The Annals of Statistics
2010, Vol. 38, No. 5, 3129–3163
DOI: 10.1214/09-AOS763

NONPARAMETRIC TESTS OF THE MARKOV HYPOTHESIS IN CONTINUOUS-TIME MODELS[1]

BY YACINE AÏT-SAHALIA, JIANQING FAN AND JIANCHENG JIANG

Princeton University and NBER, Princeton University and University of North Carolina at Charlotte

We propose several statistics to test the Markov hypothesis for β-mixing stationary processes sampled at discrete time intervals. Our tests are based on the Chapman–Kolmogorov equation. We establish the asymptotic null distributions of the proposed test statistics, showing that Wilks's phenomenon holds. We compute the power of the test and provide simulations to investigate the finite sample performance of the test statistics when the null model is a diffusion process, with alternatives consisting of models with a stochastic mean reversion level, stochastic volatility and jumps.

1. Introduction. Among stochastic processes, those that satisfy the Markov property represent an important special case. The Markov property restricts the effective size of the filtration that governs the dynamics of the process. In a nutshell, only the current value of X is relevant to determine its future evolution. This restriction simplifies model-building, forecasting and time series inference. Can it be tested on the basis of discrete observations? It is not practical to approach the testing problem in the form of a restriction on the filtration, the size of any alternative filtration being essentially unrestricted. Furthermore, the continuous-time filtration is not observable on the basis of discrete observations, especially if we do not have high-frequency data, and asymptotically the sampling interval remains fixed.

Instead, we propose to test the Markov property at the level of the discrete-frequency transition densities of the process. Given a time-homogeneous stochastic process $X = \{X_t\}_{t\geq 0}$ on $\mathbb{R}^m$, with the standard probability space $(\Omega; \mathcal{F}; P)$ and filtration $\mathcal{F}_t \subset \mathcal{F}$, we consider families of conditional probability functions $P(\cdot|x, \Delta)$ of $X_{t+\Delta}$ given $X_t = x$: for each Borel measurable function ψ, $E[\psi(X_{t+\Delta})|\mathcal{F}_t] = \int \psi(y)P(dy|X_t, \Delta)$.

If X is time-homogeneous Markovian, then its transition densities satisfy the Chapman–Kolmogorov equation

$$P(\cdot|x, \Delta + \tau) = \int_S P(\cdot|y, \Delta)P(dy|x, \tau) \tag{1}$$

Received June 2009; revised October 2009.

[1]Supported in part by NSF Grants SES-0850533, DMS-05-32370 and DMS-09-06482.

AMS 2000 subject classifications. Primary 62G10, 60J60; secondary 62G20.

Key words and phrases. Markov hypothesis, Chapman–Kolmogorov equation, locally linear smoother, transition density, diffusion.

for all $\Delta > 0$ and $\tau > 0$ and x in the support S of X. Suppose that we collect n observations on X on $[0, T]$ sampled every Δ units of time. We will assume that Δ is fixed; asymptotics are therefore with $T \to \infty$. High-frequency asymptotics, by contrast, assume that $\Delta \to 0$, and T can be fixed or T diverges. This asymptotic setup could have been considered, but it is not necessary here as we are able to test the hypothesis on the basis of discrete data at a fixed interval with no requirement for high-frequency data; high-frequency asymptotics would, of course, also generate different asymptotic properties for the tests we propose.

If we set $\tau = \Delta$ in (1), then we can estimate the transition densities at the desired frequencies on the basis of these discrete observations. On the left-hand side of the equation, the transition density at interval 2Δ can be estimated simply by retaining every other observation in the same data sample. To avoid unnecessary restrictions on the data-generating process, we will employ nonparametric estimators of the transition densities. Given these, equation (1) then becomes a testable implication of the Markov property for X.

Conversely, Kolmogorov's construction (see, e.g., [28], Chapter III, Theorem 1.5) allows one to parameterize Markov processes using transition functions. Namely, given a transition function P and a probability measure π on $\mathbb{R}^m$ serving as the initial distribution, there exists a unique probability measure such that the coordinate process X is Markovian with respect to $\sigma(X_u, u \leq t)$, has transition function P and X_0 has π as its distribution. When π is the invariant probability measure of P, the process is a stationary Markov process. Therefore, given an initial distribution, a Markov process X is determined by its transition densities.

Transition densities play a crucial role in many contexts. In mathematical finance, arbitrage considerations in finance make many pricing problems linear; as a result, they depend upon the computation of conditional expectations for which knowledge of the transition function is essential. Also, inference strategies relying on maximum-likelihood or Bayesian methods require the transition density of the process. Specification testing procedures for stochastic processes also make use of the transition densities (see, e.g., [1, 3, 7, 8, 18] and [24]). All these models, estimation methods and tests assume that the process is Markovian.

Stochastic volatility models are a very broad class of non-Markovian models, due to the latency of the volatility state variable. They have been popular in financial asset pricing and modeling (see, e.g., [17]). Parameters in stochastic volatility models are much harder to estimate and the associated pricing formulas are also different from those based on Markovian diffusion models and depend on the assumptions made on the correlation structure between the innovations to prices and volatility (as in, e.g., [23]). Other examples include models for the term structure of interest rates, which may be Markovian or not (see, e.g., [22]), and, in fact, one popular approach in mathematical finance consists of restricting term structure models to be Markovian (see, e.g., [6]). In other words, many financial econometrics models are based on the Markovian assumption and this fundamental assumption needs to be tested before they can be applied. In all these cases, testing whether the underlying process is Markovian is essential in helping to decide which family of models to use and whether a diffusion model is adequate.

We will propose test statistics for this purpose. Asymptotic null distributions of test statistics are established and we show that Wilks's phenomenon holds for several of those test statistics. The power functions of the tests are also computed for contiguous alternatives. We find that the proposed tests can detect alternatives with an optimal rate in the context of nonparametric testing procedures.

The remainder of the paper is organized as follows. In Section 2, we briefly describe the nonparametric estimation of the transition functions of the process. In Section 3, we propose several test statistics for checking the Markov hypothesis. In Section 4, we establish their asymptotic null distributions and compute their power. Simulation results are reported in Section 5. Technical conditions and proofs of the mathematical results are given in Section 6.

2. Nonparametric estimation of the transition density and distribution functions. To estimate nonparametrically the transition density of observed process X, we use the locally linear method suggested by [14]. The process X is sampled at regular time points $\{i\Delta, i = 1, \dots, n+2\}$. We make the dependence on the transition function and related quantities on Δ implicit by redefining

$$X_i = X_{i\Delta}, \qquad i = 1, \dots, n+2,$$

which is assumed to be a stationary and β-mixing process.

For ease of exposition, we describe the estimation of the transition density and distribution when $m = 1$, that is, X is a process on the line. We also define $Y_i = Y_{i\Delta} = X_{(i+1)\Delta}$ and $Z_i = Z_{i\Delta} = X_{(i+2)\Delta}$. Let b_1 and b_2 denote two bandwidths and K and W two kernel functions. Observe that as $b_2 \to 0$

$$E[K_{b_2}(Z_i - z)|Y_i = y] \approx p(z|y, \Delta), \tag{2}$$

where $K_{b_2}(z) = K(z/b_2)/b_2$ and $p(z|y, \Delta)$ is the transition density of $X_{(i+1)\Delta}$ given $X_{i\Delta}$. The left-hand side of (2) is the regression function of the random variable $K_{b_2}(Z_i - z)$ given $Y_i = y$. Hence, locally linear fit can be used to estimate this regression function. For each given x, one minimizes

$$\sum_{i=1}^{n} \{K_{b_2}(Z_i - z) - \alpha - \beta(Y_i - y)\}^2 W_{b_1}(Y_i - y) \tag{3}$$

with respect to the the local parameters α and β, where $W_{b_1}(z) = W(z/b_1)/b_1$. The resulting estimate of the conditional density is simply $\hat{\alpha}$. The estimator can be explicitly expressed as

$$\hat{p}(z|y, \Delta) = n^{-1} \sum_{i=1}^{n} W_n(Y_i - y, y; b_1) K_{b_2}(Z_i - z), \tag{4}$$

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

$$W_n(z, y; b_1) = W_{b_1}(z) \frac{s_{n,2}(y) - b_1^{-1} z s_{n,1}(y)}{s_{n,0}(y) s_{n,2}(y) - s_{n,1}^2(y)},$$

where

$$s_{n,j}(y) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{Y_i - y}{b_1} \right)^j W_{b_1}(Y_i - y).$$

Note that the effective kernel W_n depends on the sampling data points and the location y. This is the key to the design adaptation and location adaptation property of the locally linear fit.

From (4), a possible estimate of the transition distribution $P(z|y, \Delta) = P(Z_i < z|Y_i = y, \Delta)$ is given by

$$\hat{P}(z|y, \Delta) = \int_{-\infty}^{z} \hat{p}(t|y, \Delta)\, dt = \frac{1}{n} \sum_{i=1}^{n} W_n(Y_i - y, y; b_1) \bar{K}\left(\frac{Z_i - z}{b_2} \right),$$

where $\bar{K}(u) = \int_u^\infty K(t)\, dt$. Let $b_2 \to 0$, then

$$\hat{P}(z|y, \Delta) = \frac{1}{n} \sum_{i=1}^{n} W_n(Y_i - y, y; b_1) I(Z_i < z), \tag{5}$$

where we drop the term in which $Z_i = z$ would contribute the value $\bar{K}(0)$. This does not affect the asymptotic property of $\hat{P}$. Actually, (5) is really the locally linear estimator of the regression function

$$P(z|y, \Delta) = E[I(Z_i < z)|Y_i = y].$$

3. Nonparametric tests for the Markov hypothesis in discretely sampled continuous-time models. The tests we propose are based on the fact that, for X to be Markovian, its transition function must satisfy the Chapman–Kolmogorov equation in the form for densities equivalent to (1),

$$p(z|x, 2\Delta) = r(z|x, 2\Delta), \tag{6}$$

where

$$r(z|x, 2\Delta) \equiv \int_{y \in S} p(z|y, \Delta) p(y|x, \Delta)\, dy \tag{7}$$

for all $(x, z) \in S^2$.

Under time-homogeneity of the process X, the Markov hypothesis can then be tested in the form H_0 against H_1, where

$$\begin{cases} H_0 : p(z|x, 2\Delta) - r(z|x, 2\Delta) = 0 & \text{for all } (x, z) \in S^2, \\ H_1 : p(z|x, 2\Delta) - r(z|x, 2\Delta) \neq 0 & \text{for some } (x, z) \in S^2. \end{cases} \tag{8}$$

This test corresponds to a nonparametric null hypothesis versus a nonparametric alternative hypothesis.

Both $p(y|x,\Delta)$ and $p(z|x,2\Delta)$ can be estimated from data sampled at interval Δ, thanks to time homogeneity. In fact, the successive pairs of observed data $\{(X_i, Y_i)\}_{i=1}^{n+1}$ form a sample from the distribution with conditional density $p(y|x,\Delta)$ from which the estimator $\hat{p}(y|x,\Delta)$ can be constructed, and then $\tilde{r}(z|x,2\Delta)$ as indicated in equation (7) can be computed. Meanwhile, the successive pairs $(X_1, Z_1), (X_2, Z_2), \ldots,$ form a sample from the distribution with conditional density $p(z|x,2\Delta)$ which can be used to form the direct estimator by drawing a parallel to (4)

$$\hat{p}(z|x,2\Delta) = \frac{1}{n}\sum_{i=1}^{n} W_n(X_i - x, x; h_1) K_{h_2}(Z_i - z),$$

where h_1 and h_2 are two bandwidths, localizing, respectively, the x- and z-domain.

In other words, the test compares a direct estimator of the 2Δ-interval conditional density, $\hat{p}(z|x,2\Delta)$, to an indirect estimator of the 2Δ-interval conditional density, $\tilde{r}(z|x,2\Delta)$, obtained by (7). If the process is actually Markovian, then the two estimates should be close (for some distance measure) in a sense made precise by the use of the statistical distributions of these estimators.

If, instead of 2Δ transitions, we test the replicability of $j\Delta$ transitions, where j is an integer greater than or equal to 2, there is no need to explore all the possible combinations of these $j\Delta$ transitions in terms of shorter ones $(1, j-1), (1, j-2), \ldots$: verifying equation (6), for one combination is sufficient as can be seen by a recursion argument. In the event of a rejection of H_0 in (8), there is no need to consider transitions of order j. In general, a vector of "transition equalities" can be tested in a single pass in a method of moments framework with as many moment conditions as transition intervals.

We propose two classes of tests for the hypothesis problem (8) based on nonparametric estimation of the transition densities and distributions. To be more specific, since

$$r(z|x,2\Delta) = E[p(z|Y_i,\Delta)|X_i = x], \tag{9}$$

the function $r(z|x,2\Delta)$ can also be estimated by regressing nonparametrically $\hat{p}(z|Y_i,\Delta)$ on X_i. This avoids integration in (7) and makes implementation and theoretical studies easier. Employing the local linear smoother for (9), we obtain the following estimator:

$$\hat{r}(z|x,2\Delta) = n^{-1}\sum_{i=1}^{n} W_n(X_i - x, x, h_3)\hat{p}(z|Y_i,\Delta),$$

where h_3 is a bandwidth in this smoothing problem. Under H_0 in (8), the logarithm of the likelihood function is estimated as

$$\ell(H_0) = \sum_{i=1}^{n} \log \hat{r}(Z_i|X_i, 2\Delta),$$

after ignoring the initial stationary density $\pi(X_1)$. This likelihood can be compared with

$$\ell(H_1) = \sum_{i=1}^{n} \log \hat{p}(Z_i|X_i, 2\Delta),$$

which leads to the generalized likelihood ratio (GLR) test statistic (see [16])

$$\sum_{i=1}^{n} \log\{\hat{r}(Z_i|X_i, 2\Delta)/\hat{p}(Z_i|X_i, 2\Delta)\}.$$

Since the nonparametric regression functions cannot be estimated well when (X_i, Z_i) is in the boundary region, the above GLR test statistic is reduced to

$$T_0 = \sum_{i=1}^{n} \log\{\hat{r}(Z_i|X_i, 2\Delta)/\hat{p}(Z_i|X_i, 2\Delta)\} w^*(X_i, Z_i),$$

where w^* is a weight function selected to reduce the influences of the unreliable estimates in the sparse region. Admittedly, $\ell(H_1)$ is not the estimated log-likelihood under H_1 in (8), but is used to create a discrepancy measure. To see this, note that under H_0, $\hat{p}$ and $\hat{r}$ are approximately the same. By Taylor's expansion, we have

$$\begin{aligned} T_0 \approx & \sum_{i=1}^{n} \frac{\hat{p}(Z_i|X_i, 2\Delta) - \hat{r}(Z_i|X_i, 2\Delta)}{\hat{p}(Z_i|X_i, 2\Delta)} w^*(X_i, Z_i) \\ & + \frac{1}{2} \sum_{i=1}^{n} \left\{ \frac{\hat{p}(Z_i|X_i, 2\Delta) - \hat{r}(Z_i|X_i, 2\Delta)}{\hat{p}(Z_i|X_i, 2\Delta)} \right\}^2 w^*(X_i, Z_i). \end{aligned}$$

To avoid unnecessary technicalities, we ignore the first term and consider the second term

$$T_1^* = \sum_{i=1}^{n} \left\{ \frac{\hat{p}(Z_i|X_i, 2\Delta) - \hat{r}(Z_i|X_i, 2\Delta)}{\hat{p}(Z_i|X_i, 2\Delta)} \right\}^2 w^*(X_i, Z_i), \tag{10}$$

which is the χ^2-type of test statistics. A natural alternative statistic to T_1^* is

$$T_1 = \sum_{i=1}^{n} \{\hat{p}(Z_i|X_i, 2\Delta) - \hat{r}(Z_i|X_i, 2\Delta)\}^2 w(X_i, Z_i). \tag{11}$$

The resulting test statistics T_1^* and T_1 are discrepancy measures between $\hat{p}$ and $\hat{r}$ in the L_2-distance. Discrepancy-measure based test statistics receive attention and achieve success in the literature. Other discrepancy norms such as the L_∞-distance can also be investigated in the current setting. See the seminal work by [4, 5] and [21]. They are not qualitatively different as shown in the classical goodness of fit tests.

Since the testing problem (8) is equivalent to the following testing problem:

$$
(12)\qquad \begin{cases} H_0: P(z|x, 2\Delta) - R(z|x, 2\Delta) = 0 & \text{for all } (x, z) \in S^2, \\ H_1: P(z|x, 2\Delta) - R(z|x, 2\Delta) \neq 0 & \text{for some } (x, z) \in S^2, \end{cases}
$$

with, in light of (9),

$$
R(z|x, 2\Delta) = \int_{-\infty}^{z} r(t|x, 2\Delta)\, dt = E\{P(z|Y, \Delta)|X = x\},
$$

then transition distribution-based tests can be formulated too. Let $\hat{P}(z|x, 2\Delta)$ be the direct estimator for the 2Δ-transition distribution

$$
(13)\qquad \hat{P}(z|x, 2\Delta) = \frac{1}{n}\sum_{i=1}^{n} W_n(X_i - x, x; h_1) I(Z_i < z).
$$

Regressing the transition distribution $P(z|X_j, \Delta)$ on X_{j-1} yields $\hat{R}(z|x, 2\Delta)$:

$$
(14)\qquad \hat{R}(z|x, 2\Delta) = n^{-1}\sum_{i=1}^{n} W_n(X_i - x, x; h_3)\hat{P}(z|Y_i, \Delta),
$$

where $\hat{P}(z|y, \Delta) = n^{-1}\sum_{i=1}^{n} W_n(Y_i - y, y; b_1) I(Z_i < z)$. Similarly to (11), for the testing problem (12), the transition distribution-based test will be

$$
(15)\qquad T_2 = \sum_{i=1}^{n}\{\hat{P}(Z_i|X_i, 2\Delta) - \hat{R}(Z_i|X_i, 2\Delta)\}^2 \omega(X_i),
$$

where the weight function $\omega(\cdot)$ is chosen to depend on only x-variable, because $\hat{P}(z|x, 2\Delta)$ is a nonparametric estimator of the conditional distribution function, and we need only to weight down the contribution from the sparse regions in the x-coordinate.

Note that the test statistic T_2 involves only one-dimensional smoothing. Hence, it is expected to be more stable than T_1, and the null distribution of T_2 can be better approximated by the asymptotic null distribution. This will be justified by the theorems in the next section.

The choice between the transition density and distribution-based tests reflects different degrees of smoothness of alternatives that we wish to test. In a simpler problem of the traditional goodness-of-fit tests, this has been thoroughly studied in [10]. Essentially, the transition density-based tests are more powerful in detecting local deviations whereas the transition distribution-based tests are more powerful for detecting global deviations.

4. Asymptotic properties.

4.1. *Assumptions*. We assume the following conditions. These conditions are frequently imposed for nonparametric studies for dependent data.

ASSUMPTION (A1). The observed time series $\{X_i\}_{i=1}^{n+2}$ is strictly stationary with time-homogenous $j\Delta$-transition density $p(X_{i+j}|X_i, j\Delta)$.

ASSUMPTION (A2). The kernel functions W and K are symmetric and bounded densities with bounded supports, and satisfy the Lipschitz condition.

ASSUMPTION (A3). The weight function $w(x,z)$ has a continuous second-order derivative with a compact support Ω^*.

ASSUMPTION (A4). The stationary process $\{X_i\}$ is β-mixing with the exponential decay rate $\beta(n) = O(e^{-\lambda n})$ for some $\lambda > 0$.

ASSUMPTION (A5). The functions $p(y|x;\Delta)$ and $p(z|x;2\Delta)$ have continuous second-order partial derivatives with respect to (x,y) and (x,z) on the set Ω^*. The invariant density $\pi(x)$ of $\{X_i\}$ has a continuous second-order derivative for $x \in \Omega_x^*$, a project of the set Ω^* onto the x-axis. Moreover, $\pi(x) > 0$, $p(y|x,\Delta) > 0$ and $p(z|x,2\Delta) > 0$ for all $(x,y) \in \Omega^*$ and $(x,z) \in \Omega^*$.

ASSUMPTION (A6). The joint density $p_{1\ell}(x_1, x_\ell)$ of (X_1, X_ℓ) for $\ell > 1$ is bounded by a constant independent of ℓ. Put $g_{1\ell}(x_1,x_\ell) = p_{1\ell}(x_1,x_\ell) - \pi(x_1)\pi(x_\ell)$. The function $g_{1\ell}$ satisfies the Lipschitz condition: for all (x',y') and (x,y) in Ω^*,

$$|g_{1\ell}(x,y) - g_{1\ell}(x',y')| \le C\sqrt{(x-x')^2 + (y-y')^2}.$$

ASSUMPTION (A7). The bandwidths h_is and b_i are of the same order and satisfy $nh_1^3/\log n \to \infty$ and $nh_1^5 \to 0$.

ASSUMPTION (A8). The bandwidth h_1 converges to zero in such a way that $nh_1^{9/2} \to 0$ and $nh_1^{3/2} \to \infty$.

4.2. *Asymptotic null distributions*. To introduce our asymptotic results, we need the following notation. For any integrable function $f(x)$, let $\|f\|^2 = \int f^2(x)\,dx$ and

$$s(z|x,2\Delta) = \int p^2(z|y,\Delta)p(y|x,\Delta)\,dy = E[p^2(z|Y_1,\Delta)|X_1 = x].$$

Note that the sampled observations $\{X_{n+2-i}\}_{i=0}^{n+1}$ are a reverse Markov process under the null model. We also use $p^*(x|z, 2\Delta)$ to denote the 2Δ-transition density of the reverse process, and let

$$s^*(x|z, 2\Delta) = \int p^{*2}(y|z, \Delta) p^*(x|y, \Delta)\, dy.$$

Denote by

$$\Omega_{11} = \int w(x, z) p^2(z|x, 2\Delta)\, dx\, dz,$$

$$\Omega_{12} = \int w(x, z) p^3(z|x, 2\Delta)\, dx\, dz,$$

$$\Omega_{13} = \int w(x, z) s(z|x, 2\Delta) p(z|x, 2\Delta)\, dx\, dz,$$

$$\Omega_{14} = \int w(x, z) r^2(z|x, 2\Delta) p(z|x, 2\Delta)\, dx\, dz,$$

$$\Omega_{15} = \int w(x, z) s^*(x|z, 2\Delta) p^*(x|z, 2\Delta)[\pi(z)/\pi(x)]^2\, dx\, dz,$$

$$\Omega_2 = \int w^2(x, z) p^4(z|x, 2\Delta)\, dx\, dz.$$

For a kernel function $K(\cdot)$, let $K^*(\cdot) = K * K(\cdot)$ and $K_h(\cdot) = h^{-1}K(\cdot/h)$. Denote by $V(x, z)$ the conditional variance function of $P(z|Y, \Delta)$, given $X = x$. Then it is easy to see that

$$\begin{aligned}\Omega_{13} - \Omega_{14} &= \int w(x, z) V(x, z) p(z|x, 2\Delta)\, dx\, dz \\ &= E\{V(X, Z) w(X, Z) | X = x\}.\end{aligned}$$

Throughout the paper, we use the notation $T_n \overset{a}{\sim} \chi^2_{a_n}$ for a diverging sequence of constants a_n to represent that

$$(T_n - a_n)/\sqrt{2a_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

THEOREM 1. *Assume Conditions (A1)–(A7) hold. If $\{X_i\}$ is Markovian,*

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

where

$$\begin{aligned}\mu_1 = {} & \Omega_{11}\|W\|^2\|K\|^2/(h_1h_2) - \Omega_{12}\|W\|^2 h_1^{-1} \\ & + (\Omega_{13} - \Omega_{14})\|W\|^2/h_3 + \Omega_{15}\|K\|^2/b_2,\end{aligned}$$

and $\sigma_1^2 = 2\Omega_2\|W * W\|^2\|K * K\|^2/(h_1h_2)$. Furthermore, $r_1T_1 \overset{a}{\sim} \chi^2_{a_n}$, where $a_n = r_1\mu_1$ and $r_1 = 2\mu_1/\sigma_1^2$.

The test statistic T_1^*, as far as its null distribution is concerned, can be regarded as a special case of T_1, with the weight function $w(x,z) = p^{-2}(z|x, 2\Delta)w^*(x,z)$. Correspondingly, let Ω_{1j}^* denote Ω_{1j} with $w(x,z)$ replaced by $p^{-2}(z|x, 2\Delta) \times w^*(x,z)$ and Ω_2^* defined similarly. Then, we have

COROLLARY 1. *Under the conditions in Theorem 1 with w replaced by w^*, $r_1^* T_1^* \overset{a}{\sim} \chi^2_{a_n^*}$, where*

$$r_1^* = \frac{\Omega_{11}^* \|W\|^2 \|K\|^2}{\Omega_2^* \|W * W\|^2 \|K * K\|^2} \big(1 + o(1)\big),$$

$$a_n^* = \frac{{\Omega_{11}^*}^2 \|W\|^4 \|K\|^4}{\Omega_2^* \|W * W\|^2 \|K * K\|^2} \frac{1}{h_1 h_2} \big(1 + o(1)\big).$$

The r_1^* is asymptotically a constant depending on only the kernels and the weight function. The degree of freedom a_n^* is independent of nuisance parameters. This reflects that the Wilks phenomenon continues to hold in the current situation.

THEOREM 2. *Under Conditions (A1)–(A6) and (A8), if $\{X_i\}$ is Markovian,*

$$(T_2 - \mu_2)/\sigma_2 \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where

$$\mu_2 = \frac{1}{6h_1} \|W\|^2 \int \omega(x) \{1 + 6 h_1 h_3^{-1} E[V(X_\Delta, Z_\Delta) | X_\Delta = x]\} \, dx,$$

and $\sigma_2^2 = \|W * W\|^2 \|\omega\|^2 / (45 h_1)$. Furthermore, $r_2 T_2 \overset{a}{\sim} \chi^2_{b_n}$, where $b_n = r_2 \mu_2$ and $r_2 = 2\mu_2/\sigma_2^2$.

Comparing Theorems 1 and 2, it is seen that asymptotic variance of T_1 is an order of magnitude larger than that of T_2. Therefore, the null distribution of T_2 can be more stably approximated than that of T_1. On the other hand, the degrees of freedom in T_1 are larger than in T_2, and the transition density-based tests are more omnibus, capable of testing a wider class of alternative hypothesis.

4.3. *Power under contiguous alternative models.* To assess the power of the tests, we consider the following contiguous alternative sequence for T_1:

$$H_{1n} : p(z|x, 2\Delta) - r(z|x, 2\Delta) = g_n(x, z), \tag{16}$$

where g_n satisfies $E[g_n^2(X, Z)] = O(\delta_n^2)$ and $\mathrm{var}[g_n^2(X, Z)] \le M(E[g_n^2(X, Z)])^2$ for a constant $M > 0$ and a sequence δ_n going to zero as $n \to \infty$. Then the power of the test statistic T_1 can be approximated using the following theorem.

THEOREM 3. *Under Conditions (A1)–(A7), if* $nh_1h_2\delta_n^2 = O(1)$, *then under the alternative hypothesis* H_{1n},

$$(T_1 - \mu_1 - d_{1n})/\sigma_{1n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where $d_{1n} = nE\{g_n^2(X, Z)w(X, Z)\}(1 + o(1))$, *and* $\sigma_{1n} = \sqrt{\sigma_1^2 + 4\sigma_{1A}^2}$ *with*

$$\sigma_{1A}^2 = nE[g_n^2(X, Z)w^2(X, Z)\{p(Z|X, 2\Delta) - p^2(Z|X, 2\Delta)\}^2].$$

Using Theorem 1, one can construct an approximate level-α test based on T_1. Let c_α be the critical value such that

$$P\{(T_1 - \mu_1)/\sigma_1 \geq c_\alpha | H_0\} \leq \alpha.$$

Then we have the following result, which demonstrates that the test statistic T_1 can detect alternatives at rate $\delta_n = O(n^{-2/5})$.

THEOREM 4. *Under Conditions (A1)–(A6),* T_1 *can detect alternatives with rate* $\delta_n = O(n^{-2/5})$ *when* $h_1 = c_1 n^{-1/5}$ *and* $h_2 = c_2 n^{-1/5}$ *for some constants* c_1 *and* c_2. *Specifically, if* $\delta_n = dn^{-2/5}$ *for a constant* d, *then*:

(i) $\limsup_{d\to 0} \limsup_{n\to\infty} P\{(T_1 - \mu_1)/\sigma_1 \geq c_\alpha | H_{1n}\} \leq \alpha$;

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

Similarly to (16), we consider the following alternative sequence to study of the power of the test statistic T_2:

$$H_{2n} : P(z|x, 2\Delta) - R(z|x, 2\Delta) = G_n(x, z),$$

where $G_n(x, z)$ satisfies $E[G_n^2(X, Z)] = O(\rho_n^2)$ and $\mathrm{var}(G_n^2(X, Z)) \leq M \times (E[G_n^2(X, Z)])^2$ for a constant $M > 0$ and a sequence ρ_n tending to zero. Then using the following theorem one can calculate the power of the test statistic T_2.

THEOREM 5. *Under Conditions (A1)–(A6) and (A8), if* $nh_1h_3\rho_n^2 = O(1)$, *then under the alternative hypothesis* H_{2n},

$$(T_2 - \mu_2 - d_{2n})/\sigma_{2n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where $d_{2n} = nE[G_n^2(X, Z)\omega(X)] + O(nh_1^2\rho_n + \rho_n h_1^{-1})$, $\sigma_{2n}^2 = \sigma_2^2 + 4\sigma_{2A}^2$ *and*

$$\sigma_{2A}^2 = nE\left[\int G_n(X, Z)\omega(X)I(Z < z)P(dz|X, 2\Delta)\right]^2 - nE\left[\int G_n(X, Z)\omega(X)P(z|X, 2\Delta)P(dz|X, 2\Delta)\right]^2.$$

In a manner parallel to Theorem 4, the following theorem demonstrates the optimality of the test.

THEOREM 6. *Under Conditions (A1)–(A6), T_2 can detect alternatives with rate $\rho_n = O(n^{-4/9})$ when $h_1 = c_* n^{-2/9}$ for some constant c_*.*

From Theorem 6, T_2 can detect alternatives at rate $O(n^{-4/9})$. Using an argument similar to [11], we can also establish the minimax rate, $O(n^{-4/9})$, of the test. Note that the rate is optimal according to [26, 27] and [29]. Compared with Theorem 4, it is seen that T_2 is more powerful than T_1 for testing the Markov hypothesis. This is due to the fact that the alternative under consideration for T_2 is global, namely, the density under the alternative is basically globally shifted away from the null hypothesis. On the other hand, T_1 and T_1^* are more powerful than T_2 for detecting local features of the alternative hypothesis. We will now explore these features in simulations.

5. Simulations. An important application of our test methods is to verify the Markov property in the context where the null model is a diffusion process, since it is often assumed in modern financial theory and practice that the observation process comes from an underlying diffusion. Hence, we consider simulations for the diffusion models.

To use the test statistics, one needs to find their null distributions. Theoretically the asymptotic null distributions may be used to determine the p-values of the test statistics. However, in practical applications the asymptotic distributions do not necessarily give accurate approximations, since the local sample size nh_1h_2 may not be large enough. This phenomenon is shared by virtually all nonparametric kinds of tests where some form of functional estimation is used.

We will mainly focus on the finite sample performance of the test statistic T_1^*, since it possesses the Wilks property which facilitates bandwidth selection and determination of the null distribution using a bootstrap method. Since the asymptotic null distribution of T_1^* is independent of nuisance parameters/functions under the null hypothesis, for a finite sample it does not sensitively depend on the nuisance parameters/functions. Therefore, the null distribution can be approximated by bootstraps, by fixing nuisance parameters/functions at their reasonable estimates, as in [12] in a different context.

In general, different bootstrap approximations to the null distributions are needed for different null models, partially due to the large family of null models with the Markov property. We will illustrate this method for the Ornstein–Uhlenbeck model, which in financial mathematics is used for instance as the [30] model for interest rates. For other parametric models, our approach can similarly be applied.

The Ornstein–Uhlenbeck model employed as the null hypothesis is

$$dX_t = \kappa(\alpha - X_t)\,dt + \sigma\,dW_t, \tag{17}$$

where W_t is a Brownian motion, and the parameters are set as $\kappa = 0.2$, $\alpha = 0.085$, $\sigma = 0.08$, which are realistic for interest rates over long periods. We simulated the model 1000 times. In each simulation, we draw a sample with sample size $n = 2400$ and weekly sampling interval $\Delta = 1/52$ using for this purpose a higher frequency Euler approximation, or an exact discretization. The bandwidth selection for the test statistic T_1^* is performed using the simple empirical rule proposed by [25]. Alternative methods include the cross-validation approaches of [15] and [20], but their computation is intensive especially when repeated many times in Monte Carlo.

Given a sample from the model, we fit the model using the least squares method and obtain the residuals of the fit, and then generate bootstrap samples using the residual-based bootstrap method. For each simulation, we obtained three bootstrap samples (this is merely for the reduction of computation cost; using more samples will not fundamentally alter the results) and computed the test statistic T_1^* using the same bandwidths as the original sample in the simulation. Pooling together the bootstrap samples from each simulation, we obtained 3000 bootstrap statistics. Their sampling distributions, computed via the kernel density estimate, is taken as the distribution of the bootstrap method. By using the kernel density estimation method, the distribution of the realized values of the test statistic T_1^* in simulations is obtained as the true distribution (except for the Monte Carlo errors).

Figure 1 displays the estimated densities for T_1^*. Not surprisingly, the bootstrapped distributions get much closer to the true ones as the sample sizes increase. In our experience, the bootstrap approximations start to become adequate for sample sizes starting at about 2400.

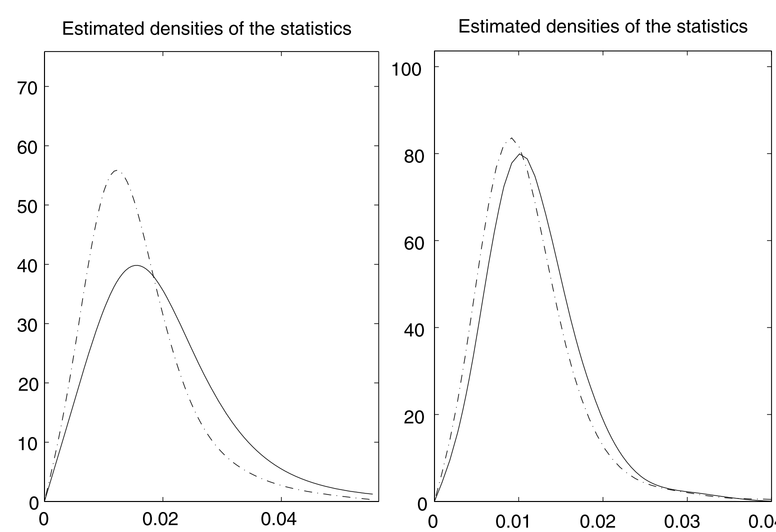

FIG. 1. *Estimated densities. Left panel*: $n = 1200$; *right panel*: $n = 2400$. *Solid—true*, *dotted—the bootstrap approximation*.

To investigate the power of the test statistics, we employ various sequences of alternatives indexed by a parameter $\theta = 0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0$. One of the main ways for an otherwise Markovian model to become non-Markovian is to restrict too much its state space. For instance, consider a bivariate diffusion model. Taken jointly, the two components are Markovian, but taken in isolation a single component may not be:

1. Alternative model with missing state variable in the drift: we first consider the situation where the null model (17) is missing a state variable, in this case X mean-revers to the stochastic level $\theta\alpha_t + (1-\theta)\alpha$ under the alternative

$$H_{1\theta} : dX_t = \kappa\big(\theta\alpha_t + (1-\theta)\alpha - X_t\big)\,dt + \sigma\,dW_t,$$

where α_t is the random process

$$d\alpha_t = \kappa_1(a - \alpha_t)\,dt + \sigma_1\,dB_t,$$

with B_t a the Brownian motion independent of W_t, $\kappa_1 = \kappa/s$, $a = s\alpha$, and $\sigma_1 = \sigma/2$, with $s = 100$ and 10. When $\theta \neq 0$, the alternatives are non-Markovian. The results in the first part of Table 1 show that the test statistic rejects the null hypothesis when the observations are drawn under $H_{1\theta}$.
2. Alternative model with missing state variable in volatility: next, we consider alternative models where volatility is stochastic,

$$H_{2\theta} : dX_t = \kappa(\alpha - X_t)\,dt + \big((1-\theta)\sigma + \theta\sigma_t\big)\,dW_t,$$

where $\sigma_t = \sqrt{Y_t}$ is a random process following the [9] model

$$dY_t = \kappa_2(b - Y_t)\,dt + \sigma_2 Y_t^{1/2}\,dB_{2t},$$

where B_{2t} is a standard Brownian motion independent of W_t, $\kappa_2 = \kappa/s$, $b = s\alpha$ and $\sigma_2 = \sigma/2$, with $s = 1000$, 100 and 10. When $\theta \neq 0$, the alternatives are also non-Markovian.
3. Alternative model with missing state variable in jumps: finally, we consider a model with compound Poisson jumps

$$H_{3\theta} : dX_t = \kappa(\alpha - X_t)\,dt + \sigma\,dW_t + J_t\,dN_t(\theta),$$

TABLE 1
Power of the test against $H_{1\theta}$

		Parameter θ					
s	**Level α**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
100	0.01	0.011	1	1	1	1	1
	0.05	0.055	1	1	1	1	1
10	0.01	0.011	0.010	0.070	0.228	0.580	0.846
	0.05	0.055	0.019	0.123	0.549	0.901	0.989

where $N_t(\theta)$ is a Poisson process with stochastic intensity θ and jump size 1, while J_t is a the jump size. We will consider two types of jump sizes:

(i) J_t is independent of $\mathcal{F}_t$ and follows $N(0, \sigma_1^2)$ with $\sigma_1 = \sigma/2$, which makes $H_{3\theta}$ Markovian;

(ii) J_t follows the CIR model

$$dJ_t = \kappa(a - J_t)\,dt + \sigma_1 J_t^{1/2}\,dB_{3t},$$

where B_{3t} is a standard Brownian motion independent of W_t, $K = 0.2$, $a = 0.085$ and $\sigma_1 = 0.08/2$. Then J_t is not independent of $\mathcal{F}_t$. This leads to alternatives $H_{3\theta}$ which are not Markovian for $\theta \neq 0$.

The alternative models considered here are β-mixing. For example, in the first alternative $H_{1\theta}$, the joint process (X_t, α_t) is an affine process and it is β-mixing. Hence, X_t is β-mixing. A similar argument applies to two other alternatives. In fact, for the first alternative $H_{1\theta}$, the time series $(X_{i\Delta}, \alpha_{i\Delta})$ can be written as a bivariate autoregressive model. Hence, it is β-mixing with the choice of parameters. Note that for all of the above alternatives, when θ is small, the null and alternative models are nearly impossible to differentiate. In the limit where $\theta = 0$, the null and the alternative are identical. Therefore, it can be expected that, when $\theta = 0$, the power of test should be close to the significance level; and as θ deviates more from 0, the power should increase. Also we can expect that our tests will be able to detect only the type (ii) jumps but not the type (i) jump, since for the type (i) jump the alternatives are Markovian.

The simulated powers are reported in Tables 1–3. The null distribution of the normalized test statistics does not depend sensitively on choice of bandwidth, whereas the power depends on the choice of bandwidth and the alternative under consideration. As expected, our test is fairly powerful for detecting non-Markovian alternatives $H_{k\theta}$ $(k = 1, 2, 3)$, at least in situations where the alternative is sufficiently far from the null. For $H_{3\theta}$, the test has, as it should, no power to identify the type (i) alternatives but is powerful for discriminating against the type (ii) alternatives. This illustrates well the sensitivity and specificity of our tests.

TABLE 2
Power of the test against $H_{2\theta}$

		Parameter θ					
s	**Level α**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
1000	0.01	0.013	0.402	0.660	0.762	0.813	0.817
	0.05	0.067	0.557	0.768	0.845	0.878	0.905
100	0.01	0.013	0.028	0.183	0.372	0.492	0.573
	0.05	0.067	0.098	0.340	0.527	0.627	0.697
10	0.01	0.013	0.007	0.020	0.017	0.032	0.088
	0.05	0.067	0.037	0.052	0.070	0.122	0.218

TABLE 3
Power of the test against $H_{3\theta}$

Jump type	Level α	Parameter θ 0.0	0.2	0.4	0.6	0.8	1.0
(i)	0.01	0.010	0.009	0.023	0.003	0.016	0.009
	0.05	0.059	0.048	0.054	0.054	0.058	0.056
(ii)	0.01	0.010	0.514	0.774	0.888	0.940	0.951
	0.05	0.059	0.533	0.796	0.894	0.946	0.961

6. Technical proofs.

6.1. *Technical lemmas.* We now introduce some technical lemmas, the proofs of which can be found in the supplemental material of this paper. To save space, some notation in the lemmas will appear later in the course of proofs of the main theorems.

LEMMA 1. *Suppose that W is symmetric and continuous with a bounded support. If $h \to 0$ and $nh \to \infty$, then*

$$W_n(z, x; h) = \left\{\frac{1}{\mu_0(W)\pi(x)} - \frac{z}{h}\frac{h\pi'(x)}{\pi^2(x)\mu_0(W)} + O_p(\rho_n(h))\right\} W_h(z) + O_p(\rho_n(h))\frac{z}{h} W_h(z),$$

uniformly for $x \in \Omega^*$, where $O_p(\rho_n(h))$ does not depend on z, where $\mu_0(W) = \int W(u)\,du$.

LEMMA 2. *Under Conditions (A1)–(A6):*

(i) *for $k = 0, 1$,*

$$\sup_{(y,z)\in\Omega^*}\left|n^{-1}\sum_{i=1}^{n} b_1^{-k}(Y_i - y)^k W_{b_1}(Y_i - y)\varepsilon_i(z)\right| = O\{\sqrt{(\log n)/(nb_1b_2)}\};$$

(ii) *for $k = 0, 1$,*

$$\sup_{(x,z)\in\Omega^*}\left|\frac{1}{n}\sum_{j=1}^{n} h_3^{-k}(X_j - x)^k W_{h_3}(X_j - x)e_j(z)\right| = O_p\{\sqrt{\log(n)/(nh_3)}\};$$

(iii) $\sup_{(x,z)\in\Omega^*}|\frac{1}{n}\sum_{j=1}^{n} q^*(x, Z_j)\varepsilon_{j+1}(z)| = O_p\{\sqrt{\log(n)/(nb_2)}\}$;

(iv) $\sup_{(x,z)\in\Omega^*}|\frac{1}{n}\sum_{j=1}^{n} W_{h_1}(X_j - x)\varepsilon_j^*(z)| = O_p\{\sqrt{(\log n)/(nh_1h_2)}\}$.

LEMMA 3. *Under Conditions (A1)–(A6), we have*

$$\xi_n(x, y) \equiv \frac{1}{n}\sum_{i=1}^{n} r_{n1}(x, Y_i)\varepsilon_i(z) = O_p\bigl(\sqrt{n^{-1}b_1 \log n}\bigr),$$

uniformly for $(x, z) \in \Omega^*$, *where* r_{n1} *is defined right after* (32).

LEMMA 4. *Suppose Conditions (A1)–(A5) hold. Then*

$$\eta_n(x, z) \equiv n^{-1}\sum_{i=1}^{n} r_n^*(x, Y_i)\varepsilon_i(z) = O\bigl\{\sqrt{[(b_1^4 + h_3^4)\log n]/(nb_2)}\bigr\},$$

uniformly for $(x, y) \in \Omega^*$, *where* $r_n^*(\cdot, \cdot)$ *is defined in* (34).

LEMMA 5. *Under Conditions (A1)–(A6):*

(i) $\sum_{1\le i<j\le n}[\tilde{\psi}(i, j) - \tilde{\psi}(i) - \tilde{\psi}(j) + \tilde{\psi}(0)] = o_p(h_1^{-1})$;
(ii) $(n-1)\sum_{i=1}^{n}[\tilde{\psi}(i) - \tilde{\psi}(0)] = o_p(h_1^{-1})$.

LEMMA 6. *Assume Conditions (A1)–(A5) hold. Then we have:*

(i) *under Condition (A6),*

$$\begin{aligned}\tfrac{1}{2}n(n-1)\tilde{\psi}(0) = {} & \Omega_{11}\|W\|^2\|K\|^2/(h_1h_2) - \Omega_{12}\|W\|^2/h_1 \\ & + \Omega_{13}\|W\|^2/h_3 - \Omega_{14}\|W\|^2/h_3 \\ & + \Omega_{15}\|K\|^2/b_2 + O(n^{-2});\end{aligned}$$

(ii) *under Condition (A7),*

$$\begin{aligned}&\frac{1}{2}n(n-1)\tilde{\phi}(0) \\ &\quad = \frac{1}{6h_1}\|W\|^2\int \omega(x)\{1 + 6h_1h_3^{-1}E[V(X_\Delta, Z_\Delta)|X_\Delta = x]\}\,dx + O(1).\end{aligned}$$

LEMMA 7. *Assume that Conditions (A1)–(A5) hold. Then we have:*

(i) *under Condition (A6),*

$$\sigma_{1n}^{-1}\sum_{1\le i<j\le n}\psi^*(i, j) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where $\sigma_{1n}^2 = 2\Omega_2\|W * W\|^2\|K * K\|^2/(n^2h_1h_2)$;
(ii) *under Condition (A7),*

$$\sigma_{2n}^{-1}\sum_{1\le i<j\le n}\phi^*(i, j) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where $\sigma_{2n}^2 = \|W * W\|^2\|w\|^2/(45n^2h_1)$.

6.2. *Preliminaries*. Since the test statistics T_1 and T_1^* compare the difference between $\hat{p}(z|x, 2\Delta)$ and $\hat{r}(z|x, 2\Delta)$, we derive an asymptotic expression for this difference under H_0 before giving the proofs of theorems. In addition, in order to streamline our arguments, we will introduce some technical lemmas and put them behind the proofs of theorems. The arguments employed here use techniques from the U-statistic and nonparametric smoothing.

First let us introduce some notation. Let $\rho_n(h) = h^2 + \sqrt{\log n/(nh)}$, $\mu_0(W) = \int W(x)\,dx$ and $\mu_2(W) = \int x^2 W(x)\,dx$. Denote by $m(y, z) = E\{K_{b_2}(Z_j - z)|Y_j = y\}$, $m^*(x, z) = E\{K_{h_2}(Z_j - z)|X_j = x\}$, $m_1(y, z) = \partial m(y, z)/\partial y$ and $m_1^*(x, z) = \partial m^*(x, z)/\partial x$.

Using an elementary property of the local linear smoother (see, e.g., [13]), we obtain that

$$\hat{p}(z|x, 2\Delta) - p(z|x, 2\Delta) = A_n^*(x, z) + B_n^*(x, z) + C_n^*(x, z), \tag{18}$$

where $\varepsilon_j^*(z) = K_{h_2}(Z_j - z) - m^*(X_j, z)$,

$$\begin{aligned} A_n^*(x, z) &= \frac{1}{n}\sum_{j=1}^n W_n(X_j - x, x; h_1)\varepsilon_j^*(z), \\ B_n^*(x, z) &= \frac{1}{n}\sum_{j=1}^n W_n(X_j - x, x; h_1) \\ &\qquad \times \{m^*(X_j, z) - m^*(x, z) - m_1^*(x, z)(X_j - x)\}, \\ C_n^*(x, z) &= m^*(x, z) - p(z|x, 2\Delta). \end{aligned} \tag{19}$$

By a second-order Taylor expansion,

$$B_n^*(x, z) = \frac{1}{n}\sum_{j=1}^n W_n(X_j - x, x; h_1)\frac{h_1^2}{2} m_2^*(\tilde{x}_j, z)\left(\frac{X_j - x}{h_1}\right)^2,$$

where $m_2^*(\tilde{x}, z) = \frac{\partial^2 m^*(x,z)}{\partial^2 x}|_{x=\tilde{x}_j}$, and $\tilde{x}_j$ lies between X_j and x. By [14], it is easy to show that

$$B_n^*(x, z) = O_p(h_1^2) \quad \text{and} \quad C_n^*(x, z) = O_p(h_2^2), \tag{20}$$

uniformly for $(x, z) \in \Omega^*$. By the definition of $\hat{r}$, we have

$$\hat{r}(z|x, 2\Delta) - r(z|x, 2\Delta) = L_{n1}(x, z) + L_{n1}^*(x, z), \tag{21}$$

where

$$L_{n1}(x, z) = \frac{1}{n}\sum_{j=1}^n W_n(X_j - x, x; h_3)\{\hat{p}(z|Y_j, \Delta) - p(z|Y_j, \Delta)\},$$

$$L_{n1}^*(x, z) = \frac{1}{n}\sum_{j=1}^n W_n(X_j - x, x; h_3)\{p(z|Y_j, \Delta) - r(z|x, 2\Delta)\}.$$

Subtracting (21) from (18), we obtain that, under $H_0: p(z|x, 2\Delta) = r(z|x, 2\Delta)$,

$$\hat{p}(z|x, 2\Delta) - \hat{r}(z|x, 2\Delta) = A_n^*(x, z) + B_n^*(x, z) + C_n^*(x, z) \\ - L_{n1}(x, z) - L_{n2}(x, z) - L_{n3}(x, z), \tag{22}$$

where

$$L_{n2}(x, z) = n^{-1} \sum_{j=1}^{n} W_n(X_j - x, x; h_3)\{p(z|Y_j, \Delta) - r(z|X_j, 2\Delta)\},$$

$$L_{n3}(x, z) = n^{-1} \sum_{j=1}^{n} W_n(X_j - x, x; h_3)\{r(z|X_j, 2\Delta) - r(z|x, 2\Delta)\}.$$

By the continuity of $\partial^2 r(z|x, 2\Delta)/\partial x^2$, it is easy to show that

$$L_{n3}(x, z) = O_p(h_3^2) \qquad \text{uniformly for } (x, z) \in \Omega^*. \tag{23}$$

Therefore, by (20), (22) and (23),

$$\hat{p}(z|x, 2\Delta) - \hat{r}(z|x, 2\Delta) \\ = [A_n^*(x, z) - L_{n2}(x, z)] - L_{n1}(x, z) + O_p\left(\sum_{i=1}^{3} h_i^2\right). \tag{24}$$

Let $e_j(z) = p(z|Y_j, \Delta) - r(z|X_j, 2\Delta)$. Then it can be rewritten that

$$L_{n2}(x, z) = \frac{1}{n} \sum_{j=1}^{n} W_n(X_j - x, x; h_3) e_j(z). \tag{25}$$

Note that $r(z|X_j, 2\Delta) = E\{p(z|Y_j, \Delta)|X_j\}$. It follows that $E[e_j(z)|X_j] = 0$ and $\mathrm{Var}[e_j(z)] = O(1)$ uniformly for z and $j = 1, \dots, n$. Applying Lemma 1 with $z = X_j - x$ and $h = h_3$, we obtain that

$$W_n(X_j - x, x; h_3) \\ = \left\{\frac{1}{\mu_0 \pi(x)} - \frac{X_j - x}{h_3} \frac{h_3 \pi'(x)}{\pi^2(x)\mu_0} + O_p(\rho_n(h_3))\right\} W_{h_3}(X_j - x) \\ + O_p(\rho_n(h_3)) \frac{X_j - x}{h_3} W_{h_3}(X_j - x), \tag{26}$$

uniformly for $x \in \Omega^*$, where $O_p(\rho_n(h_3))$ does not depend on j. Therefore,

$$L_{n2}(x, z) = L_{n21}(x, z) - L_{n22}(x, z) + L_{n23}(x, z) + L_{n24}(x, z),$$

where

$$L_{n21}(x,z) = \frac{1}{\mu_0(W)\pi(x)} n^{-1} \sum_{j=1}^{n} W_{h_3}(X_j - x)e_j(z),$$

$$L_{n22}(x,z) = \frac{h_3\pi'(x)}{\mu_0(W)\pi^2(x)} n^{-1} \sum_{j=1}^{n} \frac{X_j - x}{h_3} W_{h_3}(X_j - x)e_j(z),$$

$$L_{n23}(x,z) = O_p(\rho_n(h_3)) n^{-1} \sum_{j=1}^{n} W_{h_3}(X_j - x)e_j(z),$$

$$L_{n24}(x,z) = O_p(\rho_n(h_3)) n^{-1} \sum_{j=1}^{n} \frac{X_j - x}{h_3} W_{h_3}(X_j - x)e_j(z).$$

By Lemma 2(ii), we have $L_{n21}(x,z) = O_p\{\sqrt{(\log n)/(nh_3)}\}$ and

$$L_{n22}(x,z) = O_p\{h_3\sqrt{(\log n)/(nh_3)}\} = O_p\{\sqrt{(h_3 \log n)/n}\},$$

uniformly for $(x,z) \in \Omega^*$. Then

$$L_{n23}(x,z) = O_p\{\rho_n(h_3)\sqrt{(\log n)/(nh_3)}\} = o_p\{\sqrt{(h_3 \log n)/n}\}$$

and $L_{n24}(x,z) = o_p\{\sqrt{(h_3 \log n)/n}\}$, uniformly for $(x,z) \in \Omega^*$. Then

$$L_{n2}(x,z) = \frac{1}{\mu_0(W)\pi(x)} \frac{1}{n} \sum_{j=1}^{n} W_{h_3}(X_j - x)e_j(z) + O_p\{\sqrt{(h_3 \log n)/n}\},$$

uniformly for $(x,z) \in \Omega^*$. Note that from (19) and (25)

$$A_n^*(x,z) - L_{n2}(x,z) = \frac{1}{n} \sum_{j=1}^{n} [W_n(X_j - x, x; h_1)\varepsilon_j^*(z) \tag{27}$$

$$- W_n(Y_j - x, x; h_3)e_{j+1}(z)] + r_n(x,z), \tag{28}$$

where

$$r_n(x,z) = -\frac{1}{n} W_n(X_1 - x, x; h_3)e_1(z) + \frac{1}{n} W_n(Y_n - x, x; h_3)e_{n+1}(z),$$

which is of order $O_p(1/(nh_3)) = o_p\{\sqrt{(h_3 \log n)/n}\}$, uniformly for $(x,z) \in \Omega^*$. Let $\varepsilon_i(z) = K_{b_2}(Z_i - z) - m(Y_i, z)$. Then, similarly to (18), we have

$$\hat{p}(z|y,\Delta) - p(z|y,\Delta) = A_n(y,z) + B_n(y,z) + C_n(y,z), \tag{29}$$

where $A_n(y,z) = n^{-1}\sum_{i=1}^{n} W_n(Y_i - y, y; b_1)\varepsilon_i(z)$, $B_n(y,z) = O_p(b_1^2)$ and $C_n(y,z) = O_p(b_1^2)$, uniformly for $(y,z) \in \Omega^*$. It follows from the definition of

L_{n1} that

$$L_{n1}(x,z) = n^{-1}\sum_{i=1}^{n} W_n(X_j - x, x; h_3)A_n(Y_j, z) \tag{30}$$
$$+ n^{-1}\sum_{i=1}^{n} W_n(X_j - x, x; h_3)[B_n(Y_j, z) + C_n(Y_j, z)].$$

Using Lemma 1, we get

$$A_n(y,z) = A_{n1}(y,z) - A_{n2}(y,z) + A_{n3}(y,z) + A_{n4}(y,z),$$

where

$$A_{n1}(y,z) = \frac{1}{\mu_0 \pi(y)} n^{-1}\sum_{i=1}^{n} W_{b_1}(Y_i - y)\varepsilon_i(z),$$
$$A_{n2}(y,z) = \frac{b_1\pi'(y)}{\mu_0\pi^2(x)} n^{-1}\sum_{i=1}^{n} \frac{Y_i - y}{b_1} W_{b_1}(Y_i - y)\varepsilon_i(z),$$
$$A_{n3}(y,z) = O_p(\rho_n(b_1)) n^{-1}\sum_{i=1}^{n} W_{b_1}(Y_i - y)\varepsilon_i(z),$$
$$A_{n4}(y,z) = O_p(\rho_n(b_1)) n^{-1}\sum_{i=1}^{n} \frac{Y_i - y}{b_1} W_{b_1}(Y_i - y)\varepsilon_i(z).$$

Using Lemma 2(i), we obtain that

$$A_{n3}(y,z) = O_p(\rho_n(b_1)) O_p\left(\sqrt{\frac{\log n}{n b_1 b_2}}\right)$$

and

$$A_{n4}(y,z) = O_p(\rho_n(b_1)) O_p\left(\sqrt{\frac{\log n}{n b_1 b_2}}\right),$$

uniformly for $(y,z) \in \Omega^*$. Then

$$A_n(y,z) = A_{n1}(y,z) - A_{n2}(y,z) + O_p(\rho_n(b_1))\sqrt{(\log n)/(n b_1 b_2)},$$

uniformly for $(y,z) \in \Omega^*$. This, combined with (30) and Condition (A6), yields that

$$L_{n1}(x,z) = L_{n11}(x,z) - L_{n12}(x,z) + L_{n13}(x,z) + O_p\{(\log n)/(n b_1^{3/2})\}, \tag{31}$$

where $L_{n11}(x,z) = n^{-1}\sum_{j=1}^{n} W_n(X_j - x, x; h_3)A_{n1}(Y_j, z)$,

$$L_{n12}(x,z) = n^{-1}\sum_{j=1}^{n} W_n(X_j - x, x; h_3)A_{n2}(Y_j, z),$$
$$L_{n13}(x,z) = n^{-1}\sum_{j=1}^{n} W_n(X_j - x, x; h_3)[B_n(Y_j, z) + C_n(Y_j, z)].$$

Note that, by Lemma 2(i), $A_{n1}(y,z) = O_p\{\sqrt{(\log n)/(nb_1 b_2)}\}$, uniformly for $(y,z) \in \Omega^*$. Using Lemma 1, we obtain that

$$L_{n11}(x,z) = M_{n11}(x,z) + M_{n12}(x,z) + O_p\{(\log n)/(nb_1^{3/2})\}, \tag{32}$$

where

$$M_{n11}(x,z) = \frac{1}{\mu_0^2 \pi(x)} \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n W_{h_3}(X_j - x) W_{b_1}(Y_i - Y_j) \pi^{-1}(Y_j) \varepsilon_i(z),$$

$$\begin{aligned} M_{n12}(x,z) = \frac{h_3 \pi'(x)}{\mu_0^2 \pi^2(x)} \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n \frac{X_j - x}{h_3} W_{h_3}(X_j - x) W_{b_1}(Y_i - Y_j) \\ \times \pi^{-1}(Y_j) \varepsilon_i(z). \end{aligned}$$

Let

$$M_{n11}^*(x,y) = n^{-1} \sum_{j=1}^n W_{h_3}(X_j - x) W_{b_1}(y - Y_j) \pi^{-1}(Y_j),$$

$g_n(x,y) = E[M_{n11}^*(x,y)]$ and $r_{n1}(x,y) = M_{n11}^*(x,y) - g_n(x,y)$. Then

$$M_{n11}(x,z) = \frac{1}{\mu_0^2 \pi(x)} \frac{1}{n} \sum_{i=1}^n g_n(x, Y_i) \varepsilon_i(z) + \frac{1}{\mu_0^2 \pi(x)} \frac{1}{n} \sum_{i=1}^n r_{n1}(x, Y_i) \varepsilon_i(z).$$

By Lemma 3,

$$M_{n11}(x,z) = \frac{1}{\mu_0^2 \pi(x)} \frac{1}{n} \sum_{i=1}^n g_n(x, Y_i) \varepsilon_i(z) + O_p\{\sqrt{(b_1 \log n)/n}\}, \tag{33}$$

uniformly for $(x,z) \in \Omega^*$. Similarly to Lemma 2(iii), the first term on the right-hand side of (33) is $O_p\{\sqrt{(\log n)/(nb_2)}\}$, uniformly for $(x,z) \in \Omega^*$. Hence,

$$\sup_{(x,z)\in\Omega^*} |M_{n11}(x,z)| = O_p\{\sqrt{(\log n)/(nb_2)}\}.$$

Similarly, we have

$$\sup_{(x,z)\in\Omega^*} |M_{n12}(x,z)| = O_p\{h_3 \sqrt{(\log n)/(nb_2)}\} = O_p\{\sqrt{(b_1 \log n)/n}\}.$$

By the symmetry of the kernel function and Taylor's expansion, it can be shown that

$$\begin{aligned} g_n(x,y) &= E[\pi^{-1}(Y_1) W_{b_1}(y - Y_1) W_{h_3}(X_1 - x)] \\ &= \mu_0^2 p(y|x,\Delta) \pi(x)/\pi(y) + O(b_1^2 + h_3^2) \\ &\equiv \mu_0^2 p^*(x|y,\Delta) + O(b_1^2 + h_3^2), \end{aligned}$$

uniformly for $(x, y) \in \Omega^*$, where $p^*(x|y, \Delta)$ is the one-Δ transition density of the reverse series $\{X_{n+2-i}\}_{i=1}^{n+1}$, that is, the conditional density of X_1 given $Y_1 = y$. Note that g_n is a deterministic function. It follows that

$$g_n(x, Y_i) = \mu_0^2 p^*(x|Y_i, \Delta) + r_n^*(x, Y_i), \tag{34}$$

where $r_n^*(x, Y_i)$ is $\sigma(Y_i)$-measurable and is of order $O(b_1^2 + h_3^2)$ for $(x, Y_i) \in \Omega^*$. This combined with (33) leads to

$$\begin{aligned} L_{n11}(x, z) &= \frac{1}{n}\sum_{i=1}^{n} q^*(x, Y_i)\varepsilon_i(z) + \frac{O(1)}{n}\sum_{i=1}^{n} r_n^*(x, Y_i)\varepsilon_i(z) \\ &\quad + O_p\bigl(\{\log n/(nb_1^{3/2})\} + \{b_1(\log n)/n\}^{1/2}\bigr), \end{aligned} \tag{35}$$

where $q^*(x, y) = p(y|x, \Delta)/\pi(y)$. The first term in (35) is obviously

$$\frac{1}{n}\sum_{i=1}^{n} q^*(x, Z_i)\varepsilon_{i+1}(z) + O_p\left(\frac{1}{nb_1}\right).$$

By Lemma 4, the second term in (35) is $O_p(\sqrt{(b_1^4 + h_3^4)\log(n)/(nb_2)})$, uniformly for $(x, z) \in \Omega^*$. Then uniformly for $(x, z) \in \Omega^*$,

$$L_{n11}(x, z) = \frac{1}{n}\sum_{i=1}^{n} q^*(x, Z_i)\varepsilon_{i+1}(z) + O_p\bigl(\{\log n/(nb_1^{3/2})\} + \{b_1(\log n)/n\}^{1/2}\bigr).$$

In the same argument, $L_{n12}(x, z)$ is dominated by $L_{n11}(x, z)$ and is of order

$$b_1 L_{n11}(x, z) = O_p\bigl(\{\log n/(nb_1^{3/2})\} + \{b_1 \log n/n\}^{1/2}\bigr),$$

which combined with (31) leads to

$$\begin{aligned} L_{n1}(x, z) &= \frac{1}{n}\sum_{i=1}^{n} q^*(x, Z_i)\varepsilon_{i+1}(z) + L_{n13}(x, z) \\ &\quad + O_p\bigl(\{\log n/(nb_1^{3/2})\} + \{b_1 \log n/n\}^{1/2}\bigr), \end{aligned} \tag{36}$$

uniformly for $(x, z) \in \Omega^*$. This together with (24) and (27) yields the following asymptotic expression:

$$\hat{p}(z|x, 2\Delta) - \hat{r}(z|x, 2\Delta) = T_{n1}(x, z) + T_{n2}(x, z) + T_{n3}(x, z) + T_{n4}(x, z), \tag{37}$$

where

$$\begin{aligned} T_{n1}(x, z) = \frac{1}{n}\sum_{j=1}^{n} [&W_n(X_j - x, x; h_1)\varepsilon_j^*(z) \\ &- W_n(Y_j - x, x; h_3)e_{j+1}(z) - q^*(x, Z_j)\varepsilon_{j+1}(z)], \end{aligned}$$

$$T_{n2}(x,z) = n^{-1}\sum_{j=1}^{n} W_n(X_j - x, x; h_3)[B_n(Y_j, z) + C_n(Y_j, z)],$$

$$T_{n3}(x,z) = B_n^*(x,z) + C_n^*(x,z) + L_{n3}(x,z),$$

$$T_{n4}(x,z) = O_p\big(\{\log n/(nh_1^{3/2})\} + \{b_1 \log n/n\}^{1/2} + \{\log n/(nb_1^{3/2})\}\big),$$

uniformly for $(x,z) \in \Omega^*$.

6.3. *Proofs of theorems.* We now give the proofs of our main results.

PROOF OF THEOREM 1. (i) *Approximate T_1 by a U-statistic.* Let $w_i = w(X_i, Z_i)$. By (37) and the definition of T_1, we have

$$\begin{aligned}
T_1 &= \sum_{i=1}^{n} w_i[T_{n1}(X_i,Z_i) + T_{n2}(X_i,Z_i) + T_{n3}(X_i,Z_i) + T_{n4}(X_i,Z_i)]^2 \\
&= \sum_{i=1}^{n}\sum_{k=1}^{4} w_i T_{nk}^2(X_i,Z_i) + 2\sum_{i=1}^{n} w_i T_{n1}(X_i,Z_i)T_{n2}(X_i,Z_i) \\
&\quad + 2\sum_{i=1}^{n} w_i T_{n1}(X_i,Z_i)T_{n3}(X_i,Z_i) + 2\sum_{i=1}^{n} w_i T_{n2}(X_i,Y_i)T_{n3}(X_i,Y_i) \\
&\quad + 2\sum_{i=1}^{n} w_i[T_{n1}(X_i,Z_i) + T_{n2}(X_i,Z_i) + T_{n3}(X_i,Z_i)]T_{n4}(X_i,Z_i) \\
&\equiv T_{11} + T_{12} + T_{13} + T_{14} + T_{15}.
\end{aligned}$$

By Lemmas 1 and 2, $T_{n1}(x,z) = O_p\{\sqrt{(\log n)/(nh_1h_2)}\}$. Note that $T_{n2}(x,z) = O_p(b_1^2)$, $T_{n3}(x,z) = O_p(h_1^2)$, uniformly for $(x,z) \in \Omega^*$. It is straightforward to verify that $T_{14} = O_p(nh_1^4) = o(1/h_1)$, $T_{15} = o_p(1/\sqrt{h_1h_2})$. Using the same argument as for (B.2) in [3], we obtain $T_{12} = o_p(1/\sqrt{h_1h_2})$ and $T_{13} = o_p(1/\sqrt{h_1h_2})$. Therefore,

$$T_1 = \sum_{i=1}^{n}\sum_{k=1}^{4} w_i T_{nk}^2(X_i,Z_i) + o_p(h_1^{-1}).$$

Note that

$$\sum_{i=1}^{n} w_i T_{n2}^2(X_i,Z_i) = O_p(nh_1^4) = o_p(1/h_1),$$

$$\sum_{i=1}^{n} w_i T_{n3}^2(X_i,Z_i) = o_p(1/h_1)$$

and

$$\sum_{i=1}^{n} w_i T_{n4}^2(X_i, Z_i) = o_p(1/h_1).$$

It follows that

$$\begin{aligned} T_1 &= \sum_{i=1}^{n} w_i T_{n1}^2(X_i, Z_i) + o_p(h_1^{-1}) \\ &\equiv \tilde{T}_1 + o_p(h_1^{-1}). \end{aligned}$$

It can be rewritten that

$$\tilde{T}_1 = \sum_{i=1}^{n} w_i [B_{n1}^*(X_i, Z_i) - B_{n2}^*(X_i, Z_i) - B_{n3}(X_i, Z_i)]^2,$$

where $B_{n1}^*(x, z) = \frac{1}{n} \sum_{j=1}^{n} W_n(X_j - x, x; h_1)\varepsilon_j^*(z)$,

$$B_{n2}^*(x, z) = \frac{1}{n} \sum_{j=1}^{n} W_n(Y_j - x, x; h_3) e_{j+1}(z)$$

and

$$\begin{aligned} B_{n3}(x, z) &= \frac{1}{n} \sum_{j=1}^{n} q^*(x, Z_j)\varepsilon_{j+1}(z) \\ &= \frac{1}{n} \frac{1}{\pi(x)} \sum_{j=1}^{n} p(Z_j | x, \Delta)\pi(x)\pi^{-1}(Z_j)\varepsilon_{j+1}(z). \end{aligned}$$

Applying Lemmas 1 and 2 and using Condition (A5), we obtain that

$$\tilde{T}_1 = \sum_{i=1}^{n} w_i \{B_{n1}(X_i, Z_i) - B_{n2}(X_i, Z_i) - B_{n3}(X_i, Z_i)\}^2 + o_p(h_1^{-1}),$$

where $B_{n1}(x, z) = \frac{1}{n} \frac{1}{\pi(x)} \sum_{j=1}^{n} W_{h_1}(X_j - x)\varepsilon_j^*(z)$ and

$$B_{n2}(x, z) = \frac{1}{n} \frac{1}{\pi(x)} \sum_{j=1}^{n} W_{h_3}(Y_j - x) e_{j+1}(z).$$

Hence,

$$T_1 = \sum_{i=1}^{n} w_i \{B_{n1}(X_i, Z_i) - B_{n2}(X_i, Z_i) - B_{n3}(X_i, Z_i)\}^2 + o_p(h_1^{-1}).$$

Let $\xi(i, j) = W_{h_1}(X_j - X_i)\varepsilon_j^*(Z_i) - W_{h_3}(Y_j - X_i) e_{j+1}(Z_i) - q(X_i, Z_j)\varepsilon_{j+1}(Z_i)$ and

$$\psi(i, j, k) = n^{-2} w_i \pi^{-2}(X_i)\xi(i, j)\xi(i, k),$$

where $q(x, z) = p(z|x, \Delta)\pi(x)/\pi(z) = p^*(x|z, \Delta)$. Then

$$T_1 = \sum_{i,j,k=1}^{n} \psi(i, j, k) + o_p(h_1^{-1}).$$

(ii) *Derive the asymptotics using the asymptotic theory for the U-statistic.* Let

$$\begin{aligned} B_{11} = \sum_{i<j<k} \{&\psi(i, j, k) + \psi(i, k, j) + \psi(j, i, k) \\ &+ \psi(j, k, i) + \psi(k, i, j) + \psi(k, j, i)\}, \end{aligned}$$

$$B_{12} = \sum_{i \neq j} [\psi(i, j, j) + \psi(j, i, j) + \psi(j, j, i)]$$

and

$$B_{13} = \sum_{i=1}^{n} \psi(i, i, i).$$

Then

$$T_1 = B_{11} + B_{12} + B_{13} + o_p(h_1^{-1}). \tag{38}$$

Let $\psi^*(i, j, k) = \psi(i, j, k) + \psi(i, k, j) + \psi(j, i, k) + \psi(j, k, i) + \psi(k, i, j) + \psi(k, j, i)$. Then $\psi^*(i, j, k)$ is symmetrical about (i, j, k), and hence $B_{11} = \sum_{i<j<k} \psi^*(i, j, k)$. Using Hoeffding's decomposition, we obtain that

$$B_{11} = \sum_{i<j<k} \Phi(i, j, k) + (n-2) \sum_{1 \le i<j \le n} \psi^*(i, j), \tag{39}$$

where

$$\Phi(i, j, k) = \psi^*(i, j, k) - \psi^*(i, j) - \psi^*(i, k) - \psi^*(j, k),$$

$\psi^*(i, j) = \int \psi^*(i, j, k)\, dF(x_k, y_k, z_k)$ and F is the distribution of (X_k, Y_k, Z_k). Applying the lemma with $\delta = 1/3$ in [19], we can show that $E\{\sum_{i<j<k} \Phi(i, j, k)\}^2 = o(h_1^{-2})$. Therefore, the first term on the right-hand side of (39) is $o_p(h_1^{-1})$, so that

$$B_{11} = (n-2) \sum_{1 \le i<j \le n} \psi^*(i, j) + o_p(h_1^{-1}). \tag{40}$$

By the Markovian property of $\{X_i\}$, $E[\psi^*(i, j)] = 0$. Hence, up to a ignorable term of order $o_p(h_1^{-1})$, B_{11} is a U-statistic with mean zero. Define $\tilde{\psi}(i, j) = \psi(i, i, j) + \psi(i, j, i) + \psi(j, i, i) + \psi(j, j, i) + \psi(j, i, j) + \psi(i, j, j)$, $\tilde{\psi}(i) = \int \tilde{\psi}(i, j)\, dF(x_j, y_j, z_j)$ and $\tilde{\psi}(0) = E[\tilde{\psi}(i)]$. Then we have

$$B_{12} = \sum_{1 \le i<j \le n} \tilde{\psi}(i, j).$$

Since $\tilde{\psi}(i, j)$ is a symmetrical kernel, using the Hoeffding decomposition, we obtain that

$$
\begin{aligned}
B_{12} = & \sum_{1\le i<j\le n} [\tilde{\psi}(i, j) - \tilde{\psi}(i) - \tilde{\psi}(j) + \tilde{\psi}(0)] \\
& + (n-1)\sum_{i=1}^{n}[\tilde{\psi}(i) - \tilde{\psi}(0)] + \frac{1}{2}n(n-1)\tilde{\psi}(0).
\end{aligned} \tag{41}
$$

By Lemma 5,

$$
B_{12} = \tfrac{1}{2}n(n-1)\tilde{\psi}(0) + o_p(h_1^{-1}). \tag{42}
$$

Note that $B_{13} \ge 0$. By straightforward calculation on the mean of B_{13}, it can be shown that

$$
B_{13} = O_p\big(n/(n^2h_1^2h_2^2)\big) = o_p(h_1^{-1}). \tag{43}
$$

Therefore, a combination of (38) and (40)–(43) leads to

$$
T_1 = \frac{1}{2}n(n-1)\tilde{\psi}(0) + (n-2)\sum_{1\le i<j\le n}\psi^*(i, j) + o_p(h_1^{-1}). \tag{44}
$$

By Lemma 6(i),

$$
\tfrac{1}{2}n(n-1)\tilde{\psi}(0) = \mu_1 + o_p(h_1^{-1}).
$$

Applying Lemma 7(i), we obtain that

$$
(n-2)\sum_{i<j}\psi^*(i, j)/\sigma_1 \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),
$$

where $\sigma_1^2 = 2\Omega_2\|W * W\|^2\|K * K\|^2/(h_1h_2)$. Therefore, the result of this theorem holds. $\square$

PROOF OF THEOREM 2. The proof is similar to that of Theorem 1.

(i) *Asymptotic expression for* $\hat{P}(z|x, 2\Delta) - \hat{R}(z|x, 2\Delta)$. By the definitions in (13) and (14),

$$
\begin{aligned}
\hat{P}(z|x, 2\Delta) - P(z|x, 2\Delta) = & \frac{1}{n}\sum_{i=1}^{n} W_n(X_i - x, x; h_1) \\
& \times [I(Z_i < z) - P(z|x, 2\Delta)],
\end{aligned} \tag{45}
$$

$$
\hat{R}(z|x, 2\Delta) - R(z|x, 2\Delta) = S_{n1}(x, z) + S_{n2}(x, z), \tag{46}
$$

where $S_{n1}(x, z) = n^{-1}\sum_{i=1}^{n} W_n(X_i - x, x; h_3)[\hat{P}(z|Y_i, \Delta) - P(z|Y_i, \Delta)]$ and

$$
S_{n2}(x, z) = n^{-1}\sum_{i=1}^{n} W_n(X_i - x, x; h_3)[P(z|Y_i, \Delta) - R(z|x, 2\Delta)]. \tag{47}
$$

Let $u_i(z,\Delta) = I(Z_i < z) - P(z|Y_i,\Delta)$. Then $E[u_i(z,\Delta)] = 0$. By (5),

$$\hat{P}(z|y,\Delta) - P(z|y,\Delta) = n^{-1}\sum_{i=1}^{n} W_n(Y_i - y, y; b_1)[I(Z_i < z) - P(z|y,\Delta)].$$

This can be rewritten as

$$\hat{P}(z|y,\Delta) - P(z|y,\Delta) = P_{n1}(y,z) + P_{n2}(y,z), \tag{48}$$

where

$$P_{n1}(y,z) = n^{-1}\sum_{i=1}^{n} W_n(Y_i - y, y; b_1)u_i(z,\Delta),$$

$$P_{n2}(y,z) = n^{-1}\sum_{i=1}^{n} W_n(Y_i - y, y; b_1)[P(z|Y_i,\Delta) - P(z|y,\Delta)].$$

By Lemma 1 and the symmetry of the kernel function $W(\cdot)$, and by using Taylor's expansion, it is easy to show that

$$P_{n2}(y,z) = (\partial^2/\partial y^2)P(z|y,\Delta)b_1^2 + o_p(b_1^2) = O_p(b_1^2), \tag{49}$$

uniformly for $(y,z) \in \Omega^*$. Hence,

$$\hat{P}(z|y,\Delta) - P(z|y,\Delta) = P_{n1}(y,z) + O_p(b_1^2), \tag{50}$$

uniformly for $(y,z) \in \Omega^*$. Then

$$S_{n1}(x,z) = n^{-1}\sum_{i=1}^{n} W_n(X_i - x, x; h_3)P_{n1}(Y_i,z) + O_p(b_1^2), \tag{51}$$

uniformly for $(x,z) \in \Omega^*$. Using the same arguments as those for $L_{n11}(x,z)$ between (32) and (37), we obtain that

$$\begin{aligned} S_{n1}(x,z) &= \frac{1}{n}\sum_{i=1}^{n} q^*(x,Y_i)u_i(z,\Delta) \\ &\quad + O_p\big(\{\log n/(nb_1^{3/2})\} + \{b_1(\log n)/n\}^{1/2}\big) \\ &= \frac{1}{n}\sum_{i=1}^{n} q^*(x,Z_i)u_{i+1}(z,\Delta) \\ &\quad + O_p\bigg(\frac{\log n}{nb_1^{3/2}} + \{b_1(\log n)/n\}^{1/2}\bigg). \end{aligned} \tag{52}$$

Rewrite $S_{n2}(x,z)$ as

$$S_{n2}(x,z) = S_{n21}(x,z) + S_{n22}(x,z),$$

where

$$S_{n21}(x,z)=n^{-1}\sum_{i=1}^{n}W_n(X_i-x,x;h_3)[P(z|Y_i,\Delta)-R(z|X_i,2\Delta)],$$

$$S_{n22}(x,z)=n^{-1}\sum_{i=1}^{n}W_n(X_i-x,x;h_3)[R(z|X_i,2\Delta)-R(z|x,2\Delta)].$$

By the continuity of $\partial^2 R(z|x,2\Delta)/\partial x^2$ and the same argument as that for (49), $S_{n22}(x,z)=O_p(h_3^2)$, uniformly for $(x,z)\in\Omega^*$. Let $e_i^*(z)=P(z|Y_i,\Delta)-R(z|X_i,2\Delta)$. Then $E[e_i^*(z)|X_i]=0$, and

$$\begin{aligned}S_{n2}(x,z)&=n^{-1}\sum_{i=1}^{n}W_n(X_i-x,x;h_3)e_i^*(z)+O_p(h_3^2)\\&=n^{-1}\sum_{i=1}^{n}W_n(Y_i-x,x;h_3)e_{i+1}^*(z)+O_p(h_3^2).\end{aligned}\tag{53}$$

By (45) and (46), under H_0, we have

$$\hat{P}(z|x,2\Delta)-\hat{R}(z|x,2\Delta)=-S_{n1}(x,z)-S_{n2}(x,z)+S_{n3}(x,z),\tag{54}$$

where, with $u_j^*(z,2\Delta)=I(Z_j<z)-P(z|X_j,2\Delta)$,

$$\begin{aligned}S_{n3}(x,z)&=\frac{1}{n}\sum_{i=1}^{n}W_n(X_i-x,x;h_1)[I(Z_i<z)-P(z|x,2\Delta)]\\&=\frac{1}{n}\sum_{i=1}^{n}W_n(X_i-x,x;h_1)u_i^*(z,2\Delta)\\&\quad+\frac{1}{n}\sum_{i=1}^{n}W_n(X_i-x,x;h_1)[P(z|X_i,2\Delta)-P(z|x,2\Delta)].\end{aligned}$$

Similarly to (49), the second term above is of order $O_p(h_1^2)$,

$$S_{n3}(x,z)=\frac{1}{n}\sum_{i=1}^{n}W_n(X_i-x,x;h_1)u_i^*(z,2\Delta)+O_p(h_1^2),\tag{55}$$

uniformly for $(x,z)\in\Omega^*$. A combination of (52)–(55) yields that

$$\hat{P}(z|x,2\Delta)-\hat{R}(z|x,2\Delta)=T_{n1}^*(x,z)+T_{n2}^*(x,z)+T_{n3}^*(x,z),\tag{56}$$

where

$$\begin{aligned}T_{n1}^*(x,z)=\frac{1}{n}\sum_{j=1}^{n}[&W_n(X_j-x,x;h_1)u_j^*(z,2\Delta)\\&-W_n(Y_j-x,x;h_3)e_{j+1}^*(z)-q^*(x,Z_j)u_{j+1}(z,\Delta)],\end{aligned}$$

$$T_{n2}^*(x,z)=O_p(b_1^2+h_1^2+h_3^2),$$

uniformly for $(x, z) \in \Omega^*$, and

$$T_{n3}^*(x, z) = O_p(\{\log n/(nb_1^{3/2})\} + \{b_1(\log n)/n\}^{1/2}),$$

uniformly for $(x, z) \in \Omega^*$.

(ii) *Asymptotic normality of* T_2. Similar to (44), we have

$$T_2 = \frac{1}{2}n(n-1)\tilde{\phi}(0) + (n-2)\sum_{1\le i<j\le n}\phi^*(i, j) + o_p(h^{-1}), \tag{57}$$

where $\tilde{\phi}(0)$ and $\phi^*(i, j)$ are defined the same as $\tilde{\psi}(0)$ and $\psi^*(i, j)$, respectively, but with ψ replaced by

$$\phi(i, j, k) = n^2 w_i \pi^{-2}(X_i)\eta(i, j)\eta(i, k),$$

where

$$\begin{aligned}\eta(i, j) = {} & W_{h_1}(X_j - X_i)u_j^*(Z_i, 2\Delta) - W_{h_3}(Y_j - X_i)e_{j+1}^*(Z_i) \\ & - q(X_i, Z_j)u_{j+1}(Z_i, \Delta).\end{aligned}$$

By Lemma 6(ii), we have

$$\tfrac{1}{2}n(n-1)\tilde{\phi}(0) = \mu_2 + o_p(h_1^{-1}). \tag{58}$$

By Lemma 7(ii), we have

$$(n-2)\sum_{i<j}\phi^*(i, j)/\sigma_2 \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{59}$$

A combination of (57)–(59) completes the proof of the theorem. □

PROOF OF THEOREM 3. Under H_{1n}, $p(z|x, 2\Delta) = r(z|x, 2\Delta) + g_n(x, z)$. Similarly to (22), we have under H_{1n}

$$\hat{p}(z|x, 2\Delta) - \hat{r}(z|x, 2\Delta) = Q_n(x, z) + g_n(x, z),$$

where

$$Q_n(x, z) = A_n^*(x, z) + B_n^*(x, z) + C_n^*(x, z) - L_{n1}(x, z) - L_{n2}(x, z) - L_{n3}(x, z).$$

Then

$$\begin{aligned}T_1 = {} & \sum_{i=1}^n Q_n^2(X_i, Z_i)w_i + \sum_{i=1}^n g_n^2(X_i, Z_i)w_i \\ & + 2\sum_{i=1}^n g_n(X_i, Z_i)Q_n(X_i, Z_i)w_i.\end{aligned} \tag{60}$$

Since $\delta_n^2 = O(\frac{1}{nh_1h_2})$, it can be shown that

$$\sum_{i=1}^{n} g_n^2(X_i, Z_i)w_i = nE[g_n^2(X, Z)w(X, Z)] + o_p(1/\sqrt{h_1h_2}). \tag{61}$$

By (20) and (23), $B_n^*(x,z) = O_p(h_1^2)$, $C_n^*(x,z) = O_p(h_2^2)$ and $L_{n3}(x,z) = O_p(h_3^2)$, uniformly for $(x,z) \in \Omega^*$. It follows from the Hölder inequality that

$$\begin{aligned}
&2\sum_{i=1}^{n} w_i g_n(X_i, Z_i)[B_n^*(X_i, Z_i) + C_n^*(X_i, Z_i) - L_{n3}(X_i, Z_i)] \\
&\qquad = O_p\bigl(n\delta_n(h_1^2 + h_2^2 + h_3^2)\bigr).
\end{aligned} \tag{62}$$

A combination of (60)–(62) yields that

$$\begin{aligned}
T_1 = {}& \sum_{i=1}^{n} Q_n^2(X_i, Z_i)w_i + nE[g_n^2(X, Z)w(X, Z)] \\
& + 2\sum_{i=1}^{n} g_n(X_i, Z_i)w_i[A_n^*(X_i, Z_i) - L_{n2}(X_i, Z_i) - L_{n1}(X_i, Z_i)] \\
& + \{o_p(1/\sqrt{h_1h_2}) + O_p\bigl(n\delta_n(h_1^2 + h_2^2 + h_3^2)\bigr)\} \\
\equiv {}& T_{11} + T_{12} + T_{13} + o_p(1/\sqrt{h_1h_2}).
\end{aligned} \tag{63}$$

T_{11} can be dealt with in the same way as in the proof of Theorem 1. It is asymptotically normal with mean μ_1 and variance σ_1^2 given in Theorem 1. By the definition, $T_{12} = d_{1n}$. We now study the third term T_{13}. By (27) and (36), T_{13} admits the following decomposition:

$$\begin{aligned}
\frac{1}{2}T_{13} &= \sum_{i=1}^{n} g_n(X_i, Z_i)w_i[A_n^*(X_i, Z_i) - L_{n2}(X_i, Z_i) - L_{n1}(X_i, Z_i)] \\
&= \sum_{i=1}^{n} g_n(X_i, Z_i)w_i \frac{1}{n}\sum_{j=1}^{n}\{W_n(X_j - X_i, X_i; h_1)\varepsilon_j^*(Z_i) \\
&\qquad\qquad - W_n(Y_j - X_i; X_i; h_3)e_{j+1}(Z_i) \\
&\qquad\qquad - q^*(X_i, Z_j)\varepsilon_{j+1}(Z_i)\} \\
&\quad + o_p(1/\sqrt{h_1h_2}) + O\bigl(n\delta_n(b_1^2 + b_2^2)\bigr) + O(\delta_n h_1^{-1}h_2^{-1}) \\
&= \sum_{i\neq j} \frac{1}{n} g_n(X_i, Z_i)w_i\pi^{-1}(X_i)\{W_{h_1}(X_j - X_i)\varepsilon_j^*(Z_i) \\
&\qquad\qquad - W_{h_3}(Y_j - X_i)e_{j+1}(Z_i) \\
&\qquad\qquad - q^*(X_i, Z_j)\varepsilon_{j+1}(Z_i)\}
\end{aligned}$$

$$
\begin{aligned}
&+ o_p\bigl(1/\sqrt{h_1h_2}\bigr) + O\bigl(n\delta_n(b_1^2+b_2^2)\bigr) + O(\delta_n h_1^{-1}h_2^{-1})\\
&\equiv \sum_{i\neq j}\varphi(i,j) + o_p\bigl(1/\sqrt{h_1h_2}\bigr) + O\bigl(n\delta_n(b_1^2+b_2^2)\bigr) + O\bigl(\delta_n/(h_1h_2)\bigr).
\end{aligned}
$$

The first term above is a U-statistic with the typical element $\varphi(i,j)$. Let $\varphi^*(i,j) = \varphi(i,j)+\varphi(j,i)$. Then $\varphi^*(i,j)$ is a symmetric kernel and

$$
T_{13} = \sum_{1\le i<j\le n}\varphi^*(i,j) + O\bigl(\delta_n/(h_1h_2)\bigr) + o_p\bigl(1/\sqrt{h_1h_2}\bigr).
$$

Put $\tilde\varphi(i) = \int\varphi^*(i,j)\,dF_j$ and $\tilde\varphi(i,j) = \varphi^*(i,j) - \tilde\varphi(i) - \tilde\varphi(j)$. Then by the Hoeffding decomposition, we have

$$
\sum_{1\le i<j\le n}\varphi^*(i,j) = \sum_{1\le i<j\le n}\tilde\varphi(i,j) + (n-1)\sum_{i=1}^n\tilde\varphi(i).
$$

It is easy to show that $E[h_1h_2\tilde\varphi(i,j)]^{2(1+\delta)} = O(\delta_n^{2(1+\delta)}n^{-2(1+\delta)}h_1h_2)$. Therefore, applying the lemma with $\delta=1$ of [19], we obtain that

$$
E\Biggl\{\sum_{1\le i<j\le n}\tilde\varphi(i,j)\Biggr\}^2 = o\bigl(1/(h_1h_2)\bigr).
$$

Therefore,

$$
T_{13} = (n-1)\sum_{i=1}^n\tilde\varphi(i) + o_p\bigl(1/\sqrt{h_1h_2}\bigr) + O\bigl(\delta_n/(h_1h_2)\bigr). \tag{64}
$$

By the definition of $\tilde\varphi_i$, it can be written that

$$
\begin{aligned}
\tilde\varphi(i) &= \frac{2}{n}g_n(X_i,Z_i)w(X_i,Z_i)\pi^{-1}(X_i)\\
&\quad\times\int\{W_{h_1}(x_j-X_i)\varepsilon_j^*(Z_i) - W_{h_3}(y_j-X_i)e_{j+1}(Z_i)\\
&\qquad\qquad - q^*(X_i,z_j)\varepsilon_{j+1}(Z_i)\}\,dF_j\\
&\equiv \tilde\varphi_1(i)+\tilde\varphi_2(i)+\tilde\varphi_3(i),
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde\varphi_1(i) &= \frac{2}{n}g_n(X_i,Z_i)w(X_i,Z_i)\pi^{-1}(X_i)\int W_{h_1}(x_j-X_i)\varepsilon_j^*(Z_i)\,dF_j,\\
\tilde\varphi_2(i) &= -\frac{2}{n}g_n(X_i,Z_i)w(X_i,Z_i)\pi^{-1}(X_i)\int W_{h_3}(y_j-X_i)e_{j+1}(Z_i)\,dF_j
\end{aligned}
$$

and $\tilde\varphi_3(i) = -\frac{2}{n}g_n(X_i,Z_i)w(X_i,Z_i)\pi^{-1}(X_i)\int q^*(X_i,z_j)\varepsilon_{j+1}(Z_i)\,dF_j$. Then by the Fubini theorem and by taking iterative expectation, $E[\tilde\varphi(i)]=0$. Using the central limit theorem for the β-mixing process, we get

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

where $\sigma_{1A}^2=\frac{1}{4}nE[(n-1)^2\tilde{\varphi}^2(i)]$. By directly calculating the integration, it can be shown that

$$\tilde{\varphi}_1(i)=\frac{2}{n}g_n(X_i,Z_i)w(X_i,Z_i)[p(Z_i|X_i,2\Delta)-p^2(Z_i|X_i,2\Delta)]\bigl(1+o(1)\bigr),$$

$\tilde{\varphi}_2(i)=o(g_n(X_i,Z_i)/n)$ and $\tilde{\varphi}_3(i)=o(g_n(X_i,Z_i)/n)$. Therefore,

$$\begin{aligned}\sigma_{1A}^2&=nE[g_n^2(X_1,Z_1)w^2(X_1,Z_1)\{p(Z_i|X_i,2\Delta)-p^2(Z_i|X_i,2\Delta)\}^2]\\&\quad+o\bigl(1/(h_1h_2)\bigr).\end{aligned}$$

By straightforward calculation, it can be shown that the covariance between T_{11} and T_{13} can be ignored. It follows that the result of the theorem holds. □

PROOF OF THEOREM 4. (i) For any given small $\eta>0$, when d is small enough, $|d_{1n}/\sigma_{1n}|\le\eta$ and $\sigma_{1n}=\sigma_1(1+o(1))$. Under H_0, with the selected bandwidths,

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

Therefore, the sequence of critical values c_α (depending on n) is bounded in probability. Similarly, under H_{1n}, with the selected bandwidths,

$$(T_1-\mu_1-d_{1n})/\sigma_{1n}=O_p(1). \tag{65}$$

Note that

$$\begin{aligned}P\{(T_1-\mu_1)/\sigma_1>c_\alpha|H_{1n}\}&=P\{(T_1-\mu_1-d_{1n})/\sigma_{1n}>(c_\alpha\sigma_1-d_{1n})/\sigma_{1n}|H_{1n}\}\\&\le P\{(T_1-\mu_1-d_{1n})/\sigma_{1n}>c_\alpha\sigma_1/\sigma_{1n}-\eta|H_{1n}\}.\end{aligned}$$

It follows from Theorem 3 and Slutsky's theorem that

$$\limsup_{d\to0}\limsup_{n\to\infty}P\{(T_1-\mu_1)/\sigma_1\ge c_\alpha|H_{1n}\}\le\alpha.$$

(ii) For any given $M>0$, by taking d sufficiently large, there exists an N, when $n>N$, $d_{1n}/\sigma_{1n}\ge M$. Therefore,

$$P\{(T_1-\mu_1)/\sigma_1>c_\alpha|H_{1n}\}\ge P\{(T_1-\mu_1-d_{1n})/\sigma_{1n}>c_\alpha\sigma_1/\sigma_{1n}-M|H_{1n}\}.$$

By (65), we have

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

PROOF OF THEOREMS 5 AND 6. We put the proofs in the supplemental materials [2]. □

Acknowledgments. The authors thank the Associate Editor and the referees for constructive comments that substantially improved an earlier version of this paper.

SUPPLEMENTARY MATERIAL

Supplement: Additional technical details (DOI: 10.1214/09-AOS763SUPP; .pdf). We provide detailed proofs for Lemmas 1–7 and Theorems 5–6. Modern nonparametric smoothing techniques and theory of U-statistics are used.

REFERENCES

[1] AÏT-SAHALIA, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies* **9** 385–426.

[2] AÏT-SAHALIA, Y., FAN, J. and JIANG, J. (2010). Supplement to "Nonparametric tests of the Markov hypothesis in continuous-time models." DOI: 10.1214/09-AOS763SUPP.

[3] AÏT-SAHALIA, Y., FAN, J. and PENG, H. (2009). Nonparametric transition-based tests for jump-diffusions. *J. Amer. Statist. Assoc.* **104** 1102–1116.

[4] AZZALINI, A., BOWMAN, A. N. and HÄRDLE, W. (1989). On the use of nonparametric regression for model checking. *Biometrika* **76** 1–11. MR0991417

[5] BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviation of density function estimates. *Ann. Statist.* **1** 1071–1095. MR0348906

[6] CAVERHILL, A. (1994). When is the short rate Markovian? *Math. Finance* **4** 305–312. MR1299241

[7] CHEN, S. X. and GAO, J. (2004). On the use of the kernel method for specification tests of diffusion models. Technical report, Iowa State Univ.

[8] CHEN, S. X., GAO, J. and TANG, C. (2008). A test for model specification of diffusion processes. *Ann. Statist.* **36** 167–198. MR2387968

[9] COX, J. C., INGERSOLL, J. E. and ROSS, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53** 385–408. MR0785475

[10] FAN, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 674–688. MR1395735

[11] FAN, J. and JIANG, J. (2005). Generalized likelihood ratio tests for additive models. *J. Amer. Statist. Assoc.* **100** 890–907. MR2201017

[12] FAN, J. and JIANG, J. (2007). Nonparametric inference with generalized likelihood ratio tests (with discussion). *Test* **16** 409–478. MR2365172

[13] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York. MR1964455

[14] FAN, J., YAO, Q. and TONG, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* **83** 189–206. MR1399164

[15] FAN, J. and YIM, T.-H. (2004). A data-driven method for estimating conditional densities. *Biometrika* **91** 819–834. MR2126035

[16] FAN, J., ZHANG, C. and ZHANG, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** 153–193. MR1833962

[17] FOUQUE, J.-P., PAPANICOLAOU, G. and SIRCAR, K. R. (2000). *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge Univ. Press, London. MR1768877

[18] GAO, J. and CASAS, I. (2008). Specification testing in discretized diffusion models: Theory and practice. *J. Econometrics* **147** 131–140. MR2472987

[19] GAO, J. and KING, M. (2004). Model specification testing in nonparametric and semiparametric time series econometrics. Technical report, Univ. Western Australia.

[20] HALL, P., RACINE, J. and LI, Q. (2004). Cross-validation and the estimation of conditional probability densities. *J. Amer. Statist. Assoc.* **99** 1015–1026. MR2109491

[21] HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. MR1245774

[22] HEATH, D., JARROW, R. and MORTON, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims evaluation. *Econometrica* **60** 77–105.

[23] HESTON, S. (1993). A closed-form solution for options with stochastic volatility with applications to bonds and currency options. *Review of Financial Studies* **6** 327–343.

[24] 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.

[25] HYNDMAN, R. and YAO, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *J. Nonparametr. Statist.* **14** 259–278. MR1905751

[26] INGSTER, Y. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives I–III. *Math. Methods Statist.* **2** 85–114; **3** 171–189; **4** 249–268.

[27] LEPSKI, O. and SPOKOINY, V. (1999). Minimax nonparametric hypothesis testing: The case of an inhomogeneous alternative. *Bernoulli* **5** 333–358. MR1681702

[28] REVUZ, D. and YOR, M. (1994). *Continuous Martingales and Brownian Motion*, 2nd ed. Springer, Berlin. MR1303781

[29] SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962

[30] VASICEK, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* **5** 177–188.

Y. AÏT-SAHALIA
DEPARTMENT OF ECONOMICS
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
USA
AND
NBER
1050 MASSACHUSETTS AVE.
CAMBRIDGE, MASSACHUSETTS 02138
USA
E-MAIL: yacine@princeton.edu

J. FAN
DEPARTMENT OF ORFE
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: jqfan@princeton.edu

J. JIANG
DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF NORTH CAROLINA
AT CHARLOTTE
CHARLOTTE, NORTH CAROLINA 28223
USA
E-MAIL: jjiang1@uncc.edu