the alternative family of models.

Example 1. In this example we evaluate the power of the five tests at a sequence of alternative models,

$$dX_t = \kappa(\mu - X_t)\,dt + \{(1-\tau)\sigma + \tau\sigma(X_t)\}\,dW_t \qquad (6.2)$$

for $\tau = 0, 0.1, \ldots, 1$, where $\sigma(X_t) = \sigma^* X_t^{\gamma}$ with $\sigma^* = 0.07$ and $\gamma = 0.7$ from the parametric fit of the model of Chan et al. (1992) to the aforementioned Treasury Bill data. Note that we take the same instantaneous return function under the null hypothesis and the alternative hypothesis; this increases the degree of the difficulty of the tests.

Figure 2 shows the differences of the volatility functions between the null hypothesis and the alternative hypothesis. The power functions under three different significant levels, $\alpha = 0.01, 0.05, 0.1$, were computed. To save the space, only the results for $\alpha = 0.05$ are depicted in the figure. It is clear that T_1 and T_3 have much higher power than the other three tests. For this particular example, the transition density–based test T_1 is somewhat more powerful than the transition distribution–based test T_3.

Example 2. Here we consider the deviations in the drift. The null model is still the model of Vasicek (6.1). In the alternative model, the volatility function remains the same, but nonlinear mean revision is considered as a deviation from the null hypothesis. Specifically, we evaluate the power of the five competing tests in the following sequence of alternative models:

$$dX_t = \{(1-\tau)\kappa(\mu - X_t) + \tau C(\mu - X_t)^3/X_t\}\,dt + \sigma\,dW_t, \quad (6.3)$$

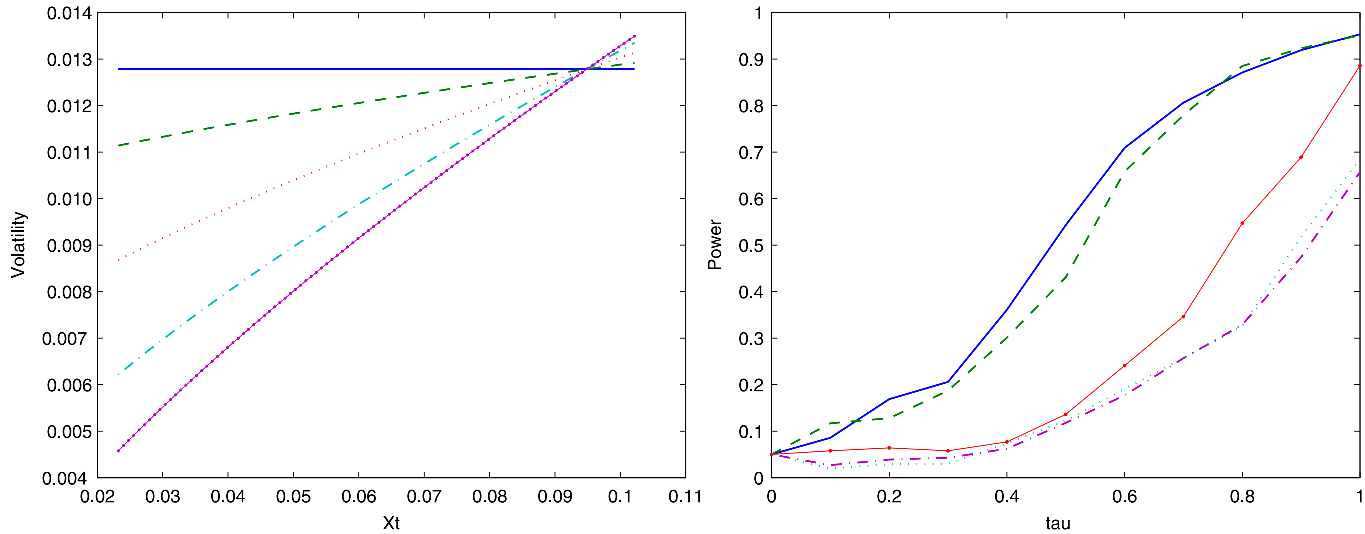

Figure 2. Example 1: Left panel: Volatility function (6.2) with $\tau = 0$ (solid), $\tau = 0.2$ (dash), $\tau = 0.5$ (dotted), $\tau = 0.8$ (dash-dotted), $\tau = 1$ (dotted-solid). Right panel: Power functions for the five competing test statistics evaluated at the sequence of alternative models (6.2). Solid line for T_1; dash line for T_3, dotted-solid line for T_4; dotted line for T_5, and dotted-dash line for T_6.

where the constants κ, μ, and σ are the same as those in the null model (6.1), and $C = 10\kappa/s_0 = 0.0294$, where $s_0 = 0.0263$ is the standard deviation of X_t under the null model (6.1).

When τ is small, the null and alterative models are nearly impossible to differentiate in terms of the drift functions (see Figure 3). Even when τ is large, the differences of the drift functions between the null and alternative models are mainly at high- and low-interest–rate regimes, which are not often visited by the process X_t. This the testing problem is intrinsically challenging.

Figure 3 also depicts the power function of the five test statistics for the alternative model (6.3) with significant level $\alpha = 0.05$. It is clear from this figure that our proposed test statistics T_1 and T_3 outperform the other three transformation-based tests, which have nearly no power in detecting the alternative models. Yet our proposed tests T_1 and T_3 are reasonably powerful, with the transition distribution–based test T_3 having the most power among the five tests.

Example 3. We now consider the power of nonparametric specification tests for (6.1) against a sequence of jump-diffusion models,

$$dX_t = \kappa(\mu - X_{t-})\,dt + \sigma(X_{t-})\,dW_t + J_{t-}\,dN_t \tag{6.4}$$

where N_t is a Poisson process with stochastic intensity $\lambda(X_t)$ and jump size 1. J_t is the jump size that is independent of $\mathcal{F}_t$ and has normal density. In the implementation, we take $\sigma(X_t) = \xi$, $\lambda(X_t) = \lambda$, and $J_t \sim N(0, \eta^2)$.

Under these specifications, the transition density for the jump diffusion model is approximately a mixture of normal distribu-

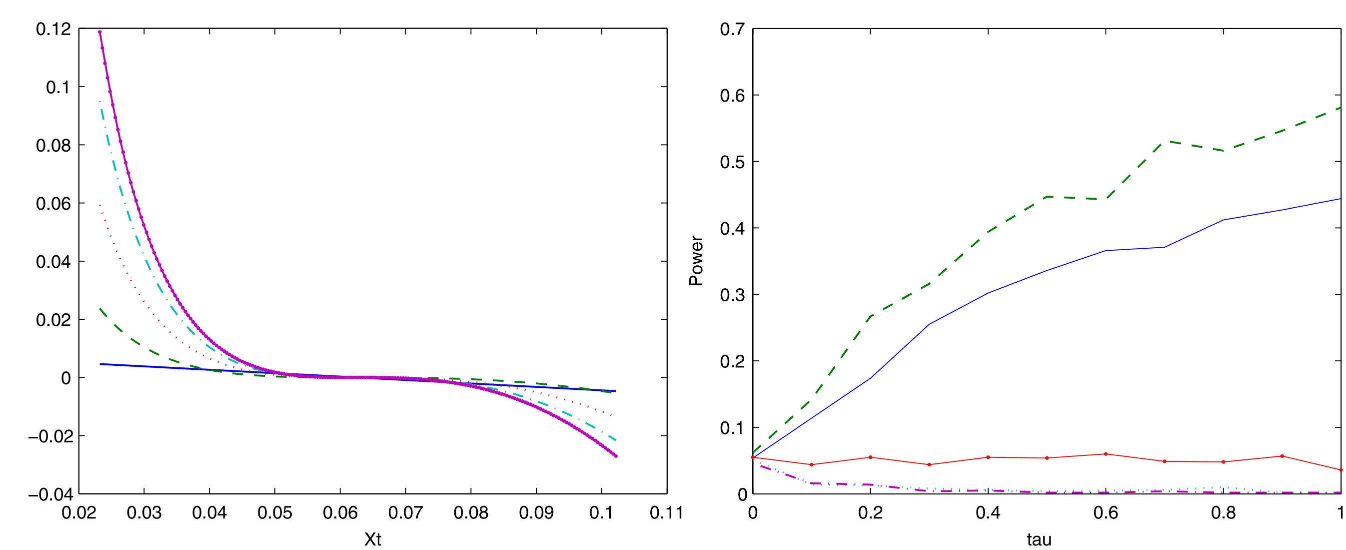

Figure 3. Example 2: Left panel: Drift function (6.3) with $\tau = 0$ (solid), $\tau = 0.2$ (dash), $\tau = 0.5$ (dotted), $\tau = 0.8$ (dash-dotted), $\tau = 1$ (dotted-solid). Right panel: Power functions for the five competing test statistics evaluated at the sequence of alternative models (6.3). Solid line for T_1; dash line for T_3, dotted-solid line for T_4; dotted line for T_5, and dotted-dash line for T_6.

tions,

$$p_\Delta(y|x) = (1-\lambda\Delta)N(\mu_\Delta(y|x), \sigma_\Delta^2) + \lambda\Delta N(\mu_\Delta(y|x), \sigma_\Delta^2 + \eta^2), \quad (6.5)$$

where $\mu_\Delta(y|x) = \mu + (x-\mu)\exp(-\kappa\Delta)$ and $\sigma_\Delta^2 = \xi^2\{1 - \exp(-2\kappa\Delta)\}/(2\kappa)$ are the conditional mean and variance from the Vasicek model.

The variance from the jump component is $\lambda\eta^2\Delta$, and that from the diffusion component is $\xi^2\Delta$. To make the testing problem more challenging and interesting, we set $\xi^2 + \lambda\eta^2 = \sigma^2$, the diffusion parameter of the Vasicek model under the null hypothesis. The ratio of the jump component to the total variance is given by $\tau^* = \lambda\eta^2/(\xi^2 + \lambda\eta^2)$. We compute the power at different ratios ranging from 0 to 0.9 or, more precisely, $\tau = 1.1\tau^*$ with $\tau = 0, 0.1, \ldots, 1.0$, ranging from no jump component (in which case the alternative hypothesis becomes the null hypothesis) to the case in which the variance of the jump component accounts for nearly 50% of the total variance.

Although there are three parameters in (6.4), we have only two equations so far:

$$\xi^2 + \lambda\eta^2 = \sigma^2; \qquad \frac{\lambda\eta^2}{\xi^2 + \lambda\eta^2} = \frac{\tau}{1.1},$$

for $\tau = 0, 0.1, \ldots, 1$. We compute the power under the following two specifications: (I) $\eta^2/\xi^2 = 2$ and (II) $\lambda = 2$ with different value of τ. The power is computed as a function of τ, of the test statistics for $\alpha = 0.01, 0.05$, and 0.10 under both model specifications (I) and (II), based on 2,400 weekly data and 1,000 simulations. Figure 4 presents only the results for $\alpha = 0.05$.

The likelihood ratio test T_0 is most powerful when τ is small, whereas the test T_7, a version of the generalized likelihood ratio test under the Rosenblatt transformation, is most powerful when τ is large. In contrast, the conditional distribution–based test T_3 has little power. To provide insight into the powerful properties, Figure 4(a) depicts the transition densities under both the null and alternative model with the conditional mean setting at 0 for model specification (I). The plot for specification (II) is similar and thus is omitted. Clearly, the deviations from the null hypothesis are locally around the conditional mean. Thus, as mentioned at the end of Sections 4 and 5.3, the conditional distribution–based tests are generally less powerful than their transition density–based counterparts. These results are consistent with those shown in Figure 4. Note that the transition density for $\tau = 1$ appears to be closer to that under the null hypothesis than than that for $\tau = 0.8$; this explains why the power decreases at $\tau = 1$.

To better demonstrate the properties of the Rosenblatt transformation–based tests T_4, T_5, and T_6, Figure 4(d) shows the density of Z_t, the Rosenblatt-transformed variable, for different values of τ. When $\tau = 0$, the null density should be uniformly distributed on [0, 1]. As τ deviates from 0, the deviations are away from uniform; however, the density appears to be closer to the uniform distribution when $\tau = 1$ than when $\tau = 0.8$. This again explains the power decreases at $\tau = 1$. These figures clearly show that the Rosenblatt-transformed methods intend to use the deviations at both the middle and the tails to enhance the power.

7. EMPIRICAL RESULTS

In this section we apply our tests to two classical data sets. The first data set comprises implied volatility data based on the Chicago Board Options Exchange's Volatility Index (VIX). The VIX data were computed using the methodology introduced by the Chicago Board Options Exchange on September 22, 2003, involving an implied volatility index based on the European S&P 500 options, the VIX is an estimate of the implied volatility of a basket of S&P 500 Index Options (SPX) constructed from different traded options in such a way that at any given time it represents the implied volatility of a hypothetical at-the-money option with 30 calendar days to expiration (or 21 trading days). The VIX options are European, simplifying the analysis. (For further details on the VIX, see Whaley 2000.) We use weekly data from June 5, 1990 to December 31, 2004, comprising 780 weekly observations. The second data set includes the yields of the 3-month Treasury Bill rate between January 8, 1982 and May 27, 2005. The data comprise 1,221 weekly observations based on the averages of the bid rates quoted on a bank discount basis by a sample of primary dealers who report to the Federal Reserve Bank of New York. Figure 5 shows these two weekly data sets. For both data sets, we are interested in testing the Constant Elasticity of Volatility (CEV) model,

$$dX_t = \kappa(\mu - X_t) + \sigma X_t^\rho\, dW_t. \quad (7.1)$$

For the VIX data, the parameters in the CEV model (7.1) are estimated using the closed-form likelihood expansion of Aït-Sahalia (2002). This results in the estimates $\hat\kappa = 2.5370$, $\hat\mu = 0.1992$, $\hat\sigma = 1.4823$, and $\hat\rho = 1.3899$. From here, the transition density and distribution under the parametric model can be evaluated. To check the degree of the departure from the null hypothesis, we estimate nonparametrically the conditional density and cumulative distribution functions. The bandwidths used here are based on the cross-validation method of Fan and Yim (2004) for estimating the conditional density and the plug-in method of Ruppert, Sheather, and Wand (1995) for estimating the conditional distribution. We use the resulting estimates to construct the test statistics T_1 and T_3. These estimates are undersmoothed (not presented here), which reduces the biases in the test statistics. This is reasonable for constructing the test statistics, because they average over these estimates. This feature is also shared by the test statistic T_4 of Hong and Li (2005) [see Figure 6(a), in which the null distribution is supposed to be uniform]. Based on 1,000 bootstrap samples, the estimated p-values are 0.443 and 0.058. They provide little evidence against the null hypothesis.

We also apply the test statistics T_4, T_5, and T_6 to this testing problem. The estimated joint density of Z_{t-1} and Z_t follows the recipes of Hong and Li (2005). To help visualize the bivariate density, Figure 6 shows several conditional distributions. Based on 1,000 bootstrap samples, the p-values are estimated and the critical values are reported. The results are summarized in Table 1. Except for the Hong–Li test T_4, all tests show little evidence against the null hypothesis. This is due to the roughness of the estimated bivariate densities caused by the bandwidth selection method of Hong and Li (2005).

We now apply the same techniques to the 3-month Treasury Bill data. To save space, we summarize our findings here. The

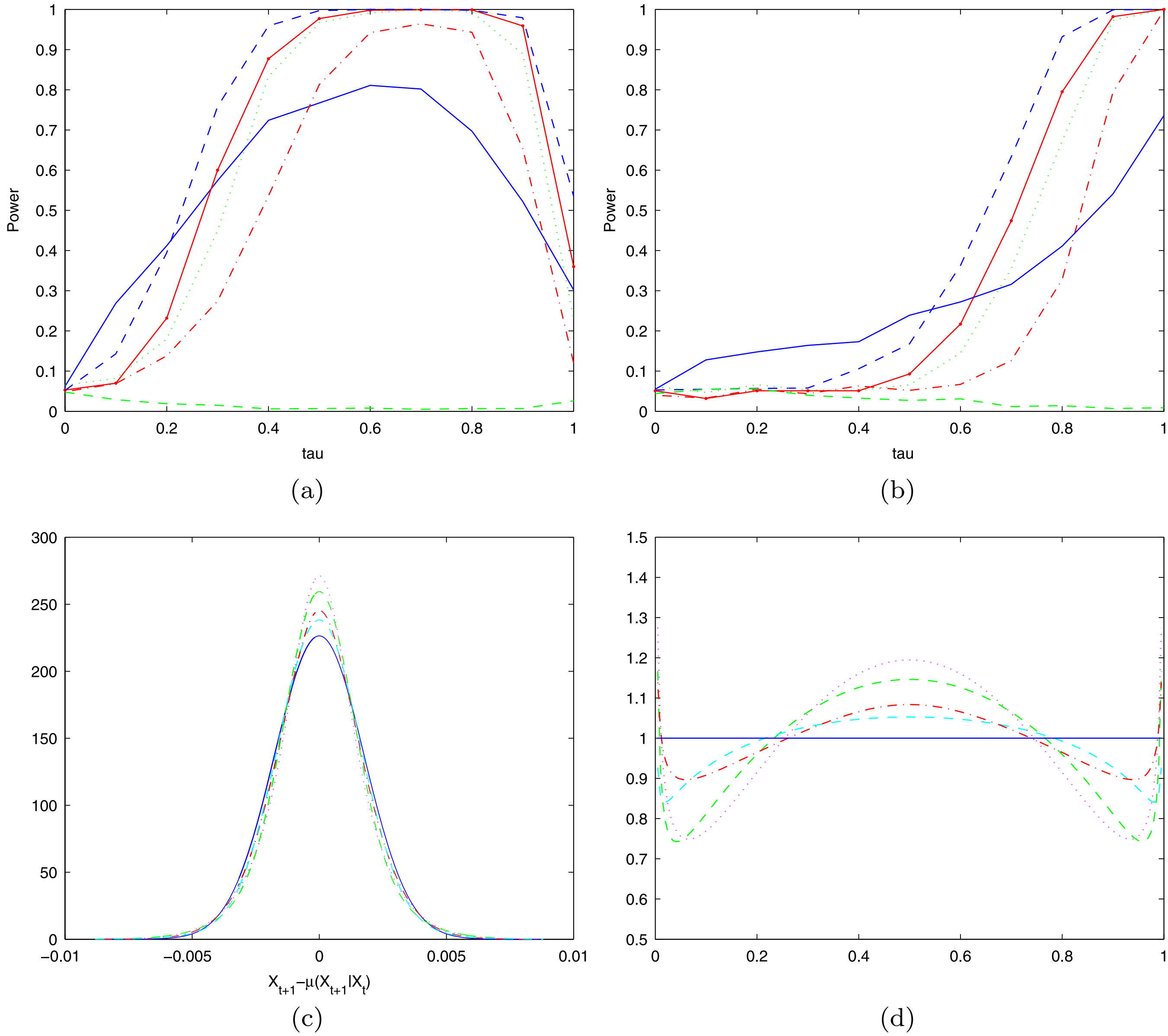

Figure 4. Example 3: (a) & (b) Power functions for the six competing test statistics evaluated at the sequence of alternative models index by τ with specification (I) (left panel) and specification (II) (right panel). Solid (blue) line for T_1; dash (green) line for T_3; dotted-solid (red) line for T_4; dotted (green) line for T_5 and dotted-dash (red) line for T_6; and dash (blue) line for T_7. (c) Transition density function for Example 3 with specification (I). (d) Density function of the Rosenblatt transformation of the observed process for Example 3 with specification (I). Solid blue line for $\tau = 0$; dash cyan line for $\tau = 0.2$; dash green line for $\tau = 0.5$; dotted red line for $\tau = 0.8$; dotted-dash red line for $\tau = 1.00$.

maximum likelihood estimates of the parameters under the null hypothesis are $\hat{\kappa} = 0.2434$, $\hat{\mu} = 0.05478$, $\hat{\sigma} = 0.1971$, and $\hat{\rho} = 1.0148$. The nonparametric estimate of the transition density and cumulative transition distributions are obtained with the bandwidths selected by the aforementioned data-driven methods. With estimated transition densities and cumulative transition distributions under both the null and alternative hypotheses, the test statistics T_1 and T_3 can be constructed. Table 2 presents the results of the test.

We also apply the transformation-based methods to 3-month Treasury Bill yields. Figure 6(c) shows the estimated bivariate density based on the Rosenblatt transformation, demonstrating a quite severe deviation from the uniform. Table 2 summarizes the test statistics, along with their critical values and p-values.

The critical values and p-values are estimated based on 1,000 bootstrap samples. All test statistics exhibit small p-values, providing stark evidence against the null hypothesis.

The Rosenblatt transformation is sensitive to the Markovian assumption. For real data sets, this assumption cannot be completely satisfied; thus the transition density–based method is more reliable.

8. CONCLUSIONS

We have proposed three model specification tests based on direct comparisons of the transition density/distribution estimated by a parametric method and that estimated directly by a nonparametric method. Our method is omnibus, relying only on the Markovian assumption. This significantly reduces the like-

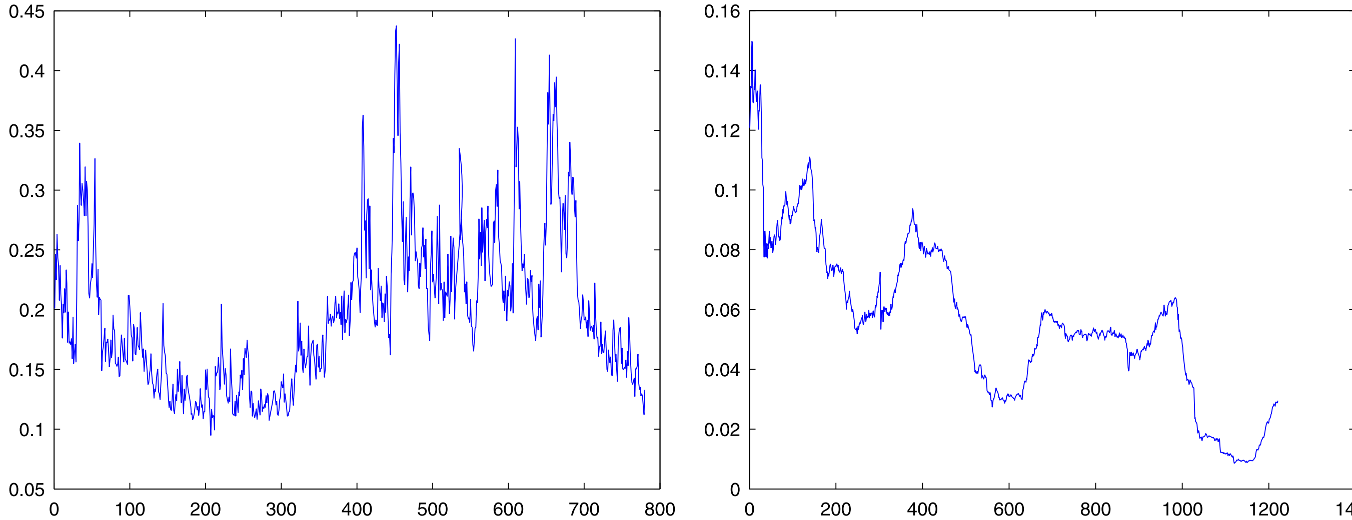

Figure 5. Left panel: The implied volatility data based on a basket of S&P 500 Index Options (SPX) of a hypothetical at-the-money option with 30 calendar days to expiration (or 22 trading days); weekly data from June 5, 1990 to December 31, 2004. Right panel: The yield of a 3-month Treasury Bill from January 8, 1982 to May 27, 2005.

lihood of model misspecification in hypothesis testing and significantly increases the applicability of our methods to a wide array of specification test problems.

We have derived the power of both transition density–based tests and transition distribution–based tests. When the deviation tends to be more global from the null hypothesis, the transition distribution–based test tends to be more powerful. Conversely, when the deviation tends to be more local, the transition density–based test tends to be more powerful. In general, the density-based tests provide sharper technical tools; however, which test is the most powerful for a particular problem remains unclear. For our simulated models, the tests do not seem to differ much; compare the powers between T_1 and T_3 and also between T_4 and T_6.

Related to our direct approaches are the Rosenblatt transformation–based tests. Although determing which methods are more powerful is difficult, the transition density– and distribution-based tests are more intuitive. They directly measure the deviations in terms of the transition density/distributions from which the data are generated. Our simulation results indicate that our proposed direct approach has some advantages. Furthermore, the direct approach has a wider scope of applicability, including testing whether a process contains jumps. In future research, we intend to develop related ideas for the problem of testing the Markov hypothesis.

APPENDIX A: TECHNICAL CONDITIONS

The first condition states that the parametric model under the null hypothesis is well specified for our purposes. Primitive conditions in terms of $(\mu, \sigma^2, \lambda, \nu)$ that lead to smoothness of $p(y|x;\theta)$ have been given by Yu (2007). We assume that

Condition 1. The specification of $(\mu, \sigma^2, \lambda, \nu)$ is such that the model (1.1) admits a unique solution and satisfies the smoothness and boundary behavior required for construction of an expansion in Δ of its parametric transition function $p(y|x;\theta)$ that Kth-order derivatives with respect to Δ is Lipschitz with respect to θ on Θ.

Condition 2. The kernel functions W and K are symmetric and bounded with a bounded support, and are Lipschitz.

Condition 3. The weight function $w(x, y)$ has continuous second-order derivatives with a compact support Ω^*.

Condition 4. The observed time series $\{X_i, i = 1, \dots, n\}$ is stationary and Markovian, with transition density $p(y|x)$. The stationary Markov process $\{X_i\}$ is β-mixing with exponential decay rate $\beta(n) = O(e^{-\lambda n})$ for some $\lambda > 0$.

Condition 5. The transition function $p(y|x)$ has continuous fourth-order partial derivatives with respect to x and y on the set Ω^*. The invariant density $\pi(x)$ has a continuous second derivative. Furthermore, $\pi(x) > 0$ and $p(y|x) > 0$ for all $(x, y) \in \Omega^*$.

Condition 6. The joint density of distinct elements of (X_1, X_ℓ), $\ell > 1$ is bounded by a constant independent of ℓ. Let

$$g(x_1, x_\ell) = f(x_1, x_\ell) - f(x_1)f(x_\ell),$$

which is Lipschitz; for all (x', y') and (x, y) in Ω^*,

$$|g(x', y') - g(x, y)| \le C(\|x' - x\|^2 + \|y' - y\|^2)^{1/2}.$$

Condition 7. The sampling interval Δ is such that $\Delta^K = o(n^{-1/2})$ as $n \to \infty$.

Condition 8. The bandwidths h_1 and h_2 converge to 0 such that $n(h_1^3 + h_2^3)/\log n \to \infty$, $n(h_1^5 + h_2^5) \to 0$, and h_1 and h_2 are of the same order.

Condition 9. The bandwidth h_1 converges to 0 such that $nh_1^{9/2} \to 0$ and $nh_1^{3/2} \to \infty$.

For a diffusion process, Condition 4 can easily be satisfied (see, e.g., lemma 4 of Aït-Sahalia and Mykland 2004). Conditions 2, 3, 5, and 6 are not the weakest possible, but they are imposed to facilitate the technical proofs. In particular, Condition 5 is set not because higher-order kernels are used, but rather to facilitate the derivation of some bounds.

APPENDIX B: PROOFS OF THE THEOREMS

The details of the proofs are necessarily tedious and complex. To highlight the key ideas of the proofs, we relegate some technical lemmas to a supplement of this article. We furnish some details for the proof of Theorem 1 and note only the differences for other similar

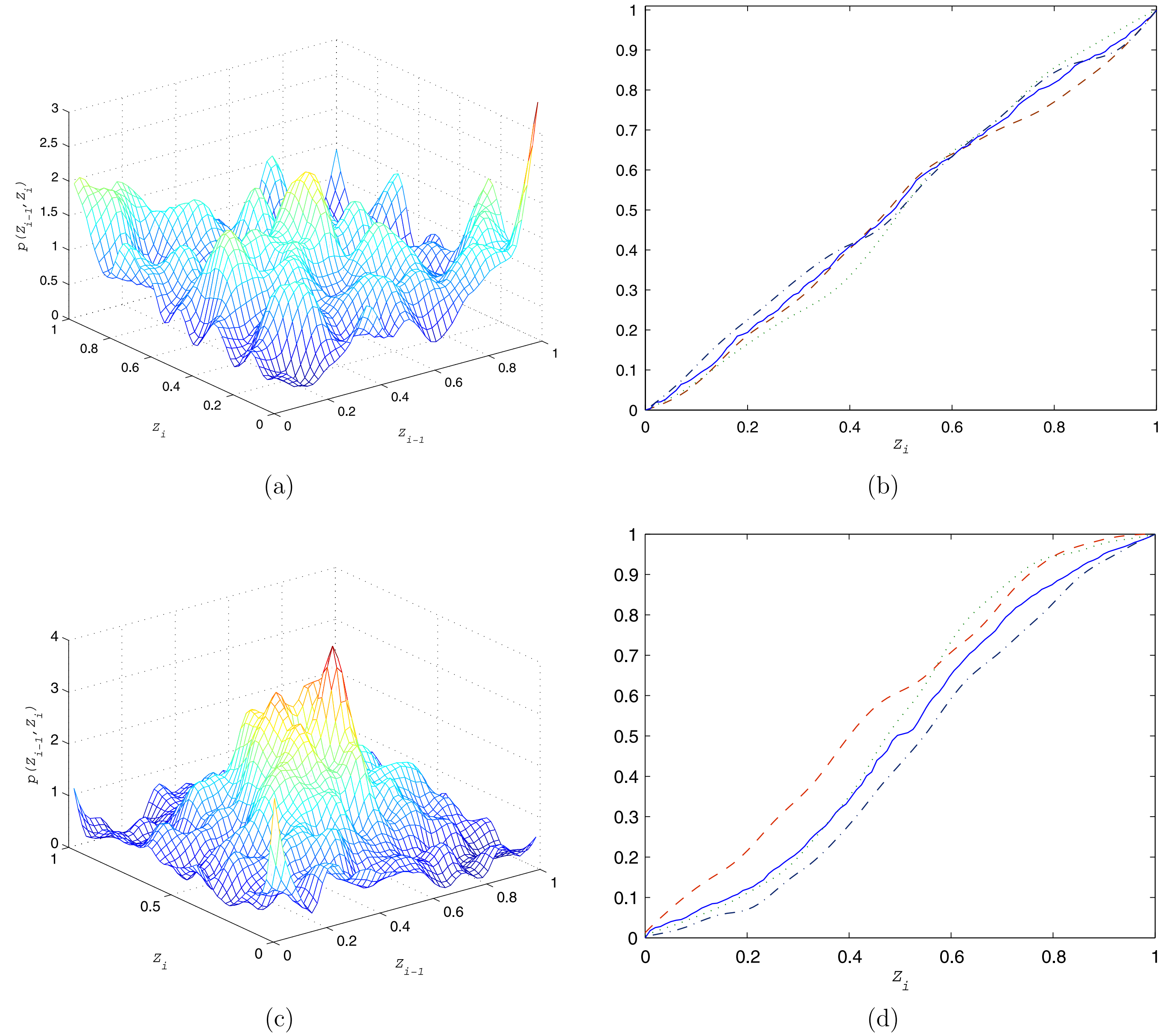

Figure 6. Estimated joint density of Rosenblatt-transformed random variables for the VIX data and yields of 3-month Treasury Bills. (a) & (c) Estimated joint density of (Z_{t-1}, Z_t) with $Z_t = P(X_t|X_{t-1}, \hat{\theta})$. (b) & (d) Conditional distribution of Z_t given $Z_{t-1} = z$ for $z = 0.25$ (dash), 0.5 (dotted), 0.75 (dotted-dash) along with the marginal cdf of Z_t (solid).

Table 1. Test statistics, critical values and p-values based on the VIX data

	T_1	T_3	T_4	T_5	T_6
Critical value ($\alpha = 0.01$)	83.6824	3.7359	3.4000	0.9139	0.9985
Critical value ($\alpha = 0.05$)	64.2587	3.0652	2.1787	0.4959	0.5871
Critical value ($\alpha = 0.10$)	56.6583	2.7749	1.6631	0.3708	0.4195
Test-stat	40.4929	3.0116	6.7640	0.3021	0.2869
p-value	0.4426	0.0582	0.000	0.1503	0.1957

proofs, which are given in the supplement. To simplify the proof of Theorem 1 without losing the essential ingredients, we replace the estimate $\hat{\theta}$ by its true value θ under the null model. This is because $\hat{\theta}$ is root-n–consistent according to Aït-Sahalia (2002) and $p(y|x;\theta)$ is Lipschitz in θ for all $(x, y) \in \Omega^*$; thus the difference between $\hat{\theta}$ and θ has no impact.

Proof of Theorem 1

Let $\rho_n = h_1^2 + \sqrt{\log n/(nh_1)}$ and $\rho_n' = h_1 + \sqrt{\log n/(nh_1)}$. Denote by $m(x, y) = \mathrm{E}K_{h_2}\{(X_2 - y)|X_1 = x\}$ and $m_1(x, y) = \frac{\partial m(x,y)}{\partial x}$. To ease the understanding of the proof, set $Y_i = X_{i+1}$. By an elementary property of the local linear fitting, we have that

$$\hat{p}(Y_i|X_i) - p(Y_i|X_i) = A_i + B_i + C_i, \tag{B.1}$$

Table 2. Test statistics, critical values and p-values based on the the 3-month Treasury Bill data

	T_1	T_3	T_4	T_5	T_6
Critical value ($\alpha = 0.01$)	151.1197	8.7614	3.9964	0.8945	0.9592
Critical value ($\alpha = 0.05$)	101.1163	7.5036	2.7461	0.5541	0.6258
Critical value ($\alpha = 0.10$)	86.9007	6.9694	2.1173	0.4162	0.4513
Test-stat	151.2424	7.3990	44.9091	4.0166	3.6301
p-value	0.0100	0.0568	0.0000	0.0000	0.0000

where, with $p(y|x) = p(y|x;\theta)$,

$$A_i = \frac{1}{nh_1}\sum_{j=1}^{n} W_n\left(\frac{X_j - X_i}{h_1}; X_i\right)\left\{K_{h_2}(Y_j - Y_i) - m(X_j, Y_i)\right\},$$

$$B_i = \frac{1}{nh_1}\sum_{j=1}^{n} W_n\left(\frac{X_j - X_i}{h_1}; X_i\right) \times \{m(X_j, Y_i) - m(X_i, Y_i) - m_1(X_i, Y_i)(X_j - X_i)\},$$

$$C_i = m(X_i, Y_i) - p(Y_i|X_i).$$

Thus the test statistic T_1 can be written as

$$T_1 = \sum_{i=1}^{n} \frac{w(X_i, Y_i)}{p^2(Y_i|X_i)}\{A_i^2 + B_i^2 + C_i^2 + 2A_iB_i + 2A_iC_i + 2B_iC_i\} \hat{=} T_1^* + T_2^* + T_3^* + T_4^* + T_5^* + T_6^*.$$

Following Fan, Yao, and Tong (1996), it is easy to show that uniformly in i, we have

$$B_i = O_p(h_1^2) \quad \text{and} \quad C_i = O_p(h_2^2).$$

Thus, under Condition 8, we have

$$\begin{cases} T_2^* = O_p(nh_1^4) = o_p(1/\sqrt{h_1h_2}), \\ T_3^* = O_p(nh_2^4) = o_p(1/\sqrt{h_1h_2}), \\ T_6^* = O_p(nh_1^2h_2^2) = o_p(1/\sqrt{h_1h_2}). \end{cases}$$

It can be shown, using the uniform convergence results similar to those given by Mack and Sliverman (1982) and lemma B of Lee (1990), that (see the full version of the article)

$$T_4^* = o_p(1/\sqrt{h_1h_2}) \quad \text{and} \quad T_5^* = o_p(1/\sqrt{h_1h_2}). \tag{B.2}$$

Thus T_1^* is the dominating term,

$$T_1 = T_1^* + o_p(1/\sqrt{h_1h_2}). \tag{B.3}$$

To deal with T_1^*, again using the uniform convergence results and the Hoeffding decomposition, it can decomposed into three terms, T_{11}, T_{12}, and T_{13},

$$T_1^* = T_{11} + T_{12} + T_{13} + o_p(1/\sqrt{h_1h_2}). \tag{B.4}$$

These terms take the following forms:

$$T_{11} = \sum_{i<j<k}^{n} \varphi(i,j,k) + \varphi(i,k,j) + \varphi(j,i,k) + \varphi(j,k,i) + \varphi(k,i,j) + \varphi(k,j,i),$$

$$T_{12} = \sum_{i \neq j}^{n} \varphi(i,j,j) + \varphi(j,i,j) + \varphi(j,j,i), \quad \text{and}$$

$$T_{13} = \sum_{i=1}^{n} \varphi(i,i,i),$$

where

$$\varphi(i,j,k) = \frac{w(X_i, Y_i)}{n^2 f^2(X_i, Y_i)} W_{h_1}(X_j - X_i)\left\{K_{h_2}(Y_j - Y_i) - m(X_j, Y_i)\right\} \times W_{h_1}(X_k - X_i)\left\{K_{h_2}(Y_k - Y_i) - m(X_k, Y_i)\right\},$$

where $f(x,y) = p(x)p(y|x)$ is the joint density of (X_i, Y_i). The rest of the proof involves the following steps, which we describe in detail later:

(a) Let $\varphi^*(i,j,k) = \varphi(i,j,k) + \varphi(i,k,j) + \varphi(j,i,k) + \varphi(j,k,l) + \varphi(k,i,j) + \varphi(k,j,i)$ be a symmetrical kernel function. By Hoeffding's decomposition and simplification, we can show that

$$T_{11} = (n-2)\sum_{i<j} \varphi^*(i,j) + o_p\left(\frac{1}{\sqrt{h_1h_2}}\right),$$

where $\varphi^*(i,j) = \int \varphi^*(i,j,k)\, dF(x_k, y_k)$, and F is the distribution of (X_k, Y_k). In other words, $\varphi^*(i,j)$ is the expectation of $\varphi^*(i,j,k)$ with respect to the variables X_k and Y_k. By the Markovian property, it is obvious that the expectation of $\varphi^*(i,j)$ is 0; thus T_{11} is a U-statistic with mean 0.

(b) Define $\tilde{\varphi}(i,j) = \varphi(i,i,j) + \varphi(i,j,i) + \varphi(j,i,i) + \varphi(j,j,i) + \varphi(j,i,j) + \varphi(i,j,j)$, and also $\tilde{\varphi}(i) = \int \tilde{\varphi}(i,j)\, dF(x_j, y_j)$, and $\tilde{\varphi}(0) = \mathrm{E}\{\tilde{\varphi}(i)\}$. Then we can see that

$$T_{12} = \frac{n(n-1)}{2}\tilde{\varphi}(0) + o_p(1/\sqrt{h_1h_2}).$$

Clearly, T_{12} converges to a constant. In fact, this constant is the mean of the limit distribution in our theorem.

(c) It can be shown that $T_{13} = o_p(1/\sqrt{h_1h_2})$ by the central limit theorem for β-mixing series.

(d) Combining the results in the first three steps and using (B.3) and (B.4), T_1 can be written as

$$T_1 = \frac{n(n-1)}{2}\tilde{\varphi}(0) + (n-2)\sum_{i<j}\varphi^*(i,j) + o_p(1/\sqrt{h_1h_2}). \tag{B.5}$$

This is the sum of a constant and a U-statistic with mean 0. To prove our theorem, we need only show that

$$\frac{n(n-1)}{2}\tilde{\varphi}(0) = \frac{\|\mathbf{1}\|_w}{h_1h_2}\|W\|^2\|K\|^2 - \frac{\Omega_x}{h_1}\|W\|^2 + o(1/\sqrt{h_1h_2}),$$

and the asymptotic normality of the U-statistic,

$$\frac{1}{\sigma_0^*}\sum_{i<j}^{n}(n-2)\varphi^*(i,j) \to \mathcal{N}(0,1),$$

where $\sigma_0^{*2} = 2\|w\|^2\|W * W\|^2\|K * K\|^2/(h_1h_2)$. To establish the asymptotic normality of the U-statistic, we need only check the conditions of Lemma 6.

We now provide some details of the proofs.

Proof of Claim (a). Let $\Phi(i,j,k) = \varphi^*(i,j,k) - \varphi^*(i,j) - \varphi^*(i,k) - \varphi^*(j,k)$. Then, by Hoeffding's decomposition, we can write

$$T_{11} = \sum_{i<j<k}\Phi(i,j,k) + (n-2)\sum_{i<j}\varphi^*(i,j).$$

Let us denote the first term as T_{11}^*. By Lemma 6(i) in the supplement with $\delta = 1/3$, we have

$$\mathrm{E}\{n^2h_1^2h_2^2T_{11}^*\}^2 \le Cn^3 \max\left\{(h_1^2h_2^2)^{3/4}\sum_{k=1}^n k^2\beta^{1/4}(k), h_1^2h_2^2\right\} \le Cn^3(h_1^2h_2^2)^{3/4},$$

for a generic constant C. Thus

$$\mathrm{E}\{T_{11}^*\}^2 \le C/\big(nh_1^{5/2}h_2^{5/2}\big),$$

which implies that

$$T_{11}^* = O_p\{\big(nh_1^{5/2}h_2^{5/2}\big)^{-1/2}\} = o_p(1/\sqrt{h_1h_2}).$$

Proofs of Claims (b) and (c). First, note that $\tilde\varphi(i,j)$ is the symmetric kernel. Following Hoeffding's decomposition, we can write T_{12} as

$$\sum_{i<j}\tilde\varphi(i,j) = \sum_{i<j}\{\tilde\varphi(i,j) - \tilde\varphi(i) - \tilde\varphi(j) + \tilde\varphi(0)\} + (n-1)\sum_{i=1}^n\{\tilde\varphi(i) - \tilde\varphi(0)\} + \frac{n(n-1)}{2}\tilde\varphi(0). \quad \text{(B.6)}$$

Then, by Lemma 6(ii) in the supplement with $\delta = 1$, we have

$$\mathrm{E}\left\{n^2h_1^2h_2^2\sum_{i<j}\{\tilde\varphi(i,j) - \tilde\varphi(i) - \tilde\varphi(j) + \tilde\varphi(0)\}\right\}^2 \le Cn^2\sqrt{h_1h_2}.$$

This entails

$$\sum_{i<j}\{\tilde\varphi(i,j) - \tilde\varphi(i) - \tilde\varphi(j) + \tilde\varphi(0)\} = O_p\left(\frac{1}{nh_1^{7/4}h_2^{7/4}}\right) = o_p(1/\sqrt{h_1h_2}). \quad \text{(B.7)}$$

For the second term in (B.6), note that $\mathrm{E}\{n^2h_1h_2[\tilde\varphi(i) - \tilde\varphi(0)]\}^2 = O(1)$. By the central limit theorem of the β-mixing process, we have

$$(n-1)\sum_{i=1}^n\{\tilde\varphi(i) - \tilde\varphi(0)\} = O_p\{\sqrt{n}/(nh_1h_2)\} = o_p(1/\sqrt{h_1h_2}).$$

This, together with (B.6) and (B.7), imply claim (b).

For claim (c), it is not difficult to show that

$$T_{13} = \sum_{i=1}^n\varphi(i,i,i) = O_p\{n/n^2h_1^2h_2^2\} = O_p(1/nh_1^2h_2^2) = o_p(1/\sqrt{h_1h_2}),$$

by applying the central limit theorem for the β-mixing process.

Proof of Claim (d)—Asymptotic Mean. The asymptotic mean comes mainly from the first term in the right side of (B.5). We now calculate its asymptotic value. By the definition of $\tilde\varphi(0)$, it is not hard to show that

$$\begin{aligned}\tilde\varphi(0) &= 2\int\frac{w(x_i,y_i)}{n^2f^2(x_i,y_i)}W_{h_1}^2(x_j-x_i)K_{h_2}^2(y_j-y_i)\,dF(x_i,y_i)\,dF(x_j,y_j)\\ &\quad -4\int\frac{w(x_i,y_i)}{n^2f^2(x_i,y_i)}W_{h_1}^2(x_j-x_i)K_{h_2}(y_j-y_i)\\ &\quad \times m(x_j,y_i)\,dF(x_i,y_i)\,dF(x_j,y_j)\\ &\quad +2\int\frac{w(x_i,y_i)}{n^2f^2(x_i,y_i)}W_{h_1}^2(x_j-x_i)m^2(x_j,y_i)\,dF(x_i,y_i)\,dF(x_j,y_j).\end{aligned}$$

Denote the first, second, and third terms by U_1, U_2, and U_3. Now we compute the asymptotic values of U_1, U_2, and U_3. By a change of variable and Taylor's expansion, we have

$$U_1 = \frac{2\|\mathbf{1}\|_w}{n^2h_1h_2}\{\|W\|^2\|K\|^2 + O(h_1^2+h_2^2)\}.$$

Taking the conditional expectation of Y_j given X_j, we get

$$U_2 = -4\int\frac{w(x_i,y_i)}{n^2f^2(x_i,y_i)}W_{h_1}^2(x_j-x_i)m^2(x_j,y_i)\,dF(x_j)\,dF(x_i,y_i),$$

where $F(x_i)$ is the cumulative distribution function of X_i. Now, by a change of variable and Taylor's expansion, we have

$$U_2 = -4\frac{\Omega_x + O(h_1^2+h_2^2)}{n^2h_1}\|W\|^2.$$

Using a similar argument, we have that

$$U_3 = 2\frac{\Omega_x + O(h_1^2+h_2^2)}{h_1}\|W\|^2.$$

Combining the results for U_1, U_2, and U_3, and ignoring terms of order $o_p(1/\sqrt{h_1h_2})$, we have

$$\frac{n(n-1)}{2}\tilde\varphi(0) = \frac{\|\mathbf{1}\|_w}{h_1h_2}\|W\|^2\|K\|^2 - \frac{\Omega_x}{h_1}\|W\|^2 + o(1/\sqrt{h_1h_2}).$$

Proof of Claim (d)—Asymptotic Normality. We now consider the asymptotic normality of $(n-2)\sum_{i<j}^n\varphi^*(i,j)$ by using a result of Hjellvik, Yao, and Tjostheim (1998), which is stated as Lemma 7 in the supplement. Toward this end, define $\Phi_{i,j} = nh_1h_2(n-2)\varphi^*(i,j)$ and $\sigma_0^2 = \int\Phi_{i,j}^2\,dF(x_i,y_i)\,dF(x_j,y_j)$, and $\sigma_n^2 = n(n-1)\sigma_0^2/2$.

First, we calculate σ_0^2. Note that from the definition of $\varphi^*(i,j)$ and the mean 0 property of $\varphi(i,j,k)$, It also is not hard to check, by Fubini's theorem and a change of variable, that

$$\begin{aligned}\varphi^*(i,j) &= 2\int\frac{1}{\pi(x_k)}W_{h_1}(x_i-x_k)W_{h_1}(x_j-x_k)\\ &\quad\times\int K_{h_2}(y_i-y_k)K_{h_2}(y_j-y_k)\frac{w(x_k,y_k)}{p(y_k|x_k)}\,dy_k\,dx_k\\ &= 2\int\frac{1}{\pi(x_k)}W_{h_1}(x_i-x_k)W_{h_1}(x_j-x_k)\\ &\quad\times\int\frac{1}{h_2}K\left(u+\frac{y_i-y_j}{2h_2}\right)K\left(u+\frac{y_j-y_i}{2h_2}\right)\\ &\quad\times\frac{w\big(x_k,\frac{(y_i+y_j)}{2}+uh_2\big)}{p\big(\frac{y_i+y_j}{2}+uh_2|x_k\big)}\,du\,dx_k.\end{aligned}$$

Letting $K^* = K*K$ and $W^* = W*W$, by Taylor's expansion, we obtain

$$\varphi^*(i,j) = \frac{2w\big(\frac{x_i+x_j}{2},\frac{y_i+y_j}{2}\big)}{f\big(\frac{x_i+x_j}{2},\frac{y_i+y_j}{2}\big)}W_{h_1}^*(x_i-x_j)K_{h_2}^*(y_i-y_j) + O_p(h_1/h_2+h_2/h_1). \quad \text{(B.8)}$$

From the definition of σ_0^2, we have

$$\sigma_0^2/(n^2h_1^2h_2^2) = \int\{(n-2)\varphi^*(i,j)\}^2\,dF(x_i,y_i)\,dF(x_j,y_j).$$

Using the same technique as before, we can easily see that

$$\begin{aligned}&\int\left\{\frac{w\big(\frac{x_i+x_j}{2},\frac{y_i+y_j}{2}\big)}{f\big(\frac{x_i+x_j}{2},\frac{y_i+y_j}{2}\big)}W_{h_1}^*(x_i-x_j)K_{h_2}^*(y_i-y_j) + O_p(h_1/h_2+h_2/h_1)\right\}^2 dF(x_i,y_i)\,dF(x_j,y_j)\\ &= \frac{\|w\|^2}{h_1h_2}\|W^*\|^2\|K^*\|^2 + o\{1/(h_1h_2)\}.\end{aligned}$$

Neglecting the terms of $o\{1/(n^2h_1h_2)\}$, it is not difficult to show that, from (B.8),

$$\sigma_n^2 = \frac{n(n-1)}{2}\sigma_0^2$$
$$= 2n(n-1)\|w\|^2h_1h_2\|W^*\|^2\|K^*\|^2 + o\{n^2h_1h_2\}. \quad \text{(B.9)}$$

We still need to check the conditions of Lemma 7 in the supplement. We omit the details of this here and refer the interested readers to the supplement.

Proofs of Theorems 2–5

See the supplement for details.

SUPPLEMENTAL MATERIALS

This supplement furnishes some details on the technical proofs of our theoretical results.

[Received April 2008. Revised July 2008.]

REFERENCES

Aït-Sahalia, Y. (1996), "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

——— (2002), "Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223–262.

——— (2008), "Closed-Form Likelihood Expansions for Multivariate Diffusions," *Annals of Statistics*, 36, 906–937.

Aït-Sahalia, Y., and Mykland, P. A. (2003), "The Effects of Random and Discrete Sampling When Estimating Continuous-Time Diffusions," *Econometrica*, 71, 483–549.

——— (2004), "Estimators of Diffusions With Randomly Spaced Discrete Observations: A General Theory," *Annals of Statistics*, 32, 2186–2222.

Altissimo, F., and Mele, A. (2009), "Simulated Non-Parametric Estimation of Dynamic Models," *The Review of Economic Studies*, 76, 413–450.

Anderson, T. W., and Darling, D. A. (1952), "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics*, 23, 193–212.

Andrews, D. (1997), "A Conditional Kolmogorov Test," *Econometrica*, 6, 1097–1128.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992), "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 48, 1209–1227.

Chen, S. X., and Gao, J. (2007), "An Adaptive Empirical Likelihood Test for Time Series Models," *Journal of Econometrics*, 141, 950–972.

Chen, S. X., Gao, J., and Tang, C. Y. (2008), "A Test for Model Specification of Diffusion Processes," *Annals of Statistics*, 36, 167–198.

Fan, J. (1992), "Design-Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

Fan, J., and Yim, T.-H. (2004), "A Data-Driven Method for Estimating Conditional Densities," *Biometrika*, 91, 819–834.

Fan, J., and Zhang, C. (2003), "A Reexamination of Diffusion Estimations With Applications to Financial Model Validation," *Journal of the American Statistical Association*, 13, 965–992.

Fan, J., and Zhang, W. (2004), "Generalized Likelihood Ratio Tests for Spectral Density," *Biometrika*, 91, 195–209.

Fan, J., Yao, Q., and Tong, H. (1996), "Estimation of Conditional Densities and Sensitivity Measures in Nonlinear Dynamical Systems," *Biometrika*, 83, 189–206.

Fan, J., Zhang, C., and Zhang, J. (2001), "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *Annals of Statistics*, 29, 153–193.

Hall, P., Racine, J., and Li, Q. (2004), "Cross-Validation and the Estimation of Conditional Probability Densities," *Journal of the American Statistical Association*, 99, 1015–1026.

Hall, P., Wolff, R., and Yao, Q. (1999), "Methods for Estimating a Conditional Distribution Function," *Journal of the American Statistical Association*, 94, 154–163.

Hjellvik, V., Yao, Q., and Tjostheim, D. (1998), "Linearity Testing Using Local Polynomial Approximation," *Journal of Statistical Planning and Inference*, 68, 295–321.

Hong, Y., and Li, H. (2005), "Nonparametric Specification Testing for Continuous-Time Models With Applications to Term Structure of Interest Rates," *Review of Financial Studies*, 18, 37–84.

Hurn, A. S., Jeisman, J., and Lindsay, K. (2007), "Seeing the Wood for the Trees: A Critical Evaluation of Methods to Estimate the Parameters of Stochastic Differential Equations," *Journal of Financial Econometrics*, 5, 390–455.

Ingster, Y. (1993), "Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives I–III," *Mathematics Methods of Statistics*, 2, 85–114; 3, 171–189; 4, 249–268.

Jensen, B., and Poulsen, R. (2002), "Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques," *Journal of Derivatives*, 9, 1–15.

Lee, A. J. (1990), *U-Statistics, Theory and Practice*, New York: Marcel Dekker.

Lepski, O., and Spokoiny, V. (1999), "Minimax Nonparametric Hypothesis Testing: The Case of an Inhomogeneous Alternative," *Bernoulli*, 5, 333–358.

Li, F., and Tkacz, G. (2006), "A Consistent Bootstrap Test for Conditional Density Functions With Time-Series Data," *Journal of Econometrics*, 133, 863–886.

Mack, Y., and Sliverman, B. (1982), "Weak and Strong Uniform Consistency of Kernel Regression and Density Estimates," *Zeitschrift für Wahrsheinlichkeitstheorie und verwandte Gebiete*, 61, 405–415.

Ruppert, D., Sheather, S., and Wand, M. (1995), "An Effective Bandwidth Selector for Local Least Squares Regression," *Journal of the American Statistical Association*, 90, 1257–1270.

Stramer, O., and Yan, J. (2007), "On Simulated Likelihood of Discretely Observed Diffusion Processes and Comparison to Closed-Form Approximation," *Journal of Computational and Graphical Statistics*, 16, 672–691.

Thompson, S. (2008), "Identifying Term Structure Volatility From the LIBOR-Swap Curve," *Review of Financial Studies*, 21, 819–854.

Wang, Y. (2002), "Asymptotic Nonequivalence of GARCH Models and Diffusions," *Annals of Statistics*, 30, 754–783.

Whaley, R. E. (2000), "The Investor Fear Gauge," *The Journal of Portfolio Management*, 26, 12–17.

Yu, J. (2007), "Closed-Form Likelihood Approximation and Estimation of Jump-Diffusions With an Application to the Realignment Risk Premium of the Chinese Yuan," *Journal of Econometrics*, 141, 1245–1280.

Nonparametric Transition-Based Tests for Jump-Diffusions *

Yacine Aït-Sahalia, Jianqing Fan and Heng Peng

December 11, 2008

Supplement

This supplement furnishes some details on the technical proofs of our theoretical results.

Completion of Proof of Theorem 1

Now we check the conditions of Lemma 7 in Supplement. To this end, let

$$M_{n1} = \max_{i<j} \max \left\{ \mathrm{E}|\Phi_{1j}\Phi_{ij}|^{1+\delta}, \int |\Phi_{1j}\Phi_{ij}|^{1+\delta} dF(x_1, y_1) dF((X_i, Y_i), (X_j, Y_j)) \right\},$$

$$\begin{aligned} M_{n2} = \max_{i<j} \max \Big\{ & \mathrm{E}|\Phi_{1j}\Phi_{ij}|^{2(1+\delta)}, \int |\Phi_{1j}\Phi_{ij}|^{2(1+\delta)} dF(x_1, y_1) dF((x_i, y_i), (x_j, y_j)), \\ & \int |\Phi_{1j}\Phi_{ij}|^{2(1+\delta)} dF((x_1, y_1), (x_i, y_i)) dF((x_j, y_j)), \\ & \int |\Phi_{1j}\Phi_{ij}|^{2(1+\delta)} dF(x_1, y_1) dF(x_i, y_i) dF(x_j, y_j) \Big\}, \end{aligned}$$

$$M_{n3} = \max_{i<j} \mathrm{E}|\Phi_{1j}\Phi_{ij}|^2,$$
$$M_{n4} = \max_{i,j,k \text{ different}} \left\{ \max_F \int |\Phi_{1j}\Phi_{ij}|^{2(1+\delta)} dF \right\},$$
$$M_{n5} = \max_{i<j} \max \left\{ \mathrm{E}\left| \int \Phi_{1j}\Phi_{1j} dF(x_1, y_1) \right|^{2(1+\delta)}, \right.$$
$$\left. \int \left| \int \Phi_{1j}\Phi_{1j} dF((x_1, y_1)) \right|^{2(1+\delta)} dF(x_i, y_i) dF(x_j, y_j) \right\},$$
$$M_{n6} = \max_{i<j} \mathrm{E}\left| \int \Phi_{1j}\Phi_{1j} dF(x_1, y_1) \right|^2.$$

Here, in the definition of M_{n4}, the maximization is taken over the four probability measures given in Supplement. By the Markovian property, we have

$$\int \Phi_{i,j} dF(x_i, y_i) = \int \Phi_{i,j} dF(x_j, y_j) = 0$$

and

$$\mathrm{E}\{\Phi_{i,j} | (x_1, y_1), \ldots, (x_{j-1}, y_{j-1})\} = 0, \text{for any} \quad i < j.$$

We let $\delta < 1$. It is easy to see that

$$M_{ni} = O(h_1^2 h_2^2), \quad i = 1, 2, 3, 4$$

and

$$M_{n5} = O(h_1^{3+2\delta} h_2^{3+2\delta}) \quad \text{and} \quad M_{n6} = O(h_1^3 h_2^3).$$

Hence, we have

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

as $n \to \infty$. On other hand, since the decay rate of β-mixing coefficients is exponential, $\sum_{k=1}^{\infty} k^2 \{\beta(k)\}^{\frac{\delta}{1+\delta}} < \infty$. Hence all conditions of Lemma 6 are satisfied. By Lemma 6, we have

$$\frac{n h_1 h_2}{\sqrt{n(n-1)/2}\sigma_0} \sum_{i<j}^{n} (n-2) \varphi^*(i, j) \to \mathcal{N}(0, 1).$$

or equivalently by (B.9)

$$\frac{1}{\sigma_0^*} \sum_{i<j}^{n} (n-2) \varphi^*(i, j) \to \mathcal{N}(0, 1),$$

where $\sigma_0^{*2} = \frac{2\|w\|^2}{h_1 h_2} \|W^*\|^2 \|K^*\|^2$. This completes the proof of Theorem 1. □

Proof of Theorem 2

The proof of this theorem is similar to that of Theorem 1. We note only the differences between them.

First, in Theorem 2, the test statistic T_3 is only related to h_1, not h_2. Thus, T_3 behaves like a generalized likelihood test statistic considered in Fan *et al.* (2001) in a nonparametric regression setting. It is a univariate smoothing problem. Secondly, $K_{h_2}(Y_j - Y_i)$ in T_1 is now replaced by $I(Y_j < Y_i)$. This does not affect the key idea of the proof, but alters the calculation of the mean and variance of the asymptotic null distribution.

First of all, T_3 can be written as

$$T_3 = \sum_{i=1}^{n} \{A_i^2 + B_i^2 + 2A_iB_i\} w(X_i) \hat{=} T_{31} + T_{32} + T_{33},$$

where $P(Y_i|X_i) = P(Y_i|X_i, \theta)$,

$$A_i = \frac{1}{nh_1} \sum_{j=1}^{n} W_n \left(\frac{X_j - X_i}{h_1}, X_i \right) \{I(Y_j < Y_i) - P(Y_i|X_j)\},$$

and

$$B_i = \frac{1}{nh_1} \sum_{j=1}^{n} W_n \left(\frac{X_j - X_i}{h_1}, X_i \right) \{P(Y_i|X_j) - P(Y_i|X_i)\}.$$

Following the standard arguments in local linear fitting, it is easy to show that

$$B_i = P_x''(Y_i|X_i) h_1^2 + o_p(h_1^2) = O_p(h_1^2),$$

where $P_x''(y|x)$ is the second derivative of $P(y|x)$ with respect to x. Hence, by Condition 9, we have $T_{32} = O_p(nh_1^4) = o_p(1/\sqrt{h_1})$. Following the same proof as that used to establish (B.2), we can show that $T_{33} = O_p(\sqrt{n} h_1^2) = o_p(1/\sqrt{h_1})$. These entail that

$$T_3 = T_{31} + o_p(1/\sqrt{h_1}). \tag{S.2}$$

Thus, we need only consider T_{31}.

By similar arguments to those used in establishing (B.2), we have

$$T_{31} = \sum_{i=1}^{n} \frac{w(X_i)}{n^2 \pi^2(X_i)} \left\{ \sum_{j=1}^{n} W_{h_1}(X_j - X_i) \{I(Y_j < Y_i) - P(Y_i|X_j)\} \right\}^2 + O_p(n\rho_n/(nh_1)).$$

The second term is of order $o_p(h_1^{\frac{1}{2}})$. Hence, by (S.2), we have

$$T_3 = \sum_{i,j,k} \phi(i,j,k) + o_p(1/\sqrt{h_1}), \tag{S.3}$$

where

$$\begin{aligned}\phi(i,j,k) = & \frac{w(X_i)}{n^2\pi^2(X_i)} W_{h_1}(X_j - X_i) \\ & \times \{I(Y_j < Y_i) - P(Y_i|X_j)\} W_{h_1}(X_k - X_i)\{I(Y_k < Y_i) - P(Y_i|X_k)\}.\end{aligned}$$

The first term of (S.3) can be decomposed as the sum of T_{311}, T_{312} and T_{313}, where

$$\begin{aligned}T_{311} &= \sum_{i<j<k}^{n} \phi(i,j,k) + \phi(i,k,j) + \phi(j,i,k) + \phi(j,k,i) + \phi(k,i,j) + \phi(k,j,i), \\ T_{312} &= \sum_{i<j}^{n} \phi(i,j,j) + \phi(i,i,j) + \phi(i,j,i) + \phi(j,i,i) + \phi(j,j,i) + \phi(j,i,j),\end{aligned}$$

and $T_{313} = \sum_{i=1}^{n} \phi(i,i,i)$. Denote the typical elements of T_{311} and T_{312} by $\phi^*(i,j,k)$ and $\tilde{\phi}(i,j)$, respectively. We can follow the proof of Theorem 1 for T_{11}^*, T_{12}^* and T_{13}^* to show that

$$\begin{aligned}T_{311} &= (n-2)\sum_{i<j} \phi^*(i,j) + o_p(1/\sqrt{h_1}), \\ T_{312} &= \frac{n(n-1)}{2}\tilde{\phi}(0) + o_p(1/\sqrt{h_1}),\end{aligned}$$

and $T_{313} = o_p(1/\sqrt{h_1})$, where $\phi^*(i,j) = \int \phi^*(i,j,k)dF(x_k,y_k)$ and $\tilde{\phi}(0) = \mathrm{E}\int \tilde{\phi}(i,j)dF(x_j,y_j)$. These results along with (S.3) entail that

$$T_3 = \frac{n(n-1)}{2}\tilde{\phi}(0) + (n-2)\sum_{i<j}^{n} \phi^*(i,j) + o_p(1/\sqrt{h_1}). \tag{S.4}$$

As in the proof of Theorem 1, the first term in the right-hand side provides the asymptotic mean of the null distribution, and the second term is a $U-$statistic which converges to a normal distribution with mean zero. Hence, after computing $\frac{n(n-1)}{2}\tilde{\phi}(0)$, similar to the proof for T_{11}^* in Theorem 1, we can apply Lemma 7 to establish the asymptotic normality of $(n-2)\sum_{i<j}^{n} \phi^*(i,j)$. Since the technical proofs are similar, we

only compute $\frac{n(n-1)}{2}\tilde{\phi}(0)$ and the variance of $(n-2)\sum_{i<j}^{n}\phi^*(i,j)$.

Computation of $\tilde{\phi}(0)$***.***

Note that $\tilde{\phi}(0) = \int\{\phi(i,j,j)+\phi(j,i,i)\}dF(x_i,y_i)F(x_j,y_j)$. Then

$$\begin{aligned} n^2\tilde{\phi}(0) &= 2\int \frac{w(x_i)}{\pi^2(x_i)}W_{h_1}^2(x_j-x_i)\{I(y_j<y_i)-P(y_i|x_j)\}^2 dF(x_i,y_i)dF(x_j,y_j) \\ &= W_1+W_2+W_3, \end{aligned} \tag{S.5}$$

where

$$W_1 = 2\int \frac{w(x_i)}{\pi^2(x_i)}W_{h_1}^2(x_j-x_i)I(y_j<y_i)dF(x_i,y_i)dF(x_j,y_j),$$

$$W_2 = -4\int \frac{w(x_i)}{\pi^2(x_i)}W_{h_1}^2(x_j-x_i)I(y_j<y_i)P(y_i|x_j)dF(x_i,y_i)dF(x_j,y_j),$$

and

$$W_3 = 2\int \frac{w(x_i)}{\pi^2(x_i)}W_{h_1}^2(x_j-x_i)P^2(y_i|x_j)dF(x_i,y_i)dF(x_j,y_j).$$

Using the change of variables and Taylor's expansion, it is easy to show that

$$\begin{aligned} W_1 &= 2\int \frac{w(x_i)}{\pi^2(x_i)}W_{h_1}^2(x_j-x_i)P(y_i|x_j)\pi(x_j)dx_j dF(x_i,y_i) \\ &= \frac{2\|W\|^2}{h_1}\int w(x_i)P(y_i|x_i)dP(y_i|x_i)dx_i + O(h_1) \\ &= h_1^{-1}\|W\|^2\|\mathbf{1}\|_w^2 + O(h_1). \end{aligned} \tag{S.6}$$

Similarly, using the same technique, we also have

$$W_2 = -\frac{4}{3h_1}\cdot\|\mathbf{1}\|_w^2\|W\|^2 + O(h_1), \tag{S.7}$$

and

$$W_3 = \frac{2}{3h_1}\|\mathbf{1}\|_w^2\|W\|^2 + O(h_1). \tag{S.8}$$

Hence, the combination of (S.5)—(S.8) lead to

$$\frac{n(n-1)}{2}\tilde{\phi}(0) = \frac{1}{6h_1}\|W\|^2\|\mathbf{1}\|_w^2 + O(h_1).$$

Computing the asymptotic variance of the $U-$statistics.

As in the proof of Theorem 1, we know that the asymptotic variance of $\sum_{i<j} \phi^*(i,j)$ is

$$\sigma^{*2} = \frac{n(n-1)}{2n^2h_1^2}\sigma^2, \tag{S.9}$$

where

$$\frac{\sigma^2}{n^2h_1^2} = \int \left\{(n-2)\int \{\phi(k,i,j)+\phi(k,j,i)\}dF(x_k,y_k)\right\}^2 dF(x_i,y_i)dF(x_j,y_j). \tag{S.10}$$

Let us first consider the term inside the square sign. It can be written as the sum of four terms:

$$\begin{aligned}
&n^2\int \{\phi(k,i,j)+\phi(k,j,i)\}dF(x_k,y_k)\\
&=2\int \frac{w(x_k)}{\pi^2(x_k)}W_{h_1}(x_j-x_k)W_{h_1}(x_i-x_k)\\
&\times \{I(y_i<y_k)-P(y_k|x_i)\}\{I(y_j<y_k)-P(y_k|x_j)\}dF(x_k,y_k)\\
&\hat{=}2(Z_1+Z_2+Z_3+Z_4).
\end{aligned}$$

We now deal with each term separately. Note that

$$\begin{aligned}
Z_1 &= \int \frac{w(x_k)}{\pi(x_k)}W_{h_1}(x_j-x_k)W_{h_1}(x_i-x_k)I(y_i\vee y_j<y_k)dP(y_k|x_k)\\
&= \frac{w(\frac{x_i+x_j}{2})}{\pi(\frac{x_i+x_j}{2})}W^*_{h_1}(x_i-x_j)\{1-P(y_i\vee y_j\Big|\frac{x_i+x_j}{2})\}+O(h_1),
\end{aligned}$$

where $x\vee y=\max(x,y)$, and $W^*=W*W$. For terms Z_2 and Z_3, recalling that W has a bounded support, the integrands below do not vanish only when $|x_i-x_j|=O(h_1)$. Hence we have following expression for Z_2 and Z_3:

$$Z_2 = -\frac{w(\frac{x_i+x_j}{2})}{2\pi(\frac{x_i+x_j}{2})}W^*_{h_1}(x_i-x_j)\{1-P^2(y_i\Big|\frac{x_i+x_j}{2})\}+O(h_1),$$

and

$$Z_3 = -\frac{w(\frac{x_i+x_j}{2})}{2\pi(\frac{x_i+x_j}{2})}W^*_{h_1}(x_i-x_j)\{1-P^2(y_j\Big|\frac{x_i+x_j}{2})\}+O(h_1).$$

Similarly, we can show that

$$Z_4 = \frac{w(\frac{x_i+x_j}{2})}{3\pi(\frac{x_i+x_j}{2})} W^*_{h_1}(x_i - x_j) + O(h_1).$$

Substituting these into (S.10), we have

$$\begin{aligned}\frac{\sigma^2}{h_1^2} &= 4\int \{Z_1 + Z_2 + Z_3 + Z_4\}^2 dF(x_i, y_i)dF(x_j, y_j) + O(h_1) \\ &= 4\int \frac{w^2(\frac{x_i+x_j}{2})}{\pi^2(\frac{x_i+x_j}{2})} {W^*}^2_{h_1}(x_i - x_j)\Big\{\frac{1}{3} + \frac{1}{2}P^2(y_j\Big|\frac{x_i+x_j}{2}) + \frac{1}{2}P^2(y_i\Big|\frac{x_i+x_j}{2}) \\ &\quad - P(y_i \vee y_j\Big|\frac{x_i+x_j}{2})\Big\}^2 dF(x_i, y_i)dF(x_j, y_j) + O(h_1).\end{aligned}$$

By a change of variable and Taylor's expansion, we have

$$\begin{aligned}\frac{\sigma^2}{h_1^2} &= \frac{4\|W^*\|^2}{h_1} \\ &\times \int w^2(x_i)\left\{\frac{1}{3} + \frac{1}{2}P^2(y_j|x_i) + \frac{1}{2}P^2(y_i|x_i) - P(y_i \vee y_j|x_i)\right\}^2 dP(y_i|x_i)dP(y_j|x_i) + O(h_1).\end{aligned}$$

By a change variable $s = P(y_i|x_i)$ and $t = P(y_j|x_i)$, and noting that $P(y_i \vee y_j|x_i) = s \vee t$, we have

$$\frac{\sigma^2}{h_1^2} = \frac{4\|W^*\|^2\|w\|^2}{h_1}\int_0^1\int_0^1 \{\frac{1}{3} + \frac{1}{2}(s^2+t^2) - s \vee t\}^2 dsdt + O(h_1). \qquad \text{(S.11)}$$

The integral in (S.11) is 1/90. Hence,

$$\frac{\sigma^2}{h_1^2} = \frac{2\|W^*\|^2\|w\|^2}{45h_1} + O(h_1).$$

Following the same proof as that of the Theorem 1, after checking the conditions of Lemma 7, we have

$$\frac{nh_1}{\sqrt{n(n-1)/2}\sigma}\sum_{i<j}^{n}(n-2)\phi^*(i,j) \to \mathcal{N}(0,1).$$

Hence,

$$\frac{1}{\sigma_0^*}\sum_{i<j}^{n}(n-2)\varphi^*(i,j) \to \mathcal{N}(0,1),$$

where $\sigma_0^{*2} = \frac{1}{45h_1}\|W * W\|^2\|w\|^2$. □

Proof of Theorem 3

First of all, under H_{n1}, the true density is $p_n(Y|X) = p(Y|X) + g_n(X,Y)$. From the definition of T_1, we have

$$T_1 = \sum_{i=1}^{n} \left\{ \frac{\hat{p}_n(Y_i|X_i) - p_n(Y_i|X_i)}{p_n(Y_i|X_i)} \right\}^2 w(X_i, Y_i) + \sum_{i=1}^{n} \left\{ \frac{g_n(X_i, Y_i)}{p_n(Y_i|X_i)} \right\}^2 w(X_i, Y_i)$$
$$+ 2\sum_{i=1}^{n} g_n(X_i, Y_i)\{\hat{p}_n(Y_i|X_i) - p_n(Y_i|X_i)\} \frac{w(X_i, Y_i)}{p_n^2(Y_i|X_i)} + o_p(\frac{1}{\sqrt{h_1 h_2}}). \quad \text{(S.12)}$$

The first term in the right side can be dealt with similarly to that in the proof of Theorem 1. It is asymptotically normal with mean μ_1 and variance σ_1^2 given in Theorem 1. For the second term, since $\delta_n^2 = O\{(nh_1h_2)^{-1}\}$, it can be shown that

$$\sum_{i=1}^{n} \left\{ \frac{g_n(X_i, Y_i)}{p_n(Y_i|X_i)} \right\}^2 w(X_i, Y_i) = n\mathrm{E}\frac{g_n^2(X, Y)w(X, Y)}{p_n^2(Y|X)} + o_p(\frac{1}{\sqrt{h_1 h_2}}). \quad \text{(S.13)}$$

Hence, it remains to deal with the cross-product term. Recall the decomposition (B.1). By the standard arguments in the local linear fit, we have

$$2\sum_{i=1}^{n} \delta_n(B_i + C_i) = O_p(nh_1^2\delta_n + nh_2^2\delta_n).$$

This together with (S.12) and (S.13) lead to

$$T_1 = \sum_{i=1}^{n} \left\{ \frac{\hat{p}_n(Y_i|X_i) - p_n(Y_i|X_i)}{p_n(Y_i|X_i)} \right\}^2 w(X_i, Y_i) + n\mathrm{E}\frac{g_n^2(X, Y)w(X, Y)}{p_n^2(Y|X)}$$
$$+ 2\sum_{i=1}^{n} \frac{g_n(X_i, Y_i)A_{ni}w(X_i, Y_i)}{p_n^2(Y_i|X_i)} + o_p(\frac{1}{\sqrt{h_1 h_2}}). \quad \text{(S.14)}$$

We now deal with the third term in (S.14). Argued the same way as that used in the proof of (B.2), we have

$$\sum_{i=1}^{n} \frac{g_n(X_i, Y_i)A_{ni}w(X_i, Y_i)}{p_n^2(Y_i|X_i)}$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{g_n(X_i, Y_i)w(X_i, Y_i)}{np_n^2(Y_i|X_i)\pi(X_i)} W_{h_1}(X_j - X_i)\{K_{h_2}(Y_j - Y_i) - m(X_j, Y_i)\} + o_p\left(\frac{1}{\sqrt{h_1 h_2}}\right)$$
$$= \sum_{i \neq j}^{n} \frac{g_n(X_i, Y_i)w(X_i, Y_i)}{np_n^2(Y_i|X_i)\pi(X_i)} W_{h_1}(X_j - X_i)\{K_{h_2}(Y_j - Y_i) - m(X_j, Y_i)\}$$
$$+ O(\delta_n h_1^{-1} h_2^{-1}) + o_p\left(\frac{1}{\sqrt{h_1 h_2}}\right).$$

The first term is a $U-$statistic with a typical element denoted by $\psi(i,j)$. Let $\psi^*(i,j)=\psi(i,j)+\psi(j,i)$ be a symmetric kernel. With the notation $\tilde{\psi}(i)=\int\psi^*(i,j)dF(x_j,y_j)$ and $\tilde{\psi}(i,j)=\psi^*(i,j)-\tilde{\psi}(i)-\tilde{\psi}(j)$, we have the following Hoeffding decomposition:

$$\sum_{i\neq j}^{n}\psi(i,j)=\sum_{i<j}^{n}\tilde{\psi}(i,j)+(n-1)\sum_{i=1}^{n}\tilde{\psi}(i). \tag{S.15}$$

It is easy to verify that $\mathrm{E}\{h_1h_2\tilde{\psi}(i,j)\}^{2(1+\delta)}=O(\delta_n^{2(1+\delta)}n^{-2(1+\delta)}h_1h_2)$. Hence, by Lemma 6 (ii) with $\delta=1$, we have that

$$\mathrm{E}\left\{\sum_{i<j}^{n}\tilde{\psi}(i,j)\right\}^2=o(h_1^{-1}h_2^{-1}). \tag{S.16}$$

For $(n-1)\sum\limits_{i=1}^{n}\tilde{\psi}(i)$, it is easy to show that $\mathrm{E}\tilde{\psi}(i)=0$. By employing the central limit theorem for β-mixing process, it can be shown that

$$\frac{(n-1)}{2\sigma_{1A}}\sum_{i=1}^{n}\tilde{\psi}(i)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1), \tag{S.17}$$

where $\sigma_{1A}^2=n\mathrm{E}(n-1)^2\tilde{\psi}^2(i)/4$. We now compute σ_{1A}^2. From the definition of $m(X_i,Y_i)$ and the Fubini theorem, we have

$$\begin{aligned}\tilde{\psi}(i)&=\frac{2}{n}\int\frac{g_n(X_j,Y_j)w(X_j,Y_j)}{p_n^2(Y_j|X_j)\pi(X_j)}W_{h_1}(X_j-X_i)\{K_{h_2}(Y_j-Y_i)-m(X_i,Y_j)\}dF(x_j,y_j)\\&=\frac{2}{n}\frac{g_n(X_i,Y_i)w(X_i,Y_i)}{p_n(Y_i|X_i)}-\mathrm{E}\left\{\frac{2\delta_n}{n}\frac{g(X_i,Y_i)w(X_i,Y_i)}{p_n(Y_i|X_i)}\Big|X_i\right\}+O(\delta_nh_1^2/n+\delta_nh_2^2/n).\end{aligned}$$

Hence

$$\sigma_{1A}^2=n\mathrm{E}\frac{g_n^2(X_i,Y_i)w^2(X_i,Y_i)}{p_n^2(Y_i|X_i)}-n\mathrm{E}\left[\mathrm{E}\left\{\frac{g_n(X_i,Y_i)w^2(X_i,Y_i)}{p_n(Y_i|X_i)}\Big|X_i\right\}\right]^2+o(1/h_1h_2).$$

Using the results in the proof of Theorem 1, it can be shown that that term $\sum_i\tilde{\psi}(i)$ and $\sum_i\varphi^*(i)$ are asymptotically uncorrelated and asymptotically jointly normal. By combining the results obtained above, we complete the proof. □

Proof of Theorem 4

With the rates specified in the theorem,

$$d_{1n}/\sigma_1^*=d^2b+dO(1)$$

where b and $O(1)$ are two constants, independent of d. For any given small η, when d is small enough, $|d_{1n}/\sigma_1^*| \le \eta$ and $\sigma_1^* = \sigma_1\{1 + o(1)\}$. First of all, since we have dropped the assumption $n(h_1^5 + h_2^5) \to 0$, inspecting the proof of Theorem 1, we have to add a term of order $O_p(n^{1/5})$ to T_1 to reflect the non-negligible terms such as T_2^* and T_3^*, which is of order $O_p\{n(h_1^4 + h_2^4)\} = O_P(n^{1/5})$. In other words, under the null hypothesis, with the selected bandwidths,

$$(T_1 - \mu_1)/\sigma_1 = O_P(1).$$

Hence, the sequence of the critical value c_α is bounded.

Similarly, since the assumption $n(h_1^5 + h_2^5) \to 0$ has been dropped, Theorem 3 now becomes

$$(T_1 - \mu_1 - d_{1n})/\sigma_1^* = O_P(1). \tag{S.18}$$

Now the power at the alternative $p_n(y|x)$ is given by

$$\begin{aligned} P\left\{\frac{T_1 - \mu_1}{\sigma_1} > c_\alpha \middle| H_{1n}\right\} &= P\left\{\frac{T_1 - \mu_1 - d_{1n}}{\sigma_1^*} > \frac{c_\alpha \sigma_1}{\sigma_1^*} - \frac{d_{1n}}{\sigma_1^*} \middle| H_{1n}\right\} \\ &\le P\left\{\frac{T_1 - \mu_1 - d_{1n}}{\sigma_1^*} > \frac{c_\alpha \sigma_1}{\sigma_1^*} - \eta \middle| H_{1n}\right\}. \end{aligned}$$

By Slutsky's theorem and Theorem 3, we have

$$\limsup_{d\to 0} \limsup_{n\to\infty} P\left\{\frac{T_1 - \mu_1}{\sigma} > c_\alpha \middle| H_{1n}\right\} \le \alpha,$$

For any given M, by taking the constant d sufficiently large, when n is large enough,

$$d_{1n}/\sigma_1^* \ge M.$$

Thus, we have

$$P\left\{\frac{T_1 - \mu_1}{\sigma_1} > c_\alpha \middle| H_{1n}\right\} \ge P\left\{\frac{T_1 - \mu_1 - d_{1n}}{\sigma_1^*} > c_\alpha - M \middle| H_{1n}\right\}.$$

By (S.18), the above random variables are tight. Hence, we have

$$\liminf_{M\to\infty} \liminf_{n\to\infty} P\left\{\frac{T_1 - \mu_1}{\sigma} > c_\alpha - M \middle| H_{1n}\right\} = 1.$$

This completes the proof of Theorem 4. $\square$

Proof of Theorem 5

First of all, decompose

$$
\begin{aligned}
T_3 &= \sum_{i=1}^{n}\{\hat{P}_n(Y_i|X_i) - P(Y_i|X_i)\}^2 w(X_i) \\
&= \sum_{i=1}^{n}\{\hat{P}_n(Y_i|X_i) - P_n(Y_i|X_i)\}^2 w(X_i) + \sum_{i=1}^{n} G_n^2(X_i, Y_i) w(X_i) \\
&+ 2\sum_{i=1}^{n} G_n(X_i, Y_i)\{\hat{P}_n(Y_i|X_i) - P_n(Y_i|X_i)\}. \qquad \text{(S.19)}
\end{aligned}
$$

The first term, as shown in Theorem 2, follows asymptotically the normal distribution with mean μ_3 and variance σ_3^2. Since $\delta_n^2 \asymp (nh_1)^{-1}$, the second term can be approximated as

$$
\sum_{i=1}^{n} G_n^2(X_i, Y_i) w(X_i) = n\mathrm{E}G_n^2(X, Y) w(X) + o_p(1/\sqrt{h_1}).
$$

To deal with the third term, let us decompose, as in the proof of Theorem 2,

$$
\hat{P}_n(Y_i|X_i) - P_n(Y_i|X_i) = A_{ni} + B_{ni}.
$$

Then, using conventional arguments in the local linear regression, we have

$$
2\sum_{i=1}^{n} G_n(X_i, Y_i) B_{ni} = O_p(nh_1^2\delta_n) = o_P(1/\sqrt{h_1}).
$$

For the term involving A_{ni}, we have

$$
\begin{aligned}
&\sum_{i=1}^{n} G_n(X_i, Y_i) A_{ni} w(X_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{G_n(X_i, Y_i) w(X_i)}{nf(X_i)} W_{h_1}(X_j - X_i)\{I(Y_j < Y_i) - P_n(Y_i|X_j)\} + o_p(1/\sqrt{h}) \\
&= \sum_{i \neq j}^{n} \frac{G_n(X_i, Y_i) w(X_i)}{nf(X_i)} W_{h_1}(X_j - X_i)\Big\{I(Y_j < Y_i) - P_n(Y_i|X_j)\Big\} + O(\delta_n h_1^{-1}) + o(1/\sqrt{h_1}).
\end{aligned}
$$

(S.20)

Similar to the proof of Theorem 3, the first term is a $U-$statistic with zero mean, which can be shown to be asymptotically normal with mean zero and variance $4\sigma_{3A}^2$. Since the proof is nearly identical to that in the proof of Theorem 3, we only compute σ_{3A}^2 here.

Let $\Psi(i,j)$ be the typical element of (S.20). As the proof of Theorem 3, we also define that $\Psi^*(i,j) = \Psi(i,j) + \Psi(j,i)$ and $\Psi^*(i) = \int \Psi^*(i,j)dF(x_j, y_j)$. Then.

$$\begin{aligned}\sigma_{3A}^2 &= n\mathrm{E}\left\{\int G_n(X_i, Y_j)w(X_i)I(Y_i < Y_j)dP(Y_j|X_i)\right\}^2 \\ &\quad - n\left\{\mathrm{E}\int G_n(X_i, Y_j)w(X_i)P(Y_j|X_i)dP(Y_j|X_i)\right\}^2.\end{aligned}$$

Since $\mathrm{E}\Psi^*(i) = 0$, and this $U-$statistic in (S.20) is asymptotically independent of the $U-$statistic in the first term of (S.19). Consequently, we obtain the resulting normality as shown. □

Technical lemmas

The following two lemmas are extension of a uniform convergence result of Mack and Sliverman (1982) to geometrically mixing processes. The proofs are similar to those of Theorem 5.3 and Lemma 5.1 in Fan and Yao (2003).

Lemma 1. *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a stationary sequence satisfying the mixing condition $|\beta(l)| = O(e^{-\lambda l})$ for some $\lambda > 0$ with an invariant density $p(x, y)$. Then, under Condition 2, we have*

$$\begin{aligned}&\sup_{(x,y)\in\Omega}\left|n^{-1}\sum_{t=1}^{n}\{W_{h_1}(X_t - x)K_{h_2}(Y_t - y) - \mathrm{E}[W_{h_1}(X_t - x)K_{h_2}(Y_t - y)]\}\right| \\ &= O_p[\{nh_1h_2/\log(n)\}^{-1/2}],\end{aligned}$$

provided that $h_1 \to 0$ and $h_2 \to 0$ in such a way that $nh_1h_2/\log n \to \infty$, where Ω is a compact set over which $p(x, y)$ is bounded and Lipschiz continuous.

Lemma 2. *Under the conditions of Lemma 1 and Condition 2, if the marginal density $\pi(x)$ of X and the function $m(x, y)$ are Lipschitz in x for all (x, y) on a compact set Ω, then we have*

$$\sup_{(x,y)\in\Omega} \left| n^{-1} \sum_{t=1}^{n} \{W_h(X_t - x)m(X_t, y) - \mathrm{E}[W_h(X_t - x)m(X_t, y)]\} \right| = O_p[\{nh/\log(n)\}^{-1/2}],$$

provided that $h \to 0$ and $nh/\log n \to \infty$.

The following lemma is the uniform convergence result for the local linear fit of the conditional density. The proof is similar to those for Theorems 5.3 and 6.5 of Fan and Yao (2003).

Lemma 3. *For the local linear fit of the conditional density, under Conditions 2, 4 and 5, we have*

$$\sup_{(x,y)\in\Omega^*} |\hat{p}(y|x) - p(y|x) - \theta_{n,0}| = O_p(\{nh_1h_2/\log(n)\}^{-1/2}),$$

provided that h_1 and h_2 converge to zero in such a way that $nh_1h_2/\log n \to \infty$, where

$$\theta_0(x, y) = \frac{h_1^2\mu_2}{2}\frac{\partial^2 f(y|x)}{\partial x^2} + \frac{h_2^2\mu_K}{2}\frac{\partial^2 f(y|x)}{\partial y^2} + o(h_1^2 + h_2^2).$$

The following lemma is a part of Lemma 1.3 of Bosq (1998).

Lemma 4. *If (X, Y) has a absolutely continuous distribution with respect to the Lebesgue measure on $\mathcal{R}^{2d}$ and*

$$g_{(X,Y)}(x, y) = f_{(X,Y)}(x, y) - f_X(x)f_Y(y); \quad x, y \in \mathcal{R}^d$$

satisfies Lipschitz's condition

$$|g(x', y') - g(x, y)| \le C(\|x' - x\|^2 + \|y' - y\|^2)^{1/2}$$

for some constant C, then there exists a constant $\gamma(d, C)$ such that

$$\|g\|_\infty \le \gamma(d, C)\alpha^{1/(2d+1)},$$

where $\alpha = \alpha(\sigma(X), \sigma(Y)) = \sup_{B\in\sigma(X),C\in\sigma(Y)} |P(B \cap C) - P(B)P(C)|$ and $\sigma(X), \sigma(Y)$ are the sigma fields generated by X and Y respectively.

The following lemma, due to Lee (1990) (see Lemma B there), is often used for $U-$statistics of weakly dependent stationary sequences.

Lemma 5. *Let $t_1 < t_2 < \cdots < t_k$, $\{X_i\}$ be a stationary sequence, and F, G_j and H_j be the distribution function of $(X_{t_1}, \ldots, X_{t_k}), (X_{t_1}, \ldots, X_{t_j})$ and $(X_{t_{j+1}}, \ldots, X_{t_k})$ respectively. Then for any measurable function h, we have*

$$\left|\int hdF - \int\int hdG_j dH_j\right| \leq 3M^{\frac{1}{1+\delta}}\beta^{\frac{\delta}{1+\delta}}(t_{j+1} - t_j).$$

provided that for some $\delta > 0$

$$M = \max\left(\int |h|^{1+\delta}dF, \int\int |h|^{1+\delta}dG_j dH_j\right) < \infty.$$

The following two lemmas are proved in Hjellvik *et al.* (1998). Gao and King (2004) have extended those results to α-mixing sequences.

Lemma 6. *(i) Let $\varphi(\cdot,\cdot,\cdot)$ be a symmetric Borel function defined on $R^p \times R^p \times R^p$. Assume that for any constants $x, y \in R^p, \mathrm{E}\{\varphi(\xi_1, x, y)\} = 0$. Then, there exists a constant c such that*

$$\mathrm{E}\left\{\sum_{1\leq i<j<k\leq n}\varphi(\xi_i,\xi_j,\xi_k)\right\}^2 \leq cn^3 \max\left\{M^{\frac{1}{1+\delta}}\sum_{k=1}^{n}k^2\beta^{\frac{\delta}{1+\delta}}(k), \max_{j>i>1}\mathrm{E}[\varphi(\xi_1,\xi_i,\xi_j)]^2\right\},$$

where $\delta > 0$ is a constant for which

$$M = \max_{1\leq i<j\leq n}\max\Big\{\mathrm{E}|\varphi(\xi_1,\xi_i,\xi_j)|^{2(1+\delta)}, \int |\varphi(\xi_1,\xi_i,\xi_j)|^{2(1+\delta)}dP(\xi_1)dP(\xi_i,\xi_j),$$
$$\int |\varphi(\xi_1,\xi_i,\xi_j)|^{2(1+\delta)}dP(\xi_1,\xi_i)dP(\xi_j), \int |\varphi(\xi_1,\xi_i,\xi_j)|^{2(1+\delta)}dP(\xi_1)dP(\xi_i)dP(\xi_j)\Big\}.$$

(ii) Let $\varphi(\cdot,\cdot)$ be a symmetric Borel function defined in $R^p \times R^p$. Assume that for any constant $x \in R^p$, $\mathrm{E}\{\varphi(\xi_1, x)\} = 0$. Then, there exists a constant c such that

$$\mathrm{E}\left\{\sum_{1\leq i<j\leq n}\varphi(\xi_i,\xi_j)\right\}^2 \leq cn^2 \max\left\{M^{\frac{1}{1+\delta}}\sum_{k=1}^{n}k\beta^{\frac{\delta}{1+\delta}}(k), \max_{i>1}\mathrm{E}[\varphi(\xi_1,\xi_i)]^2\right\},$$

where $\delta > 0$ is a constant for which

$$M = \max_{1<i\leq n}\max\Big\{\mathrm{E}|\varphi(\xi_1,\xi_i)|^{2(1+\delta)}, \int |\varphi(\xi_1,\xi_i)|^{2(1+\delta)}dP(\xi_1)dP(\xi_i)\Big\}.$$

Suppose that $\varphi_n(\cdot,\cdot)$ is a symmetric Borel function defined on $R^p \times R^p$, which may depend on the sample size n. Assume further that there exists a sequence of σ-algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ for which $\xi_j \in \mathcal{F}_j$, and

$$\text{(i) } \mathrm{E}\{\varphi_n(x,\xi_1)\} = 0 \quad \text{for any } x \in R^p,$$
$$\text{(ii) } \mathrm{E}\{\varphi_n(\xi_i,\xi_j)|\mathcal{F}_{j-1}\} = 0, \quad \text{for any } i < j.$$

Let $\varphi_{ij} = \varphi_n(\xi_i,\xi_j), \sigma_{ij}^2 = \mathrm{Var}(\varphi_{ij})$, and $\sigma_n^2 = \sum\limits_{1\le i<j\le n} \sigma_{ij}^2$. For some constant $\delta > 0$, define

$$M_{n1} = \max_{i<j} \max\left\{ \mathrm{E}|\varphi_{1j}\varphi_{ij}|^{1+\delta}, \int |\varphi_{1j}\varphi_{ij}|^{1+\delta} dP(\xi_1)dP(\xi_i,\xi_j) \right\},$$
$$M_{n2} = \max_{i<j} \max\Big\{ \mathrm{E}|\varphi_{1j}\varphi_{ij}|^{2(1+\delta)}, \int |\varphi_{1j}\varphi_{ij}|^{2(1+\delta)} dP(\xi_1)dP(\xi_i,\xi_j),$$
$$\int |\varphi_{1j}\varphi_{ij}|^{2(1+\delta)} dP(\xi_1,\xi_i)dP(\xi_j), \int |\varphi_{1j}\varphi_{ij}|^{2(1+\delta)} dP(\xi_1)dP(\xi_i)dP(\xi_j)\Big\},$$

$$M_{n3} = \max_{i<j} \mathrm{E}|\varphi_{1j}\varphi_{ij}|^2, \quad M_{n4} = \max_{i,j,k \text{ different}} \left\{ \max_P \int |\varphi_{1j}\varphi_{ij}|^{2(1+\delta)} dP \right\},$$
$$M_{n5} = \max_{i<j} \max\left\{ \mathrm{E}\Big| \int \varphi_{1j}\varphi_{1j} dP((x_1,y_1)) \Big|^{2(1+\delta)}, \int \Big| \int \varphi_{1j}\varphi_{1j} dP(\xi_1) \Big|^{2(1+\delta)} dP(\xi_i))dP(\xi_j) \right\},$$
$$M_{n6} = \max_{i<j} \mathrm{E}\Big| \int \varphi_{1j}\varphi_{1j} dP(\xi_1) \Big|^2.$$

where the maximization on P in the equation for M_{n4} is taken over the four probability measures
$P(\xi_1,\xi_i,\xi_j,\xi_k), P(\xi_1)P(\xi_i,\xi_j,\xi_k), P(\xi_1)P(\xi_{i_1})P(\xi_{i2},\xi_{i3})$, and $P(\xi_1)P(\xi_i)P(\xi_j)P(\xi_k)$, where (i_1,i_2,i_3) is the permutation of (i,j,k) in the ascending order.

Lemma 7. *If for some $\delta > 0, \sum_{k=1}^{\infty} k^2\{\beta(k)\}^{\frac{\delta}{1+\delta}} < \infty$, and*

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

as $n \to \infty$, then $\sigma_n^{-1} \sum\limits_{1\le i<j\le n} \varphi(\xi_i,\xi_j)$ is asymptotically normal with mean value 0 and variance 1.

References

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes, Lecture Notes in Statistics 110*, second edn, New York: Springer.

Fan, J., and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer-Verlag.

Fan, J., Zhang, C., and Zhang, J. (2001), “Generalized likelihood ratio statistics and Wilks phenomenon,” *The Annals of Statistics*, 29, 153-193.

Gao, J., and King, M. (2004), “Model specification testing in nonparametric and semi- parametric time series econometrics,” Technical report, The University of Western Australia.

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

Lee, A. J. (1990),*U-statistics, theory and practice*, New York: Marcel Dekker.

Mack, Y., and Sliverman, B. (1982), “Weak and strong uniform consistency of kernel regression and density estimates”, *Zeitschriftfür Wahrsheinlichkeitstheorie und verwandte Gebiete*, 61, 405-415.