Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 44–64
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Testing the Markov property with high frequency data

João Amaro de Matos[a,*], Marcelo Fernandes[b]

[a]Faculdade de Economia, Universidade Nova de Lisboa, Rua Marquês de Fronteira, 20, 1099-038 Lisbon, Portugal
[b]Economics Department, Queen Mary, University of London, Mile End Road, E1 4NS, London, UK

Available online 6 March 2007

Abstract

This paper develops a framework to nonparametrically test whether discrete-valued irregularly spaced financial transactions data follow a Markov process. For that purpose, we consider a specific optional sampling in which a continuous-time Markov process is observed only when it crosses some discrete level. This framework is convenient for it accommodates the irregular spacing that characterizes transactions data. Under such an observation rule, the current price duration is independent of a previous price duration given the previous price realization. A simple nonparametric test then follows by examining whether this conditional independence property holds. Monte Carlo simulations suggest that the asymptotic test has huge size distortions, though a bootstrap-based variant entails reasonable size and power properties in finite samples. As for an empirical illustration, we investigate whether bid–ask spreads follow Markov processes using transactions data from the New York Stock Exchange. The motivation lies on the fact that asymmetric information models of market microstructures predict that the Markov property does not hold for the bid–ask spread. We robustly reject the Markov assumption for two out of the five stocks under scrutiny. Finally, it is reassuring that our results are consistent with two alternative measures of asymmetric information.

JEL classification: C14; C52; G10; G19

Keywords: Bid–ask spread; Conditional independence; Duration; Nonparametric testing; U-statistic

*Corresponding author.
E-mail addresses: amatos@fe.unl.pt (J. Amaro de Matos), m.fernandes@qmul.ac.uk (M. Fernandes).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.01.007

1. Introduction

Despite the innumerable studies in financial economics rooted in the Markov property, there are only two tests available in the literature to check such an assumption: Aït-Sahalia (2000) and Fernandes and Flôres (2004). To build a nonparametric testing procedure, the first test uses the fact that the Chapman–Kolmogorov equation must hold in order for a process to be compatible with the Markov assumption (see also Aït-Sahalia, 2002). Although the Chapman–Kolmogorov representation involves a quite complicated nonlinear functional relationship among transition probabilities of the process, it brings about several advantages. First, estimating transition distributions is straightforward and does not require any prior parameterization of conditional moments. Second, a test based on the whole transition density is obviously preferable to tests based on specific conditional moments. Third, the Chapman–Kolmogorov representation is well defined, even within a multivariate context.

Fernandes and Flôres (2004) develop alternative ways of testing whether discretely recorded observations are consistent with an underlying Markov process. Instead of using the highly nonlinear functional characterization provided by the Chapman–Kolmogorov equation, they rely on a simple characterization out of a set of necessary conditions for Markov models. As in Aït-Sahalia (2000), the testing strategy boils down to measuring the closeness of density functionals that are nonparametrically estimated by kernel-based methods.

Both testing procedures assume, however, that the data are evenly spaced in time. Financial transactions data do not satisfy such an assumption and hence these tests are not appropriate. To design a test for the Markov property that is suitable to high frequency data, we build on the theory of Markov processes with stochastic time changes. We consider a particular optional sampling for the underlying continuous-time Markov process $\{\tilde{X}_t;\ t>0\}$ that yields discrete-time realizations $\{\tilde{X}_{t_1},\ldots,\tilde{X}_{t_n}\}$ as the cumulative change in $\tilde{X}_t$ adds to a discrete level c. Accordingly, we accommodate the irregular spacing that characterizes transactions data. Further, such an optional sampling scheme implies that consecutive spells between the observed changes in $\tilde{X}_t$ are conditionally independent given the discrete-time realization of $\tilde{X}_t$. We then develop a simple nonparametric test for the Markov property by testing whether the conditional independence property holds.

There is a large literature on how to test either unconditional independence (e.g., Hoeffding, 1948; Rosenblatt, 1975; Pinkse, 1999) or serial independence (e.g., Robinson, 1991; Skaug and Tjøstheim, 1993; Pinkse, 1998). However, there are only a few papers discussing tests of conditional independence: Linton and Gozalo (1999) and, more recently, Su and White (2002, 2003a,b). Linton and Gozalo (1999) test for conditional independence between iid random variables by looking at the restrictions on the cumulative distribution function under a quadratic measure of distance. Su and White (2002, 2003a,b) extend their approach by considering weakly dependent stochastic processes as well as different metrics. In particular, Su and White (2002) verify whether the density restriction implied by conditional independence holds using the Hellinger distance, whereas Su and White (2003a,b) check restrictions on the characteristic function and on the empirical likelihoods, respectively. Our setting combines the setups of Linton and Gozalo (1999) and Su and White (2002). As in Su and White (2002), we derive tests under mixing conditions so as to deal with the time-series dependence of the data. However, we gauge how well the density restriction implied by the conditional independence property fits the data using a quadratic measure of distance as in Linton and Gozalo (1999).

A relevant application of our testing procedure is to check whether bid–ask spreads follow Markov processes. Asymmetric information models of market microstructures predict that the bid–ask spread depends on the whole trading history, and hence the Markov property does not hold (e.g., Easley and O'Hara, 1992). Our nonparametric approach to test the Markov property is consistent with Hasbrouck's (1991) goal to uncover the extent of adverse selection costs in a framework that is robust to deviations from the assumptions of the formal models of market microstructure. Bearing that in mind, we examine transactions data from five stocks actively traded on the New York Stock Exchange (NYSE), namely, Boeing, Coke, Disney, Exxon, and IBM.

The results reveal that the Markov assumption is consistent with the Coke, Disney and Exxon bid–ask spreads, whereas the converse is true for Boeing and IBM. This indicates that the latter stocks presumably have higher rates of return in equilibrium, since uninformed traders require a compensation to hold stocks with greater private information (Easley et al., 2002). The usual objection that the actions of arbitrageurs remove any chance of higher returns does not apply because adverse selection risk is systematic. An uninformed investor indeed is always at a disadvantage relative to traders with better information. Our findings thus imply that the standard asset-pricing framework under symmetric information is not valid to examine the Boeing and IBM returns, though it may work for Coke, Disney and Exxon.

The remainder of this paper is organized as follows. Section 2 discusses how to design a nonparametric test for Markovian dynamics that is suitable to high frequency data. The asymptotic normality of the test statistic is then derived both under the null hypothesis that the Markov property holds and under a sequence of local alternatives. Section 3 reports a simulation study that evinces that, although our asymptotic test exhibits palpable size distortions, a bootstrap variant seems to entail reasonable size and power properties. Section 4 applies the above ideas to test whether the bid–ask spreads of five actively traded stocks on the NYSE follow a Markov process with stochastic time changes. Section 5 summarizes the results and offers some concluding remarks. For ease of exposition, we collect all proofs and technical lemmas in the appendix.

2. Markov processes with stochastic time changes

Let t_i $(i = 1, 2, \ldots)$ denote the observation times of the continuous-time strong stationary Markov process $\{\tilde{X}_t,\ t>0\}$ and assume that $t_0 = 0$. We consider discrete-time observations of $\{\tilde{X}_t,\ t>0\}$ stemming from a particular optional sampling that marks the continuous-time process only if the cumulative change in $\tilde{X}_t$ is at least c. More precisely, the time elapsed between two consecutive discrete-time observations is

$$\tilde{d}_{i+1} \equiv t_{i+1} - t_i = \inf_{\tau>0}\{|\tilde{X}_{t_i+\tau} - \tilde{X}_{t_i}| \geqslant c\} \tag{1}$$

for $i = 0, \ldots, n-1$. The data available for statistical inference are the durations $(\tilde{d}_1, \ldots, \tilde{d}_N)$ and the corresponding discrete-time realizations $(\tilde{X}_1, \ldots, \tilde{X}_N)$, with $\tilde{X}_i = \tilde{X}_{t_i}$.

The observation times $\{t_i,\ i = 1, 2, \ldots\}$ form a sequence of increasing stopping times of the continuous-time Markov process $\{\tilde{X}_t,\ t>0\}$, hence the discrete-time process $\{\tilde{X}_i,\ i = 1, 2, \ldots\}$ satisfies the Markov property as well. Furthermore, the duration $\tilde{d}_{i+1}$ is a measurable function of the path of $\{\tilde{X}_t,\ 0 < t_i \leqslant t \leqslant t_{i+1}\}$, and thus depends on the information available at time t_i only through $\tilde{X}_i$ (see Burgayran and Darolles, 1997). In other words, the current duration is independent of the previous duration conditional on the previous discrete-time realization. It is therefore natural to test the Markov assumption by checking whether the property of conditional independence between consecutive durations holds. The next section pursues such a testing strategy.

2.1. A simple nonparametric test

As the conditional independence property is invariant to monotonic transformations, we carry out two data transformations to simplify matters. The first refers to a log-transformation and aims at circumventing the boundary bias inherent to symmetric kernels due to support boundness. The second relates to normalizing the log-transformed data to zero mean and unit variance. This permits using the same bandwidth for the resulting duration and spread series $\{(d_i, X_i),\ i = 1, \ldots, N\}$.

Assume the existence of the joint density f_{iXj} of (d_i, X_j, d_j) for $i > j$, and let $f_{i|X}$ and f_{Xj} denote the conditional density of d_i given X_j and the joint density of (X_j, d_j), respectively. The null hypothesis of conditional independence then reads

$$\mathrm{H}_0^* : f_{iXj}(a_1, x, a_2) = f_{i|X}(a_1|x) f_{Xj}(x, a_2) \quad \text{a.e. for every } j < i.$$

It is of course unfeasible to test such a restriction for all past realizations d_j of the duration process. Accordingly, it is convenient to fix j as in the pairwise approach taken by the serial independence literature (see Skaug and Tjøstheim, 1993). The resulting null hypothesis thus is the necessary condition

$$\mathrm{H}_0 : f_{iXj}(a_1, x, a_2) = f_{i|X}(a_1|x) f_{Xj}(x, a_2) \quad \text{a.e. for a fixed } j. \tag{2}$$

To keep the nonparametric nature of the testing procedure, we employ kernel smoothing to estimate both the right- and left-hand sides of (2). Next, it suffices to gauge how well the density restriction in (2) fits the data by the means of some discrepancy measure.

For the sake of simplicity, we consider a test statistic based on a weighted mean squared difference, namely,

$$\Lambda_f = \mathbb{E}\{w_{iXj}[f_{iXj}(d_i, X_j, d_j) - f_{i|X}(d_i|X_j) f_{Xj}(X_j, d_j)]^2\}, \tag{3}$$

where $w_{iXj} = w_{iXj}(d_i, X_j, d_j) = \mathbf{1}[(d_i, X_j, d_j) \in \mathscr{S}]$ trims the data down to a compact support $\mathscr{S}$ through the indicator function $\mathbf{1}(\cdot)$ that takes value one if the argument is true, zero otherwise. We employ a weighting scheme so as to avoid the lack of precision that afflicts conditional density estimation in areas of low density of the conditioning variables. The mean squared distance is convenient because it facilitates the derivation of the asymptotic theory. Bickel and Rosenblatt (1973), Hall (1984), Fan (1994), Aït-Sahalia (1996), Aït-Sahalia et al. (2001), and Fernandes and Grammig (2005) use similar distance measures. One could also consider other distance measures such as the integrated squared difference (Rosenblatt, 1975), the Kullback–Leibler contrast (Robinson, 1991), and the Hellinger metric (Su and White, 2002). See Hong and White (2004) for other entropic measures.

Letting $\jmath = i - j$ and $n = N - \jmath$, the sample analog of (3) is

$$\Lambda_{\hat{f}} = \frac{1}{n} \sum_{k=1}^{n} w_{iXj}(d_{k+\jmath}, X_k, d_k)[\hat{f}_{iXj}(d_{k+\jmath}, X_k, d_k) - \hat{g}_{iXj}(d_{k+\jmath}, X_k, d_k)]^2,$$

where $\hat{g}_{iXj}(d_{k+J}, X_k, d_k) = \hat{f}_{i|X}(d_{k+J}|X_k)\hat{f}_{Xj}(X_k, d_k) = (\hat{f}_{iX}(d_{k+J}, X_k)/\hat{f}_X(X_k))\hat{f}_{Xj}(X_k, d_k)$,

$$\hat{f}_{iXj}(a_1, x, a_2) = \frac{1}{n\,b_n^3}\sum_{k=1}^{n} K\left(\frac{a_1 - d_{k+J}}{b_n}\right)K\left(\frac{x - X_k}{b_n}\right)K\left(\frac{a_2 - d_k}{b_n}\right),$$

$$\hat{f}_{iX}(a_1, x) = \frac{1}{n\,b_n^2}\sum_{k=1}^{n} K\left(\frac{a_1 - d_{k+J}}{b_n}\right)K\left(\frac{x - X_k}{b_n}\right),$$

$$\hat{f}_{Xj}(x, a_2) = \frac{1}{n\,b_n^2}\sum_{k=1}^{n} K\left(\frac{x - X_k}{b_n}\right)K\left(\frac{a_2 - d_k}{b_n}\right),$$

$$\hat{f}_X(x) = \frac{1}{n\,b_n}\sum_{k=1}^{n} K\left(\frac{x - X_k}{b_n}\right),$$

with $K(\cdot)$ denoting the kernel function and b_n the bandwidth. At first glance, deriving the limiting distribution of $\Lambda_{\hat{f}}$ seems to involve a number of complex steps since one must deal with the cross-correlation among $\hat{f}_{iXj}$, $\hat{f}_{i|X}$ and $\hat{f}_{Xj}$. Happily, the fact that the rates of convergence of the three estimators are different simplifies things substantially. In particular, $\hat{f}_{iXj}$ converges at a slower rate than $\hat{f}_{i|X}$ and $\hat{f}_{Xj}$ due to its higher dimensionality. As such, the estimation of $g_{iXj} = f_{i|X}f_{Xj}$ does not play a major role in the asymptotic behavior of the test statistic, contributing only to the bias term (see Appendix). To derive the asymptotic theory, we impose the following regularity conditions.

Assumption A1. The sequence $\{Z_k = (d_{k+J}, X_k, d_k)\}$ is strictly stationary and β-mixing with $\beta_\tau = \mathrm{O}(\rho^\tau)$, where $0<\rho<1$.

Assumption A2. Let $e_K \equiv \int K^2(u)\,\mathrm{d}u$ and $v_K \equiv \{\int[\int K(u)K(u+v)\,\mathrm{d}u]^2\,\mathrm{d}v\}^3$, where the univariate kernel function K is symmetric around zero and of order s (even integer). We also assume that K is continuously differentiable up to the sth order on $\mathbb{R}$ with derivatives in $L^2(\mathbb{R})$.

Assumption A3. The bandwidth b_n is of order $\mathrm{O}(n^{-1/p})$, where $6<p<2s+3/2$.

Assumption A4. The density functions f_{iXj}, f_{iX}, f_{Xj} and f_X are continuously differentiable up to the sth order and their derivatives are bounded and square integrable. In addition, the marginal density f_X is bounded away from zero and the joint density function of $(Z_{k_1}, \ldots, Z_{k_\varsigma})$ exists and satisfies a Lipschitz-type condition: $|f(z_{k_1} + \Delta, \ldots, z_{k_\varsigma} + \Delta) - f(z_{k_1}, \ldots, z_{k_\varsigma})| \leqslant D(z_{k_1}, \ldots, z_{k_\varsigma})\,\|\Delta\|$, where D is integrable and $1 \leqslant \varsigma \leqslant 4$.

Assumption A1 restricts the amount of data dependence, requiring that the stochastic process is absolutely regular with geometric decay rate. Alternatively, one could assume α-mixing conditions as in Gao and King (2004), though the conditions under which a diffusion process satisfies Assumption A1 are quite weak (Aït-Sahalia, 1996). See Chen et al. (2001) for some advantages of the β-mixing assumption relative to the α-mixing condition in the context of kernel density estimation. Assumption A2 focuses on higher-order kernels so as to reduce the bias in the kernel density estimation. Assumption A3 restricts the rate at which the bandwidth must converge to zero and, in particular, requires the use of higher-order kernels. Other bandwidth conditions are also applicable, but they would result in different terms for the bias as in Härdle and Mammen (1993). Assumption A4 requires that the joint density f_{iXj} is smooth enough to admit a functional Taylor expansion and that the conditional density $f_{i|X}$ is everywhere well defined.

The following proposition documents the asymptotic normality of the test statistic.

Proposition 1. *Under the null hypothesis* H_0 *and Assumptions* A1–A4, *the statistic*

$$\hat{\lambda}_n = \frac{n b_n^{3/2} \Lambda_{\hat{f}} - b_n^{-3/2} \hat{\delta}_\Lambda - b_n^{-1/2} \hat{\xi}_\Lambda + b_n^{-1/2} \hat{\zeta}_\Lambda}{\hat{\sigma}_\Lambda} \tag{4}$$

weakly converges to a standard normal distribution provided that $\hat{\delta}_\Lambda$, $\hat{\xi}_\Lambda$, $\hat{\zeta}_\Lambda$ *and* $\hat{\sigma}^2_\Lambda$ *consistently estimate* $\delta_\Lambda = 2e_K^3 \mathbb{E}(w_{iXj} f_{iXj})$, $\xi_\Lambda = 2e_K^2 \mathbb{E}(w_{iXj}((f^2_{iX} + f^2_{Xj})/f^2_X) f_{iXj})$, $\zeta_\Lambda = 4e_K^2 \mathbb{E}(w_{iXj}((f_{iX} + f_{Xj})/f_X) f_{iXj})$ *and* $\sigma^2_\Lambda = 2v_K \mathbb{E}(w^2_{iXj} f^3_{iXj})$, *respectively.*

A test that rejects the null at level α when $\hat{\lambda}_n$ is greater or equal to the $(1-\alpha)$-quantile of a standard normal distribution thus is locally strictly unbiased. As for $\hat{\delta}_\Lambda$, $\hat{\xi}_\Lambda$, $\hat{\zeta}_\Lambda$ and $\hat{\sigma}^2_\Lambda$, we substitute the sample average for the expectation operator and plug in the kernel density estimates to obtain

$$\hat{\delta}_\Lambda = \frac{e_K^3}{n} \sum_{k=1}^n w_{iXj}(d_{k+J}, X_k, d_k) \hat{f}_{iXj}(d_{k+J}, X_k, d_k),$$

$$\hat{\xi}_\Lambda = \frac{e_K^2}{n} \sum_{k=1}^n w_{iXj}(d_{k+J}, X_k, d_k) \frac{\hat{f}^2_{iX}(d_{k+J}, X_k) + \hat{f}^2_{Xj}(X_k, d_k)}{\hat{f}^2_X(X_k)} \hat{f}_{iXj}(d_{k+J}, X_k, d_k),$$

$$\hat{\zeta}_\Lambda = \frac{e_K^2}{n} \sum_{k=1}^n w_{iXj}(d_{k+J}, X_k, d_k) \frac{\hat{f}_{iX}(d_{k+J}, X_k) + \hat{f}_{Xj}(X_k, d_k)}{\hat{f}_X(X_k)} \hat{f}_{iXj}(d_{k+J}, X_k, d_k),$$

$$\hat{\sigma}^2_\Lambda = \frac{v_K}{n} \sum_{k=1}^n w^2_{iXj}(d_{k+J}, X_k, d_k) \hat{f}^3_{iXj}(d_{k+J}, X_k, d_k).$$

To examine the local power of our testing procedure, we first define the sequence of densities $f^{[n]}_{iXj}$ and $g^{[n]}_{iXj}$ such that $\|f^{[n]}_{iXj} - f_{iXj}\| = (n^{-1} b_n^{-3/2})$ and $\|g^{[n]}_{iXj} - g_{iXj}\| = (n^{-1} b_n^{-3/2})$. As in Aït-Sahalia et al. (2001), we consider the sequence of local alternatives

$$H_1^{[n]}: \ \sup |f^{[n]}_{iXj}(a_1, x, a_2) - g^{[n]}_{iXj}(a_1, x, a_2) - \varepsilon_n \ell(a_1, x, a_2)| = \mathrm{o}(\varepsilon_n), \tag{5}$$

where $\varepsilon_n = n^{-1/2} b_n^{-3/4}$ and the function $\ell(\cdot,\cdot,\cdot)$ is such that $\ell_1 \equiv \mathbb{E}[\ell(d_i, X_j, d_j)] = 0$ and $\ell_2 \equiv \mathbb{E}[\ell^2(d_i, X_j, d_j)] < \infty$. The next result illustrates the fact that the testing procedure entails nontrivial power under local alternatives that shrink to the null at rate ε_n.

Proposition 2. *Under the sequence of local alternatives* $H_1^{[n]}$ *and Assumptions* A1–A4, *it follows that* $\hat{\lambda}_n \xrightarrow{\mathrm{d}} \mathrm{N}(\ell_2/\sigma_\Lambda, 1)$.

It is also possible to derive alternative testing procedures that rely on the restrictions imposed by the conditional independence property on the cumulative probability functions. For instance, Linton and Gozalo (1999) propose two nonparametric tests for conditional independence restrictions rooted in a generalization of the empirical distribution function. They show that, in an iid setup, the asymptotic null distribution of the test statistic is a quite complicated functional of a Gaussian process. Unfortunately, extending their results to the time-series context is not simple as opposed to the case of tests based on smoothing techniques. This is due to the fact that smoothing methods effectively use the nearest neighbors in the state space, which are unlikely to be the neighbors in the time space under Assumption A1.

3. Finite sample properties

It is well known that the asymptotic behavior of kernel-based tests is sometimes of little value in finite samples (see Fan, 1995; Fan and Linton, 2003). It is therefore natural to consider a bootstrap-version of our test that relies on a Markov resampling scheme that satisfies the null hypothesis (Horowitz, 2003). Our bootstrap-based test consists of three steps:

(S1) Draw the initial observation $X_0^{(b)}$ from the kernel-based nonparametric estimate of the stationary distribution of the bid–ask spreads and then draw the remaining artificial sample $\{d_k^{(b)}, X_k^{(b)}\}_{k=1}^m$ from the kernel estimates $\hat{F}(X_k, d_k|X_{k-1} = X_{k-1}^{(b)})$ of the conditional distribution of (d_k, X_k) given the previous realization of the bid–ask spread. This is the bootstrap sample, for which the null hypothesis in (2) holds conditional on the original sample.
(S2) Compute the test statistic $\hat{\lambda}_m^{(b)}$ as in (4) using the bootstrap sample.
(S3) Repeat the steps S1 and S2 for a large number of times, say B, and obtain the empirical distribution function of $\{\hat{\lambda}_m^{(b)}\}_{b=1}^B$.

Note that, as suggested by Bickel et al. (1997), we resample only m out of n observations so as to cope with the fact that the U-statistic implied by (3) is degenerate (see Appendix). The m/n bootstrap is consistent for $m \to \infty$ and $m/n \to 0$ as $n \to \infty$ even if the n/n bootstrap fails (Politis et al., 1999).

To evaluate the finite-sample performance of our asymptotic and bootstrap-based tests, we conduct a simple Monte Carlo study. As our empirical interest lies on testing for adverse selection costs by checking whether the bid–ask spread satisfies the Markov property, we simulate Easley and O'Hara's (1992) model with empirically plausible estimates of the model parameters. In their setup, there is a single market maker, who is risk neutral and acts competitively. Let V denote the value of the asset and define an information event as the occurrence of a signal ψ about V. The signal can take on one of two values, L and H, with probabilities $\delta > 0$ and $1 - \delta > 0$. The expected value of the asset conditional on the signal is $\mathbb{E}(V \mid \psi = L) = V_L$ or $\mathbb{E}(V \mid \psi = H) = V_H$. If no information event occurs ($\psi = 0$), then the expected value of the asset remains at $V_* = \delta V_L + (1 - \delta) V_H$.

Information events occur with probability $\alpha \in (0, 1)$ before the start of the trading day. There are two types of traders: uninformed and informed. The informed traders are risk neutral and price takers. As such, their optimal trading strategy reads: If a high (low) signal occurs, the insider buys (sells, respectively) the stock if the current quote is below V_H (above V_L). The uninformed market participants trade for nonspeculative reasons, with a fraction γ of potential sellers and a fraction $1 - \gamma$ of potential buyers. Uninformed buyers trade with probability ε_B, whereas an uninformed seller's trading probability is ε_S.

Transactions occur throughout the day along discrete intervals of time that are long enough to accommodate at most one trade. The exact length of a trading interval is arbitrary and could even approach zero so as to reformulate the statistical model in terms of Poisson arrivals. At each trading interval t, the market maker announces the bid and ask prices at which she is willing to trade one unit of the asset. Easley and O'Hara (1992) show that the spread $X_{d,t+1}$ at time $t+1$ on a particular day d is

$$\begin{aligned} X_{d,t+1} = & [\Pr(\psi = L \mid N_{d,t}, S_{d,t}+1, B_{d,t}) - \Pr(\psi = L \mid N_{d,t}, S_{d,t}, B_{d,t}+1)]V_L \\ & + [\Pr(\psi = H \mid N_{d,t}, S_{d,t}+1, B_{d,t}) - \Pr(\psi = H \mid N_{d,t}, S_{d,t}, B_{d,t}+1)]V_H \\ & + [\Pr(\psi = 0 \mid N_{d,t}, S_{d,t}+1, B_{d,t}) - \Pr(\psi = 0 \mid N_{d,t}, S_{d,t}, B_{d,t}+1)]V_*, \end{aligned}$$

where $N_{d,t}$ is the number of intervals with no trades, $S_{d,t}$ is the number of sells, and $B_{d,t}$ is the number of buys up to time t on the mth day. It is straightforward to compute the above probabilities in terms of the tree parameters $(\alpha, \delta, \mu, \gamma, \varepsilon_S, \varepsilon_B)$.

We simulate 66 days with 96 trading intervals of 5 min using the parameter estimates in Easley et al. (1997): $\alpha = \frac{3}{4}$, $\delta = \frac{1}{2}$, $\mu = \frac{1}{6}$, $\gamma = \frac{1}{3}$, and $\varepsilon_S = \varepsilon_B = \frac{1}{3}$. As for the stochastic process of the asset value, we use a simple binomial model in which the asset value today equals the asset value yesterday plus an error term, which may take values plus or minus two with equal probabilities. We fix the initial condition for the asset value process at $V_0 = 50$ and then simulate the trading outcomes for each interval $t = 1, \ldots, 96$ on each day $d = 1, \ldots, 66$ according to the tree diagram in Fig. 1. The output then includes 66 daily observations (about 3 months) of the asset value as well as 6,336 (66×96) intraday

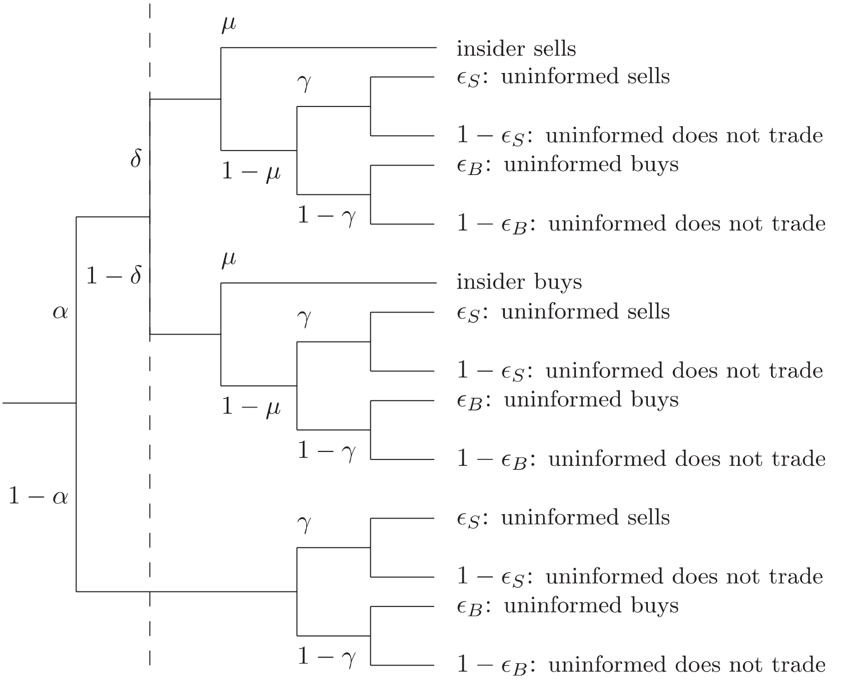

Fig. 1. Tree diagram of the trading process. Notation: α is the probability of an information event, δ is the probability of a low signal, μ is the probability a trade comes from an informed trader, γ is the probability that an uninformed trader is a seller, $1-\gamma$ is the probability that an uninformed trader is a buyer, ε_S is the probability that the uninformed trader will sell, and ε_B is the probability that the uninformed trader will buy. Nodes to the left of the dotted line occur only at the beginning of the trading day; nodes to the right occur at each trading interval.

observations of the bid–ask spread, from which we construct a sample of bid–ask spreads and their durations according to the optional sampling given by (1) with $c = \frac{1}{16}$. We consider 2,000 replications and the sample size of the resulting series of bid–ask spreads is, on average, about 3,200 observations.

Letting $j = 1$, we compute the test statistic in (4) by carrying out all density estimations using the product fourth-order kernel $\boldsymbol{K}(u_1, u_2, u_3) = K(u_1)K(u_2)K(u_3)$, where

$$K(u) = \frac{3}{\sqrt{8\pi}}\left(1 - \frac{u^2}{3}\right)\exp(-u^2/2), \tag{6}$$

for which $e_K = 27/(32\sqrt{\pi})$ and $v_K = [7881\sqrt{2}/(16384\sqrt{\pi})]^3$. We set the bandwidth in a similar fashion to Aït-Sahalia and Lo (1998), viz.

$$b_n = b_{0n} n^{-1/9} \tag{7}$$

with $b_{0n} = 2.04/(\log n)$. In the ambit of the bootstrap algorithm, we substitute $m = 1,000$ for n in (7) not only to generate $B = 399$ bootstrap samples in step S1, but also to compute the test statistics $\hat{\lambda}_m^{(b)}$ in step S2. In all instances, we focus on the compact support $\mathscr{S} = [-3/2, 3/2]^{\otimes 3}$.

The simulation results suggest that our asymptotic and bootstrap-based tests perform equally well in that both tests always reject the null hypothesis that the Markov property holds for the bid–ask spread. It rests to check whether the seemingly excellent finite-sample performance is an artifact due to size distortions of the tests. We take a similar approach to Easley et al. (1997), who test their specification against a simpler trinomial model in which the probabilities of buy, sell or no-trade are constant over time. We simulate such a trinomial model with the following constant probabilities:

$$\begin{aligned}
&\Pr(\text{buy}) = \alpha(1-\delta)\mu + (1-\gamma)\varepsilon_B[\alpha(1-\mu) + (1-\alpha)],\\
&\Pr(\text{sell}) = \alpha\delta\mu + \gamma\varepsilon_S[\alpha(1-\mu) + (1-\alpha)],\\
&\Pr(\text{no-trade}) = [\gamma(1-\varepsilon_S) + (1-\gamma)(1-\varepsilon_B)][\alpha(1-\mu) + (1-\alpha)].
\end{aligned}$$

Using the above set of parameters, the probabilities of buy and sell are both equal to $\frac{5}{24}$, whereas the probability of no-trade is $\frac{7}{12}$.

The Monte Carlo results evince that, at the 1% level, the asymptotic test never rejects the null, whereas the rejection frequency for the bootstrap-based test amounts to 0.55%. At the 5% level, the rejection frequency of the asymptotic and bootstrap-based tests increase to 0.15% and 4.05%, respectively. Altogether, we find that, although the asymptotic test exhibits a huge difference between the empirical and nominal sizes, the bootstrap version of our test has reasonable size and power properties.

4. Empirical exercise

We illustrate the above ideas using transactions data on bid and ask quotes. Information-based models of market microstructure, such as Glosten and Milgrom (1985), Easley and O'Hara (1987) and Easley and O'Hara (1992), predict that the quote-setting process depends on the whole trading history rather than exclusively on the most recent quote. This implies that the bid and ask prices, as well as the bid–ask spread, are non-Markovian. One could therefore indirectly test for the presence of asymmetric information by checking, for instance, whether the bid–ask spread satisfies the Markov property.

We focus on New York Stock Exchange (NYSE) transactions data ranging from September to November 1996. In particular, we look at five actively traded stocks: Boeing, Coke, Disney, Exxon, and IBM. See Giot (2000) and Bauwens et al. (2004) for a thorough description of the data, which originally come from the NYSE's Trade and Quote (TAQ) database. Trading on the NYSE is organized as a combined market maker/order book system. A designated specialist composes the market for each stock by managing the trading and quoting processes and providing liquidity. Apart from an opening auction, trading is continuous from 9:30 to 16:00. Table 1 reports however that the bid and ask quotes are both integrated of order one, and hence nonstationary. In contrast, there is no evidence of unit roots in the bid–ask spread processes, and hence we restrict attention to the bid–ask spread data.

We define spread durations as the time interval needed to observe a change in the bid–ask spread (i.e., $c = \frac{1}{16}$). For all stocks, we remove durations between events recorded outside the regular opening hours of the NYSE, as well as overnight spells. As documented by Giot (2000), durations feature a strong time-of-day effect related to predetermined market characteristics, such as trade opening and closing times and lunch time. To account for this feature, we also consider seasonally adjusted spread durations $\hat{d}_i = \tilde{d}_i/\phi(t_i)$, where $\tilde{d}_i$ is the original spread duration in seconds and $\phi(\cdot)$ denotes a time-of-day factor determined by averaging durations over 30-min intervals for each day of the week and

Table 1
Phillips and Perron's (1988) unit root tests

Stock	Sample size	Truncation lag	Test statistic
Boeing			
Ask	6,317	10	−1.6402
Bid	6,317	10	−1.6655
Spread	6,317	10	−115.3388
Coke			
Ask	3,823	8	−2.1555
Bid	3,823	8	−2.1615
Spread	3,823	8	−110.2846
Disney			
Ask	5,801	9	−1.2639
Bid	5,801	9	−1.2318
Spread	5,801	9	−112.1909
Exxon			
Ask	6,009	9	−0.6694
Bid	6,009	9	−0.6405
Spread	6,009	9	−121.8439
IBM			
Ask	15,124	12	−0.2177
Bid	15,124	12	−0.2124
Spread	15,124	12	−163.0558

Both ask and bid prices are in logs, whereas the spread refers to the difference of the logarithms of the ask and bid prices. The truncation lag ℓ of the Newey and West's (1987) heteroskedasticity and autocorrelation consistent estimate of the spectrum at zero frequency is based on the automatic criterion $\ell = \lfloor 4(n/100)^{2/9} \rfloor$, where $\lfloor z \rfloor$ denotes the integer part of z.

fitting a cubic spline with nodes at each half hour (see Giot, 2000). With such a transformation we aim at controlling for possible time heterogeneity of the underlying Markov process. As before, we first log-transform and normalize data to zero mean and unit variance and then estimate the density functions using the kernel in (6) with bandwidth as in (7). For $\jmath = 1$ and $\mathscr{S} = [-3/2, 3/2]^{\otimes 3}$, we compute the asymptotic and the bootstrap-based tests based on $B = 399$ artificial Markov samples of $m = 1,000$ observations.

Table 2 reports mixed test results in the sense that the Markov hypothesis seems to suit only three out of the five bid–ask spreads under consideration. We clearly reject the Markov property for the Boeing and IBM bid–ask spreads, indicating that adverse selection may play a role in the formation of their prices. In contrast, there is no indication of non-Markovian behavior in the Coke, Disney, and Exxon bid–ask spreads. Interestingly, the results are quite robust. First, they do not depend on whether the spread durations are adjusted or not for the time-of-day effect. This is surprising because the Markov property is not invariant under such a transformation and one could argue that conflicting results could cast doubts on the outcome of the analysis. Second, further analysis also shows that the results are not sensitive to small changes (up to 50%) in the bandwidths, as well.

For the sake of comparison, we also look at two alternative measures of adverse selection cost. The first proxy relates to earnings uncertainty. If information is not transparent and private information plays a role, then one would expect the distribution of the analysts' earnings-per-share estimates to display large dispersion. The I/B/E/S summary database provides some descriptive statistics on the analysts' estimates of the earnings-per-shares on November 1996. Table 3 documents the main features of the distribution of the analysts' estimates. Although the discrepancies between mean and median are quite small, their distances to the highest and lowest estimates are pretty different. This means that the distributions are not symmetric. As for the dispersion, the

Table 2
Nonparametric tests of the Markov property

Stock	Duration		Adjusted duration	
	$\hat{\lambda}_n$	p-value	$\hat{\lambda}_n$	p-value
Boeing	52.4478	(0.000) [0.0075]	55.1480	(0.000) [0.0025]
Coke	32.6271	(0.000) [0.4425]	33.2193	(0.000) [0.4275]
Disney	68.5487	(0.000) [0.0900]	71.2109	(0.000) [0.0825]
Exxon	41.8214	(0.000) [0.4300]	42.3974	(0.000) [0.3975]
IBM	88.4903	(0.000) [0.000]	94.4011	(0.000) [0.000]

Adjusted durations refer to the correction for time-of-day effects. Asymptotic p-values are in parentheses, whereas the p-values in brackets are based on 399 Markov bootstrap samples of 1,000 observations.

Table 3
Alternative measures of adverse selection costs

	Boeing	Coke	Disney	Exxon	IBM
PIN	0.0909	0.0594	0.0617	0.0784	0.1178
Earning-per-share estimates					
Number	23	26	32	35	23
Mean	1.48	1.39	0.74	1.33	2.76
Median	1.42	1.40	0.73	1.33	2.76
Maximum	1.72	1.42	0.78	1.36	2.91
Minimum	1.34	1.31	0.72	1.27	2.54
Range	0.38	0.11	0.06	0.09	0.37
Standard deviation	0.11	0.02	0.01	0.02	0.09

The descriptive statistics of the earning-per-share estimates are from the I/B/E/S summary database of November 1996, whereas PIN refers to Easley et al.'s (2002) estimates of the unconditional probability of informed trading in 1996. There are no PIN estimates for Coke and Disney in 1996, hence we report PIN estimates from 1995 for Coke and from 1997 for Disney.

standard deviation and range figures are consistent with our results. The distribution of earnings-per-share estimates for stocks with non-Markovian prices (i.e., Boeing and IBM) indeed are relatively more disperse.

However, such a dispersion does not necessarily relate to adverse selection cost. For instance, it could merely reflect genuine divergence of opinions (Diether et al., 2002). We therefore consider a more direct measure of adverse selection cost based on the parametric structural model by Easley and O'Hara (1992). More precisely, we employ Easley et al.'s (2002) maximum likelihood estimates of the probability of informed trading given by $\text{PIN} = \alpha\,\mu/(\alpha\,\mu + \varepsilon_B + \varepsilon_S)$. There are no PIN estimates for Coke and Disney in 1996, hence we report PIN estimates from 1995 for Coke and from 1997 for Disney. Table 3 shows that the PIN values also corroborate our results in that the stocks for which we cannot reject the Markov property exhibit the lowest PIN values.

5. Conclusion

This paper develops a test for Markovian dynamics that is particularly tailored to high frequency data. Although we derive the asymptotic normality of our test statistic, we also propose a bootstrap-based variant of the test so as to enhance the finite-sample properties of the testing procedure. Monte Carlo simulations show indeed that our bootstrap-based test seems to have reasonable size and power properties.

Our testing procedures are especially interesting in the context of information-based models of market microstructure. For instance, Easley and O'Hara (1992) predict that the price discovery process is such that the Markov assumption does not hold for the bid–ask spread set by the market maker. We therefore check whether the Markov hypothesis is reasonable for the bid–ask spread of five stocks actively traded on the New York Stock Exchange. The results show that the Markov assumption seems inadequate for the Boeing and IBM bid–ask spreads, indicating that the market maker may account for asymmetric information in the quote-setting process. In contrast, a Markovian character appears to suit the Coke, Disney and Exxon bid–ask spreads well, suggesting low adverse selection costs.

Acknowledgments

We are indebted to Valentina Corradi, the anonymous referees, and many seminar participants for valuable comments. We also thank Pedro Saffi for providing the I/B/E/S data, Soeren Hvidkjaer for helping us with the PIN data, as well as Luc Bauwens and Pierre Giot for granting access to their clean NYSE data. The authors gratefully acknowledge the financial support from FCT-Portugal and CNPq-Brazil, respectively. The usual disclaimer applies.

Appendix A. Proofs

In the following, we employ an index in the expectation operator to individuate the random quantity that it refers to and let its absence denote expectation over the whole sample $(Z_1, \ldots, Z_n)$, where $Z_k = (d_{k+j}, X_k, d_k)$. To simplify notation, we denote by f any density function that relates to Z_k, e.g., $f(z_i, z_j)$ corresponds to the joint density function of (Z_i, Z_j) evaluated at (z_i, z_j). We also let $\{\check{Z}_k,\ k = 1, \ldots, n\}$ denote an iid sequence with the same marginal distribution as Z_k.

Lemma 1. *Consider the functional*

$$I_n = \int \varphi(z)[\hat{f}_{iXj}(z) - f_{iXj}(z)]^2 \, \mathrm{d}z,$$

where the function φ includes an indicator function on the compact support $\mathscr{S}$, $\hat{f}_{iXj}$ is the kernel estimate of the joint density function f_{iXj} of $Z_k = (d_{k+j}, X_k, d_k)$ evaluated at $z = (a_1, x, a_2)$. Under Assumptions A1–A4,

$$nb_n^{3/2} I_n - b_n^{-3/2} e_K^3 \mathbb{E}_{Z_1}[\varphi(Z_1)] \xrightarrow{\mathrm{d}} \mathrm{N}(0, v_K \mathbb{E}_{Z_1}[\varphi^2(Z_1) f_{iXj}(Z_1)]),$$

provided that the above expectations are finite.

Proof. To simplify notation, let $f = f_{iXj}$, $\boldsymbol{K}_{b_n}(z) = b_n^{-3} \boldsymbol{K}(z/b_n)$, $r_n(z, Z_1) = \varphi(z)^{1/2} \boldsymbol{K}_{b_n}(z - Z_1)$, and $\check{r}_n(z, Z_1) = r_n(z, Z_1) - \mathbb{E}_{Z_1}[r_n(z, Z_1)]$. Consider then the following decomposition:

$$\begin{aligned} I_n = & \int \varphi(z)[\hat{f}(z) - \mathbb{E}\hat{f}(z)]^2 \, \mathrm{d}z + \int \varphi(z)[\mathbb{E}\hat{f}(z) - f(z)]^2 \, \mathrm{d}z \\ & + 2 \int \varphi(z)[\hat{f}(z) - \mathbb{E}\hat{f}(z)][\mathbb{E}\hat{f}(z) - f(z)] \, \mathrm{d}z, \end{aligned}$$

or equivalently, $I_n = I_{1n} + I_{2n} + I_{3n} + I_{4n}$, where

$$\begin{aligned} I_{1n} &= \frac{2}{n^2} \sum_{i<j} \int \check{r}_n(z, Z_i) \, \check{r}_n(z, Z_j) \, \mathrm{d}z, \\ I_{2n} &= \frac{1}{n^2} \sum_i \int \check{r}_n^2(z, Z_i) \, \mathrm{d}z, \\ I_{3n} &= \int \varphi(z)[\mathbb{E}\hat{f}(z) - f(z)]^2 \, \mathrm{d}z, \\ I_{4n} &= 2 \int \varphi(z)[\hat{f}(z) - \mathbb{E}\hat{f}(z)][\mathbb{E}\hat{f}(z) - f(z)] \, \mathrm{d}z. \end{aligned}$$

We next show that the first term is a degenerate U-statistic and contributes with the variance in the limiting distribution, while the second gives the asymptotic bias. Assumption A3 ensures, in turn, that the third and fourth terms are negligible. We start with the first moment of $r_n(z, Z_1)$:

$$\begin{aligned}\mathbb{E}_{Z_1}[r_n(z, Z_1)] &= \varphi^{1/2}(z) \int \boldsymbol{K}_{b_n}(z - z_1) f(z_1)\,\mathrm{d}z_1 \\ &= \varphi^{1/2}(z) \int \boldsymbol{K}(v) f(z - v b_n)\,\mathrm{d}v \\ &= \varphi^{1/2}(z) f(z) + \mathrm{O}(b_n^s),\end{aligned}$$

where $v = (z - z_1)/b_n$ and the last equality follows from a sth-order Taylor expansion of $f(z - v b_n)$ around z under Assumptions A2 and A3. As for the second moment, a similar argument yields

$$\begin{aligned}\mathbb{E}_{Z_1}[r_n^2(z, Z_1)] &= \varphi(z) \int \boldsymbol{K}_{b_n}^2(z - z_1) f(z_1)\,\mathrm{d}z_1 \\ &= b_n^{-3} \varphi(z) \int \boldsymbol{K}^2(v) f(z - v b_n)\,\mathrm{d}v \\ &= b_n^{-3} e_K^3 \varphi(z) f(z) + \mathrm{O}(b_n^{-1}),\end{aligned}$$

given the symmetric nature of the kernel. This means that

$$\begin{aligned}\mathbb{E}(I_{2n}) &= \frac{1}{n} \int \mathbb{E}_{Z_1}[r_n^2(z, Z_1)]\,\mathrm{d}z - \frac{1}{n} \int \mathbb{E}_{Z_1}^2[r_n(z, Z_1)]\,\mathrm{d}z \\ &= \frac{1}{n} \int [b_n^{-3} e_K^3 \varphi(z) f(z) + \mathrm{O}(b_n^{-1})]\,\mathrm{d}z + \mathrm{O}(n^{-1}) \\ &= n^{-1} b_n^{-3} e_K^3 \int \varphi(z) f(z)\,\mathrm{d}z + \mathrm{O}(n^{-1} b_n^{-1}),\end{aligned}$$

whereas

$$\begin{aligned}\mathrm{Var}(I_{2n}) &= \mathrm{Var}\left[\frac{1}{n^2} \sum_{i=1}^{n} \int \check{r}_n^2(z, Z_i)\,\mathrm{d}z\right] \\ &= \frac{1}{n^3} \mathrm{Var}\left[\int \check{r}_n^2(z, Z_1)\,\mathrm{d}z\right] + \frac{2}{n^4} \sum_{i<j} \mathrm{Cov}\left(\int \check{r}_n^2(z, Z_i)\,\mathrm{d}z, \int \check{r}_n^2(y, Z_j)\,\mathrm{d}y\right). \qquad (8)\end{aligned}$$

As for the variance term on the right-hand side of (8), it follows that

$$\begin{aligned}\mathrm{Var}\left[\int \check{r}_n^2(z, Z_1)\,\mathrm{d}z\right] &= \mathbb{E}_{Z_1}\left[\int\!\!\int \check{r}_n^2(z, Z_1) \check{r}_n^2(y, Z_1)\,\mathrm{d}z\,\mathrm{d}y\right] - \left\{\int \mathbb{E}_{Z_1}[\check{r}_n^2(z, Z_1)]\,\mathrm{d}z\right\}^2 \\ &= \mathbb{E}_{Z_1}\left[\int\!\!\int \check{r}_n^2(z, Z_1) \check{r}_n^2(y, Z_1)\,\mathrm{d}z\,\mathrm{d}y\right] + \mathrm{O}(b_n^{-6}) \\ &= \int\!\!\int \mathbb{E}_{Z_1}\{(r_n(z, Z_1) - \mathbb{E}_{Z_1}[r_n(z, Z_1)])(r_n(y, Z_1) \\ &\quad - \mathbb{E}_{Z_1}[r_n(y, Z_1)])\}^2\,\mathrm{d}z\,\mathrm{d}y + \mathrm{O}(b_n^{-6})\end{aligned}$$

$$
\begin{aligned}
&= \int\int \mathbb{E}_{Z_1}[r_n^2(z,Z_1)r_n^2(y,Z_1)]\,\mathrm{d}z\,\mathrm{d}y \\
&\quad - 3\int\int \mathbb{E}_{Z_1}^2[r_n(z,Z_1)]\mathbb{E}_{Z_1}^2[r_n(y,Z_1)]\,\mathrm{d}z\,\mathrm{d}y \\
&\quad - 4\int\int \mathbb{E}_{Z_1}[r_n^2(z,Z_1)r_n(y,Z_1)]\mathbb{E}_{Z_1}[r_n(y,Z_1)]\,\mathrm{d}z\,\mathrm{d}y \\
&\quad + 2\int\int \mathbb{E}_{Z_1}[r_n^2(z,Z_1)]\mathbb{E}_{Z_1}^2[r_n(y,Z_1)]\,\mathrm{d}z\,\mathrm{d}y \\
&\quad + 4\int\int \mathbb{E}_{Z_1}[r_n(z,Z_1)r_n(y,Z_1)]\mathbb{E}_{Z_1}[r_n(z,Z_1)]\mathbb{E}_{Z_1}[r_n(y,Z_1)]\,\mathrm{d}z\,\mathrm{d}y \\
&\quad + \mathrm{O}(b_n^{-6}).
\end{aligned}
$$

It is straightforward to show that the first term on the right-hand side of the last equality equals $b_n^{-6}\mathbb{E}_{Z_1}[\varphi^2(Z_1)]\int\int \boldsymbol{K}^2(u)\boldsymbol{K}^2(u+v)\,\mathrm{d}u\,\mathrm{d}v + \mathrm{O}(b_n^{-5})$, the second term is of order O(1), the third term is proportional to $b_n^{-3}\mathbb{E}_{Z_1}[\varphi^2(Z_1)f(Z_1)]\int\int \boldsymbol{K}^2(u)\boldsymbol{K}(u+v)\,\mathrm{d}u\,\mathrm{d}v + \mathrm{O}(b_n^{-2})$, the fourth term is of order $\mathrm{O}(b_n^{-3})$, and the fifth term equals $4\mathbb{E}_{Z_1}[\varphi^2(Z_1)f^2(Z_1)]\int\int \boldsymbol{K}(u)\boldsymbol{K}(u+v)\,\mathrm{d}u\,\mathrm{d}v + \mathrm{O}(b_n)$. Altogether this means that $\mathrm{Var}[\int \breve{r}_n^2(z,Z_1)\,\mathrm{d}z] = \mathrm{O}(b_n^{-6})$. We now determine the order of the covariance term on the right-hand side of (8) using Dette and Spreckelsen's (2004) Lemma A1. In particular, for some $\delta > 0$,

$$
\frac{2}{n^4}\sum_{i<j}\left|\mathbb{E}\left(\int\int \breve{r}_n^2(z,Z_i)\breve{r}_n^2(y,Z_j)\,\mathrm{d}z\,\mathrm{d}y\right) - \mathbb{E}\left(\int \breve{r}_n^2(z,Z_i)\,\mathrm{d}z\right)\mathbb{E}\left(\int \breve{r}_n^2(y,Z_j)\,\mathrm{d}y\right)\right|
$$

is at most of order $\mathrm{O}(n^{-4}b_n^{-6}\sum_{i<j}\beta_{j-i}^{\delta/(1+\delta)}) = \mathrm{O}(n^{-3}b_n^{-6})$ under Assumptions A1 and A2. The variance of $nb_n^{3/2}I_{2n}$ thus is of order $\mathrm{O}(n^{-1}b_n^{-3}) = \mathrm{o}(1)$ by Assumption A3, and hence

$$
nb_n^{3/2}I_{2n} - b_n^{-3/2}e_K^3\mathbb{E}_{Z_1}[\varphi(Z_1)] = \mathrm{o_p}(1)
$$

given Chebyshev's inequality. In turn, the deterministic term I_{3n} is proportional to the integrated squared bias of the fixed kernel density estimation, and hence it is of order $\mathrm{O}(b_n^{2s})$. This implies that $nb_n^{3/2}I_{3n} = \mathrm{O}(nb_n^{(3+4s)/2})$. As for the fourth term, we note that $I_{4n} = 2[I_{4n}^* - \mathbb{E}(I_{4n}^*)]$, where

$$
\begin{aligned}
I_{4n}^* &= \int \varphi(z)\hat{f}(z)[\mathbb{E}\hat{f}(z) - f(z)]\,\mathrm{d}z \\
&= n^{-1}b_n^{-3}\sum_{i=1}^n \int \varphi(z)\boldsymbol{K}\left(\frac{z-Z_i}{b_n}\right)[\mathbb{E}\hat{f}(z) - f(z)]\,\mathrm{d}z \\
&= n^{-1}b_n^{-3}\sum_{i=1}^n Y_i.
\end{aligned}
$$

As in Hall's (1984) Lemma 1, $\mathbb{E}(Y_1^r) = \mathrm{O}(b_n^{(3+s)r})$ for any $r \geqslant 1$, and hence

$$
\begin{aligned}
\mathrm{Var}(I_{4n}^*) &= 2n^{-2}b_n^{-6}\sum_{i<j}\mathrm{Cov}(Y_i, Y_j) + n^{-1}\,b_n^{-6}\,\mathrm{Var}(Y_i) \\
&\leqslant n^{-2}b_n^{-6}\times \mathrm{O}\left(b_n^{2(3+s)}\sum_{i<j}\beta_{j-i}^{\delta/(1+\delta)}\right) + n^{-1}\,b_n^{-6}\times \mathrm{O}(b_n^{2(3+s)}) = \mathrm{O}(n^{-1}b_n^{2s})
\end{aligned}
$$

by Dette and Spreckelsen's (2004) Lemma A1. Assumption A3 then ensures that $n b_n^{3/2} I_{4n} = \mathrm{o_p}(1)$ by Chebyshev's inequality. Lastly, recall that $I_{1n} = \sum_{i<j} H_n(Z_i, Z_j)$, where

$$H_n(Z_i, Z_j) = 2n^{-2} \int \check{r}_n(z, Z_i)\, \check{r}_n(z, Z_j)\, \mathrm{d}z.$$

Because $H_n(Z_i, Z_j)$ is symmetric, centered and such that $\mathbb{E}[H_n(Z_i, z_j)] = 0$ almost surely, I_{1n} is a degenerate U-statistic. Assumptions A1–A4 ensure that Fan and Li's (1999) central limit theorem for degenerate U-statistics hold and hence $nb_n^{3/2} I_{1n} \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Omega)$, where

$$\begin{aligned}
\Omega &= \frac{n^4 b_n^3}{2} \mathbb{E}[H_n^2(\check{Z}_1, \check{Z}_2)] \\
&= 2b_n^3 \int \int \left[\int \check{r}_n(z, \check{z}_1) \check{r}_n(z, \check{z}_2)\, \mathrm{d}z \right]^2 f(\check{z}_1) f(\check{z}_2)\, \mathrm{d}\check{z}_1\, \mathrm{d}\check{z}_2 \\
&= 2b_n^3 \int \int \left[\int \check{r}_n(z, \check{z}_1) \check{r}_n(y, \check{z}_1) f(\check{z}_1)\, \mathrm{d}\check{z}_1 \right]^2 \mathrm{d}z\, \mathrm{d}y \\
&= 2 \int \int \varphi(z) \varphi(z + vb_n) \left[\int \boldsymbol{K}(u) \boldsymbol{K}(u+v) f(z - ub_n)\, \mathrm{d}u \right. \\
&\qquad \left. - b_n \int \boldsymbol{K}(u) f(z - ub_n)\, \mathrm{d}u \int \boldsymbol{K}(u) f(z + vb_n - ub_n)\, \mathrm{d}u \right]^2 \mathrm{d}z\, \mathrm{d}v \\
&= 2 \int \varphi^2(z) \left[\int \boldsymbol{K}(u) \boldsymbol{K}(u+v) f(z) \right]^2 \mathrm{d}z\, \mathrm{d}v + \mathrm{O}(b_n) \\
&= 2\, v_K \mathbb{E}[\varphi^2(Z_1) f(Z_1)] + \mathrm{O}(b_n).
\end{aligned}$$

To complete the proof, it now rests to verify that Assumptions (A1)–(A3) in Fan and Li (1999) hold under Assumptions A1–A4. In the following, we employ the notation $a_n \sim b_n$ to denote that a_n and b_n are of the same order of magnitude. Letting i and j denote nonnegative integers, a change-of-variables argument then yields that

$$\begin{aligned}
\gamma_{ijk} &\equiv \mathbb{E}[H^i(Z_1, Z_2)\, H^j(Z_{k+1}, Z_{k+2})] \\
&\sim n^{-2(i+j)} \mathbb{E}\left[\left(\int r_n(z, Z_1)\, r_n(z, Z_2)\, \mathrm{d}z \right)^i \left(\int r_n(z, Z_{k+1})\, r_n(z, Z_{k+2})\, \mathrm{d}z \right)^j \right] \\
&= n^{-2(i+j)}\, b_n^{-6(i+j)} \int \cdots \int \prod_{\ell=1}^{i} \varphi(y_\ell) \boldsymbol{K}\left(\frac{y_\ell - z_1}{b_n} \right) \boldsymbol{K}\left(\frac{y_\ell - z_2}{b_n} \right) \\
&\quad \times \prod_{\ell'=i+1}^{i+j} \varphi(y_{\ell'})\, \boldsymbol{K}\left(\frac{y_{\ell'} - z_{k+1}}{b_n} \right) \boldsymbol{K}\left(\frac{y_{\ell'} - z_{k+2}}{b_n} \right) \\
&\quad \times f(z_1, z_2, z_{k+1}, z_{k+2})\, \mathrm{d}y_1 \cdots \mathrm{d}y_{i+j}\, \mathrm{d}z_1\, \mathrm{d}z_2\, \mathrm{d}z_{k+1}\, \mathrm{d}z_{k+2} \\
&= \mathrm{O}(n^{-2(i+j)} b_n^{-3(i+j-k-1)})
\end{aligned}$$

for $k = 0, 1, 2$. It is also easy to appreciate that $\tilde{\gamma}_{ijk} \equiv \mathbb{E}[H^i(\check{Z}_1, \check{Z}_2) H^j(\check{Z}_{k+1}, \check{Z}_{k+2})]$ has the same order of magnitude. Fan and Li's (1999) Assumptions (A1)–(A3) then hold if one restricts attention to small blocks of length $r_1 = \lfloor C \log n \rfloor$, where C is a sufficiently large constant and $\lfloor \cdot \rfloor$ denotes the integer part, and large blocks with length $r_2 = \lfloor n^{1/4} \rfloor$ as in

Fan and Ullah (1999). To appreciate that, it suffices to observe that, under Assumptions A1–A4,

$$\begin{aligned}
&r_1 r_2^2 \tilde{\gamma}_{400}/(n\tilde{\gamma}_{200})^2 = \mathrm{O}(n^{-3/2+3/p}\log n) = \mathrm{o}(1),\\
&r_1 \gamma_{112}/\tilde{\gamma}_{200} = \mathrm{O}(n^{-6/p}\log n) = \mathrm{o}(1),\\
&r_1 \gamma_{111}/\tilde{\gamma}_{200} = \mathrm{O}(n^{-3/p}\log n) = \mathrm{o}(1),\\
&r_1 \tilde{\gamma}_{221}/\tilde{\gamma}_{200} = \mathrm{O}(n^{-4+3/p}\log n) = \mathrm{o}(1),\\
&r_1 \tilde{\gamma}_{200}^{-1} \int \mathbb{E}[H^2(z_1, Z_2)H^2(z_1, Z_3)]f(z_1)\,\mathrm{d}z_1 = \mathrm{O}(n^{-4+3/p}\log n) = \mathrm{o}(1),\\
&r_1^5 r_2 \tilde{\gamma}_{222}/(n\tilde{\gamma}_{200})^2 = \mathrm{O}(n^{-7/4-3/p}[\log n]^5) = \mathrm{o}(1),\\
&r_1^5 r_2 \tilde{\gamma}_{132}/(n\tilde{\gamma}_{200})^2 = \mathrm{O}(n^{-7/4-3/p}[\log n]^5) = \mathrm{o}(1),
\end{aligned}$$

as required by Assumption (A1). The validity of Assumption (A2) ensues from a similar change-of-variables argument:

$$\begin{aligned}
&r_1^4 \tilde{\gamma}_{200}^{-2} \int \mathbb{E}^2[H(Z_1, z_2)H(Z_1, z_3)]f(z_2, z_3)\,\mathrm{d}z_2\,\mathrm{d}z_3 = \mathrm{O}(n^{-3/p}[\log n]^4) = \mathrm{o}(1),\\
&r_1^4 \tilde{\gamma}_{200}^{-2} |\mathbb{E}\{\mathbb{E}_{Z_1}[H^2(Z_1, Z_2)]\mathbb{E}_{Z_1}[H(Z_1, Z_3)H(Z_1, Z_4)]\}| = \mathrm{O}(n^{-3/p}[\log n]^4) = \mathrm{o}(1),\\
&r_1^4 \tilde{\gamma}_{200}^{-2} |\mathbb{E}\{\mathbb{E}_{Z_1}[H(Z_1, Z_2)H(Z_1, Z_3)]\mathbb{E}_{Z_1}[H(Z_1, Z_2)H(Z_1, Z_4)]\}| = \mathrm{O}(n^{-6/p}[\log n]^4) = \mathrm{o}(1),\\
&r_1^4 \tilde{\gamma}_{200}^{-2} |\mathbb{E}\{\mathbb{E}_{Z_1}[H(Z_1, Z_2)H(Z_1, Z_3)]\mathbb{E}_{Z_1}[H(Z_1, Z_4)H(Z_1, Z_5)]\}| = \mathrm{O}(n^{-6/p}[\log n]^4) = \mathrm{o}(1),\\
&r_1 n^{-1} \tilde{\gamma}_{200}^{-2} \mathbb{E}_{Z_1}\{\mathbb{E}_{Z_2}[H^2(Z_1, Z_2)]\} = \mathrm{O}(n^{-1}\log n) = \mathrm{o}(1).
\end{aligned}$$

Assumption (A3) holds if the constant in the definition of the small block length r_1 satisfies

$$C > \max\left\{ \left(10 - \frac{6}{p}\right) \frac{1+\delta}{\delta \log(1/\rho)}, \left(12 - \frac{6}{p}\right) \frac{1+\delta}{2\delta \log(1/\rho)} \right\}$$

in view that $0 < \rho < 1$,

$$n^2 \beta_{r_1}^{\delta/(1+\delta)} (r_1^2 + n^2 \beta_{r_1}^{\delta/(1+\delta)})/\tilde{\gamma}_{200}^2 = \mathrm{O}(n^{10+C(\delta\log\rho/(1+\delta))-6/p} + n^{12+2C(\delta\log\rho/(1+\delta))-6/p}),$$

and that M_n as defined by Fan and Li (1999) is of order o(1) under Assumptions A1–A4. □

Proof of Proposition 1. Consider the second-order functional Taylor expansion

$$\Lambda_{f+h} = \Lambda_f + \mathrm{D}\Lambda_f(h) + \tfrac{1}{2}\mathrm{D}^2\Lambda_f(h, h) + \mathrm{O}(\|h\|^3),$$

where h denotes the perturbation $h_{iXj} = \hat{f}_{iXj} - f_{iXj}$. Under the null hypothesis that $f_{iXj} = g_{iXj}$, both Λ_f and $\mathrm{D}\Lambda_f$ equal zero. To appreciate the singularity of the latter, it suffices to compute the Gâteaux derivative of $\Lambda_{f,h}(\lambda) = \Lambda_{f+\lambda h}$ with respect to λ evaluated at $\lambda \to {}^{+}0$. Let

$$g_{iXj}(\lambda) = \frac{\int [f_{iXj} + \lambda h_{iXj}](a_1, x, a_2)\,\mathrm{d}a_2 \int [f_{iXj} + \lambda h_{iXj}](a_1, x, a_2)\,\mathrm{d}a_1}{\int [f_{iXj} + \lambda h_{iXj}](a_1, x, a_2)\,\mathrm{d}(a_1, a_2)}.$$

It then follows that

$$\frac{\partial \Lambda_{f,h}(0)}{\partial \lambda} = 2\int w_{iXj}[f_{iXj} - g_{iXj}][h_{iXj} - \mathrm{D}g_{iXj}]f_{iXj}(a_1, x, a_2)\,\mathrm{d}(a_1, x, a_2) + \int w_{iXj}[f_{iXj} - g_{iXj}]^2 h_{iXj}(a_1, x, a_2)\,\mathrm{d}(a_1, x, a_2),$$

where $\mathrm{D}g_{iXj}$ is the functional derivative of g_{iXj} with respect to f_{iXj}, namely

$$\mathrm{D}g_{iXj} = \left(\frac{h_{iX}}{f_{iX}} + \frac{h_{Xj}}{f_{Xj}} - \frac{h_X}{f_X}\right) g_{iXj},$$

where $h_{iX} = \hat{f}_{iX} - f_{iX}$, $h_{Xj} = \hat{f}_{Xj} - f_{Xj}$, and $h_X = \hat{f}_X - f_X$. As is apparent, imposing the null hypothesis induces singularity in the first functional derivative $\mathrm{D}\Lambda_f$. To complete the proof, it suffices to derive the second-order derivative and bound the remainder term. As for the latter, Gyorfi et al. (1989) show that

$$\|\hat{f}_{iXj} - f_{iXj}\|_\infty = \mathrm{O}(b_n^s + n^{-1/2} b_n^{-3/2} \log n)$$

in the mixing context (see also Aït-Sahalia et al., 2001, p. 409). The bandwidth condition in Assumption A3 then ensures that

$$n b_n^{3/2} \|\hat{f}_{iXj} - f_{iXj}\|_\infty^3 = \mathrm{O}(n b_n^{3s+3/2} + n^{-1/2} b_n^{-3} [\log n]^3) = \mathrm{o}(1).$$

It then suffices to appreciate that, under the null, the second-order derivative reads

$$\mathrm{D}^2 \Lambda_f(h, h) = 2\int w_{iXj}\,[h_{iXj}(a_1, x, a_2) - \mathrm{D}g_{iXj}(a_1, x, a_2)]^2\,\mathrm{d}F_{iXj}(a_1, x, a_2)$$

given that all other terms will depend on $f_{iXj} - g_{iXj}$. In view that $\mathrm{D}g_{iXj}$ converges at a faster rate than h_{iXj} due to its lower dimensionality, it contributes only to the asymptotic bias term.[1] In a similar spirit to the proof of Lemma 1, it is straightforward to show that

$$\begin{aligned}
J_{1n} &\equiv \int w_{iXj} h_{iX}^2(a_1, x) \frac{f_{Xj}^2(x, a_2)}{f_X^2(x)}\,\mathrm{d}F_{iXj}(a_1, x, a_2) \\
&= n^{-1} b_n^{-2} e_K^2 \int w_{iXj} \frac{f_{Xj}^2(x, a_2) f_{iXj}(a_1, x, a_2)}{f_X^2(x)}\,\mathrm{d}F_{iXj}(a_1, x, a_2) + \mathrm{O_p}(n^{-1} b_n^{-1/2}), \\
J_{2n} &\equiv \int w_{iXj} h_{Xj}^2(x, a_2) \frac{f_{iX}^2(a_1, x)}{f_X^2(x)}\,\mathrm{d}F_{iXj}(a_1, x, a_2) \\
&= n^{-1} b_n^{-2} e_K^2 \int w_{iXj} \frac{f_{iX}^2(a_1, x) f_{iXj}(a_1, x, a_2)}{f_X^2(x)}\,\mathrm{d}F_{iXj}(a_1, x, a_2) + \mathrm{O_p}(n^{-1} b_n^{-1/2}), \\
J_{3n} &\equiv \int w_{iXj} h_X^2(x) \frac{f_{iX}^2(a_1, x) f_{Xj}^2(x, a_2)}{f_X^4(x)}\,\mathrm{d}F_{iXj}(a_1, x, a_2) \\
&= n^{-1} b_n^{-1} e_K \int w_{iXj} \frac{f_{iXj}^3(a_1, x, a_2)}{f_X^2(x)}\,\mathrm{d}F_{iXj}(a_1, x, a_2) + \mathrm{O_p}(n^{-1}),
\end{aligned}$$

[1] We thank an anonymous referee and Valentina Corradi for calling our attention to that. See Corradi et al. (2006) for a similar discussion.

which implies that

$$n\,b_n^{3/2}\,J_n \equiv nb_n^{3/2}\int w_{iXj}[\mathrm{D}g_{iXj}(a_1,x,a_2)]^2\,\mathrm{d}F_{iXj}(a_1,x,a_2)$$
$$= b_n^{-1/2}e_K^2\int w_{iXj}\,\frac{f_{iX}^2(a_1,x)+f_{Xj}^2(x,a_2)}{f_X^2(x)}\,f_{iXj}(a_1,x,a_2)\,\mathrm{d}F_{iXj}(a_1,x,a_2)+\mathrm{o_p}(1).$$

As for the cross-product term of $\mathrm{D}^2\Lambda_f(h,h)$, we must derive the asymptotic behavior of

$$I_{1n}^* \equiv \int \varphi(a_1,x,a_2)[\hat{f}_{iXj}(a_1,x,a_2)-f_{iXj}(a_1,x,a_2)][\hat{f}_{iX}(a_1,x)-f_{iX}(a_1,x)]\,\mathrm{d}(a_1,x,a_2),$$
$$I_{2n}^* \equiv \int \varphi(a_1,x,a_2)[\hat{f}_{iXj}(a_1,x,a_2)-f_{iXj}(a_1,x,a_2)][\hat{f}_{Xj}(x,a_2)-f_{Xj}(x,a_2)]\,\mathrm{d}(a_1,x,a_2),$$
$$I_{3n}^* \equiv \int \varphi(a_1,x,a_2)[\hat{f}_{iXj}(a_1,x,a_2)-f_{iXj}(a_1,x,a_2)][\hat{f}_{X}(x)-f_{X}(x)]\,\mathrm{d}(a_1,x,a_2),$$

where φ is a generic function of the weighting scheme w_{iXj} and of the above density functions. Following the same steps as in the proof of Lemma 1, it is easy to show that the first two terms are asymptotically equivalent in that $nb_n^{3/2}I_{1n}^* = nb_n^{3/2}I_{2n}^* = b_n^{-1/2}e_K^2\,\mathbb{E}[\varphi(Z_1)]+\mathrm{o_p}(1)$, whereas the third term is negligible given that $nb_n^{3/2}I_{3n}^* = b_n^{1/2}e_K^2\,\mathbb{E}[\varphi(Z_1)]+\mathrm{o_p}(b_n^{1/2})=\mathrm{o_p}(1)$. This thus means that

$$nb_n^{3/2}\,I_n^* \equiv nb_n^{3/2}\int w_{iXj}(z)\,h_{iXj}(z)\mathrm{D}g_{iXj}(z)\,\mathrm{d}F_{iXj}(z)$$
$$= b_n^{-1/2}e_K^2\int w_{iXj}\,\frac{f_{iX}(a_1,x)+f_{Xj}(x,a_2)}{f_X(x)}\,f_{iXj}(a_1,x,a_2)\,\mathrm{d}F_{iXj}(a_1,x,a_2)+\mathrm{o_p}(1).$$

It now only remains to deal with $\int w_{iXj}(a_1,x,a_2)h_{iXj}^2(a_1,x,a_2)\,\mathrm{d}F_{iXj}(a_1,x,a_2)$. The result then follows from a direct application of Lemma 1 with $\varphi(a_1,x,a_2)=w_{iXj}(a_1,x,a_2)\,f_{iXj}(a_1,x,a_2)$. □

Proof of Proposition 2. The conditions imposed are such that the second-order functional Taylor expansion is also valid in the double array case $(d_{i,n},X_{i,n},d_{j,n})$. Thus, under $H_1^{[n]}$ and Assumptions A1–A4,

$$\hat{\lambda}_n-\frac{b_n^{1/2}}{\hat{\sigma}_\Lambda}\sum_{k=1}^{n-m} w_{iXj}[f_{iXj}(d_{k+j,n},X_{k+j,n},d_{k,n})-g_{iXj}(d_{k+j,n},X_{k+j,n},d_{k,n})]^2$$

converges weakly to a standard normal distribution under $f^{[n]}$. The result then follows by noting that $\hat{\sigma}_\Lambda \xrightarrow{p^{[n]}} \sigma_\Lambda$ and

$$\Lambda_{f^{[n]}} = \mathbb{E}\{w_{iXj}(d_{i,n},X_{j,n},d_{j,n})[f^{[n]}(d_{i,n},X_{j,n},d_{j,n})-g^{[n]}(d_{i,n},X_{j,n},d_{j,n})]^2\}+\mathrm{O_p}(n^{-1/2})$$
$$= n^{-1}b_n^{-3/2}\ell_2+\mathrm{o_p}(n^{-1}b_n^{-3/2}).\qquad \square$$

References

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

Aït-Sahalia, Y., 2000. Testing the Markov hypothesis in discretely-sampled continuous-time processes. Princeton University.

Aït-Sahalia, Y., 2002. Telling from discrete data whether the underlying continuous-time model is a diffusion. Journal of Finance 57, 2075–2112.

Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. Journal of Finance 53, 499–547.

Aït-Sahalia, Y., Bickel, P.J., Stoker, T.M., 2001. Goodness-of-fit tests for kernel regression with an application to option implied volatilities. Journal of Econometrics 105, 363–412.

Bauwens, L., Giot, P., Grammig, J., Veredas, D., 2004. A comparison of financial duration models via density forecasts. International Journal of Forecasting 20, 589–609.

Bickel, P.J., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. Annals of Statistics 1, 1071–1095.

Bickel, P.J., Götze, F., van Zwet, W.R., 1997. Resampling fewer than n observations: gains, losses, and remedies for losses. Statistica Sinica 7, 1–31.

Burgayran, E., Darolles, S., 1997. Nonparametric estimation of a diffusion equation from tick observations. CREST Working Paper 9756, INSEE.

Chen, X., Linton, O., Robinson, P.M., 2001. The estimation of conditional densities. In: Puri, M.L. (Ed.), Asymptotics in Statistics and Probability. VSP International Science Publishers, Leiden, pp. 71–84.

Corradi, V., Distaso, W., Fernandes, M., 2006. International market links and realized volatility transmission. Queen Mary, University of London.

Dette, H., Spreckelsen, I., 2004. Some comments on specification tests in nonparametric absolutely regular processes. Journal of Time Series Analysis 25, 159–172.

Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. Journal of Finance 57, 2113–2141.

Easley, D., O'Hara, M., 1987. Price, trade size, and information in security markets. Journal of Financial Economics 19, 69–90.

Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. Journal of Finance 47, 577–605.

Easley, D., Kiefer, N.M., O'Hara, M., 1997. One day in the life of a very common stock. Review of Financial Studies 10, 805–835.

Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns? Journal of Finance 57, 2185–2221.

Fan, Y., 1994. Testing the goodness-of-fit of a parametric density function by kernel method. Econometric Theory 10, 316–356.

Fan, Y., 1995. Bootstrapping a consistent nonparametric goodness-of-fit test. Econometric Reviews 14, 367–382.

Fan, Y., Li, Q., 1999. Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification tests. Journal of Nonparametric Statistics 10, 245–271.

Fan, Y., Linton, O., 2003. Some higher-order theory for a consistent non-parametric model specification test. Journal of Statistical Planning and Inference 109, 125–154.

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

Fernandes, M., Flôres, R.G., 2004. Nonparametric tests for the Markov property. Getulio Vargas Foundation.

Fernandes, M., Grammig, J., 2005. Nonparametric specification tests for conditional duration models. Journal of Econometrics 127, 35–68.

Gao, J., King, M., 2004. Model specification testing in nonparametric and semiparametric time series econometrics. Research Report 2004/02, School of Mathematics and Statistics, University of Western Australia.

Giot, P., 2000. Time transformations, intraday data and volatility models. Journal of Computational Finance 4, 31–62.

Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.

Gyorfi, L., Härdle, W., Sarda, P., Vieu, P., 1989. Nonparametric Curve Estimation from Time Series. Lecture Notes in Statistics. Springer, Berlin.

Hall, P., 1984. Central limit theorem for integrated squared error multivariate nonparametric density estimators. Journal of Multivariate Analysis 14, 1–16.
Härdle, W., Mammen, E., 1993. Comparing nonparametric vs. parametric regression fits. Annals of Statistics 21, 1926–1947.
Hasbrouck, J., 1991. Measuring the information content of stock trades. Journal of Finance 46, 179–207.
Hoeffding, W., 1948. A non-parametric test of independence. Annals of Mathematical Statistics 19, 546–557.
Hong, Y., White, H., 2004. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. Econometrica 73, 837–901.
Horowitz, J.L., 2003. Bootstrap methods for Markov processes. Econometrica 71, 1049–1082.
Linton, O., Gozalo, P., 1999. Conditional independence restrictions: testing and estimation. London School of Economics and Brown University.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.
Pinkse, J., 1998. A consistent nonparametric test for serial independence. Journal of Econometrics 84, 205–231.
Pinkse, J., 1999. Nonparametric misspecification testing. University of British Columbia.
Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer, New York.
Robinson, P.M., 1991. Consistent nonparametric entropy-based testing. Review of Economic Studies 58, 437–453.
Rosenblatt, M., 1975. A quadratic measure of deviation of two-dimensional density estimates and a test of independence. Annals of Statistics 3, 1–14.
Skaug, H.J., Tjøstheim, D., 1993. Nonparametric tests of serial independence. In: Rao, T.S. (Ed.), Developments in Time Series Analysis. Chapman & Hall, London.
Su, L., White, H., 2002. Testing conditional independence via empirical likelihood. University of California, San Diego.
Su, L., White, H., 2003a. A consistent characteristic-function-based test for conditional independence. University of California, San Diego.
Su, L., White, H., 2003b. A nonparametric Hellinger metric test for conditional independence. University of California, San Diego.