TESTING THE MARKOV PROPERTY WITH ULTRA-HIGH FREQUENCY FINANCIAL DATA*

JOÃO AMARO DE MATOS

Faculdade de Economia

Universidade Nova de Lisboa

Rua Marquês de Fronteira, 20

1099-038 Lisbon, Portugal

Tel: +351 21 3826100

Fax: +351 21 3873973

amatos@fe.unl.pt

MARCELO FERNANDES

Economics Department

Queen Mary, University of London

Mile End Road

E1 4NS London, UK

Tel: +44 20 7882 5082

Fax: +44 20 89833580

m.fernandes@qmul.ac.uk

1

TESTING THE MARKOV PROPERTY WITH ULTRA HIGH FREQUENCY FINANCIAL DATA

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 not only the irregular spacing of transactions data, but also price discreteness. Under such an observation rule, the current price duration is independent of previous price durations given the current price realization. A simple nonparametric test then follows by examining whether this conditional independence property holds. Finally, we investigate whether or not 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. The results are mixed in the sense that the Markov assumption is rejected for three out of the five stocks we have analyzed.

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 (1999). To build a nonparametric testing procedure, the first uses the fact that the Chapman-Kolmogorov equation must hold in order for a Markov process compatible with the data to exist. If, on the one hand, the Chapman-Kolmogorov representation involves a quite complicated nonlinear functional relationship among transition probabilities of the process, on the other hand, 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 (1999) 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 consistent test for the Markov property that is suitable to ultra-high frequency data, we build on the theory of Markov processes with stochastic time changes. We assume that there is an underlying continuous-time Markov process that

3

is observed only when it crosses some discrete level. Accordingly, we accommodate not only the irregular spacing of transaction data, but also price discreteness. Further, such an optional sampling scheme implies that consecutive spells between price changes are conditionally independent given the current price realization. This paper then develops a simple nonparametric test for the Markov property by testing whether this conditional independence property holds.

There is an extensive literature on how to test either unconditional independence (e.g., Hoeffding, 1948; Rosemblatt, 1975; Pinkse, 1999) or serial independence (e.g., Robinson, 1991; Skaug and Tjøstheim, 1993; Pinkse, 1998). However, there are only a few works 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 so as to consider weakly dependent stochastic processes as well as different metrics. In particular, Su and White (2003a,b) respectively check restrictions on the characteristic function and on the empirical likelihoods, whereas Su and White (2002) verify whether the density restriction implied by conditional independence hold using the Hellinger distance. Our setting can be seen as combining the setups that Linton and Gozalo (1999) and Su and White (2002) consider. As in Su and White (2002), we derive tests under mixing conditions so as to deal with the time-series dependence associated with the Markov property. 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 microstructure 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): Boeing, Coca-Cola, Disney, Exxon, and IBM.

The results reveal that the Markov assumption is consistent with the Disney and Exxon bid-ask spreads, whereas the converse is true for Boeing, Coca-Cola and IBM. This indicates that the latter stocks presumably have higher rates of return for, in equilibrium, uninformed traders require compensation to hold stocks with greater private information (Easley, Hvidkjaer and O'Hara, 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 results thus imply that the standard asset-pricing framework is not suitable to examine the Boeing, Coca-Cola and IBM returns, though it may work for 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 huge size distortions, a bootstrap variant of the test 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. Testing Markov processes with stochastic time changes

Let t_i $(i = 1, 2, \ldots)$ denote the observation times of the continuous-time price process $\{X_t, \ t > 0\}$ and assume that $t_0 = 0$. Suppose that the shadow price $\{X_t, \ t > 0\}$ follows a strong stationary Markov process. To account for price discreteness, we assume that prices are observed only when the cumulative change in the shadow price is at least c, say a basic tick. The price duration then reads

$$d_{i+1} \equiv t_{i+1} - t_i = \inf_{\tau > 0} \{|X_{t_i + \tau} - X_{t_i}| \geq c\} \tag{1}$$

for $i = 0, \ldots, n-1$. The data available for statistical inference are the price durations $(d_1, \ldots, d_n)$ and the corresponding realizations $(X_1, \ldots, X_n)$, where $X_i = X_{t_i}$.

The observation times $\{t_i, \ i = 1, 2, \ldots\}$ form a sequence of increasing stopping times of the continuous-time Markov process $\{X_t, \ t > 0\}$, hence the discrete-time price process $\{X_i, \ i = 1, 2, \ldots\}$ satisfies the Markov property as well. Further, the price duration d_{i+1} is a measurable function of the path of $\{X_t, \ 0 < t_i \leq t \leq t_{i+1}\}$, and thus depends on the information available at time t_i only through X_i (Burgayran and Darolles, 1997). In other words, the sequence of price durations are conditionally independent given the observed price (Dawid, 1979). Therefore, one can test the Markov assumption by checking the property of conditional independence between consecutive durations given the current price realization.

Assume the existence of the joint density $f_{iXj}(\cdot, \cdot, \cdot)$ of (d_i, X_i, d_j), and let $f_{i|X}(\cdot)$ and $f_{Xj}(\cdot, \cdot)$ denote the conditional density of d_i given X_i and the joint density of (X_i, d_j), respectively. The null hypothesis of conditional independence implied by the Markov character of the price process then reads

$H_0^*: \quad f_{iXj}(a_1, x, a_2) = f_{i|X}(a_1) f_{Xj}(x, a_2)$ 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). Thus, the resulting null hypothesis is the necessary condition

$$H_0: \quad f_{iXj}(a_1, x, a_2) = f_{i|X}(a_1) f_{Xj}(x, a_2) \quad \text{a.e. for a fixed } j. \tag{2}$$

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 the mean squared difference,[1] yielding the following test statistic

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

The sample analog then is

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

where $\hat{g}_{iXj}(d_{k+i-j}, X_{k+i-j}, d_k) = \hat{f}_{i|X}(d_{k+i-j}|X_{k+i-j}) \hat{f}_{Xj}(X_{k+i-j}, d_k)$.

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 slower than $\hat{f}_{i|X}$ and $\hat{f}_{Xj}$ due to its higher dimensionality. As such, estimating the conditional density $f_{i|X}$ and the joint density f_{Xj} does not play a role in the asymptotic behavior of the test statistic.

To derive the asymptotic theory, we impose the following regularity conditions as in Aït-Sahalia (1994), Fan and Li (1999), and Fan and Ullah (1999).

ASSUMPTION A1: The sequence $\{d_i, X_i, d_j\}$ is strictly stationary and β-mixing with $\beta_\tau = O(\rho^\tau)$, where $0 < \rho < 1$.

---

[1] One could also consider other distance measures such as the integrated squared difference (Rosemblatt, 1975), the Kullback-Leibler contrast (Robinson, 1991), and the Hellinger metric (Su and White, 2002).

ASSUMPTION A2: The density function f_{iXj} is continuously differentiable up to order $s + 1$ and its derivatives are bounded and square integrable. In addition, the marginal density f_X is bounded away from zero and the following Lipschitz-type conditions hold: $|f_{iXj}(a_1 + \Delta_1, x + \Delta_2, a_2 + \Delta_3) - f_{iXj}(a_1, x, a_2)| \leq D(a_1, x, a_2) \|\Delta\|$, where $D(\cdot, \cdot, \cdot)$ is integrable.

ASSUMPTION A3: Let $e_K \equiv \int |K(u)|^2 \, du$ and $v_K \equiv \int \left[\int K(u)K(u+v) \, du\right]^2 \, dv$, where the kernel K is of order s (even integer) and is continuously differentiable up to order s on $\mathbb{R}^3$ with derivatives in $L^2(\mathbb{R}^3)$.

ASSUMPTION A4: The bandwidths $b_{d,n}$ and $b_{x,n}$ are of order $o\left(n^{-1/(2s+3)}\right)$ as the sample size n grows.

Assumption A1 restricts the amount of dependence allowed in the observed data sequence to ensure that the central limit theorem holds. It requires that the stochastic process is absolutely regular with geometric decay rate (see, e.g., Meitz and Saikkonen, 2004). Assumption A2 requires that the joint density function f_{iXj} is smooth enough to admit a functional Taylor expansion, and that the conditional density $f_{i|X}$ is everywhere well defined. Although assumption A3 provides enough room for higher order kernels, hereinafter, we implicitly assume that the kernel is of second order ($s = 2$). Assumption A4 restricts the rate at which the bandwidth must converge to zero. In particular, it induces a slight degree of undersmoothing in the density estimation, since the optimal bandwidth is of order $O\left(n^{-1/(2s+3)}\right)$. Other limiting conditions on the bandwidth are also applicable, but they would result in different terms for the bias as in Härdle and Mammen (1993).

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

PROPOSITION 1: *Under the null and assumptions A1 to A4, the statistic*

$$\hat{\lambda}_n = \frac{n \, b_n^{1/2} \Lambda_{\hat{f}} - b_n^{-1/2} \hat{\delta}_\Lambda}{\hat{\sigma}_\Lambda} \xrightarrow{d} N(0,1), \tag{4}$$

where $b_n = b_{d,n}^2\, b_{x,n}$ is the bandwidth for the kernel estimation of the joint density f_{iXj}, and $\hat{\delta}_\Lambda$ and $\hat{\sigma}_\Lambda^2$ are consistent estimates of $\delta_\Lambda = e_K\, E(f_{iXj})$ and $\sigma_\Lambda^2 = v_K\, E(f_{iXj}^3)$, respectively.

Thus, a test that rejects the null hypothesis at level α when $\hat{\lambda}_n$ is greater or equal to the $(1-\alpha)$-quantile $z_{1-\alpha}$ of a standard normal distribution is locally strictly unbiased.

To examine the local power of our testing procedure, we first define the sequence of densities $f_{iXj}^{[n]}$ and $g_{iXj}^{[n]}$ such that $\left\| f_{iXj}^{[n]} - f_{iXj} \right\| = \left(n^{-1} b_n^{-1/2} \right)$ and $\left\| g_{iXj}^{[n]} - g_{iXj} \right\| = \left(n^{-1} b_n^{-1/2} \right)$. We then consider the sequence of local alternatives

$$H_1^{[n]}: \quad \sup \left| f_{iXj}^{[n]}(a_1, x, a_2) - g_{iXj}^{[n]}(a_1, x, a_2) - \epsilon_n \ell(a_1, x, a_2) \right| = o(\epsilon_n), \tag{5}$$

where $\epsilon_n = n^{-1/2} b_n^{-1/4}$ and the function $\ell(\cdot, \cdot, \cdot)$ is such that $\ell_1 \equiv E[\ell(a_1, x, a_2)] = 0$ and $\ell_2 \equiv E[\ell^2(a_1, x, a_2)] < \infty$. 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 to A4, $\hat{\lambda}_n \xrightarrow{d} N\left(\ell_2/\sigma_\Lambda, 1 \right)$.*

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 the mixing condition in Assumption A1.

9

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 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). More precisely, 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 $\left\{ d_j^{(b)}, X_j^{(b)} \right\}_{j=1}^{m}$ from the kernel estimates of the conditional distribution $F\left(X_j, d_j \,\middle|\, X_{j-1} = X_{j-1}^{(b)} \right)$ of the random vector (d_j, X_j) 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 $T_m^{(b)}$ as in (4) using the bootstrap sample rather than the original data.

S3   Repeat the steps S1 and S2 for a large number of time, say B, and obtain the empirical distribution function of $\left\{ T_m^{(b)} \right\}_{b=1}^{B}$.

Note that, as suggested by Bickel, Götze and van Zwet (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).

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 for the parameters in the model. 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 $E(V \mid \psi = L) = V_L$ or $E(V \mid \psi = H) = V_H$. 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 current 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

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

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 $(\alpha, \delta, \mu, \gamma, \epsilon_S, \epsilon_B)$.

We simulate 66 days with 96 trading intervals of 5 minutes using the parameter estimates in Easley, Kiefer and O'Hara (1997): $\alpha = 3/4$, $\delta = 1/2$, $\mu = 1/6$, $\gamma = 1/3$, and $\epsilon_S = \epsilon_B = 1/3$. 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 Figure 1. The output then includes 66 daily observations (about 3 months) of the asset value as well as 6,336 (66×96) intraday 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 $c = 1/16$. We consider 10,000 replications and the sample size of the resulting series of bid-ask spreads is, on average, about 3,200 observations.

To compute the test statistic in (4), we carry out all density estimations using the product Gaussian kernel, namely

$$K(u) = (2\pi)^{-3/2} \exp\left(-\frac{u_1^2 + u_2^2 + u_3^2}{2} \right), \tag{6}$$

which implies that $e_K = (4\pi)^{-3/2}$ and $v_K = (8\pi)^{-3/2}$. As for the bandwidths, we adjust Silverman's (1986) rule of thumb so as to conform to the degree of undersmoothing required by Assumption A4. More precisely, we set

$$b_{u,n} = \frac{\hat{\sigma}_u}{\log(n)} (7n/4)^{-1/7}, \qquad u = d, x \tag{7}$$

where $\hat{\sigma}_d$ and $\hat{\sigma}_x$ denote the standard errors of the spread duration d_i and bid-ask spread X_i data, respectively. As for the bootstrap-based tests, we compute the test statistics using the product Gaussian kernel and the bandwidths as in (7) for $B = 499$ bootstrap samples of size $m = 1,000$.

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 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:

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

Using the above set of parameters, the probabilities of buy and sell are both equal to 5/24 and the probability of no-trade is 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 about 0.2%. At the 5% level, the rejection frequency of the asymptotic and bootstrap-based tests increase to 0.4% and 4.2%, 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 properties.

4. EMPIRICAL EXERCISE

We illustrate the above ideas using transactions data on bid and ask quotes. The motivation for such an exercise is natural. Information-based models of market microstructure, such as Glosten and Milgrom (1985) and Easley and O'Hara (1987, 1992), predict that the quote-setting process depends on the whole trading history rather than exclusively on the most recent quote, and thus both bid and ask prices, as well as the bid-ask spread,

are non-Markovian. One can 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 from the Dow Jones index: Boeing, Coca-Cola, Disney, Exxon, and IBM.[2] 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. As kernel density estimation relies on the assumption of stationarity (see Assumption A1), spread data are therefore more convenient to serve as input for the subsequent analysis.

Spread durations are defined as the time interval needed to observe a change in the bid-ask spread (i.e., $c = 1/16$). For all stocks, durations between events recorded outside the regular opening hours of the NYSE, as well as overnight spells, are removed. 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 for traders. To account for this feature, we also consider seasonally adjusted spread durations $d_i^* = d_i/\phi(t_i)$, where d_i is the original spread duration in seconds and $\phi(\cdot)$ denotes a time-of-day factor determined by averaging durations over thirty-minutes intervals for each day of the week and fitting a cubic spline with nodes at each half hour. With such a transformation we aim at controlling for possible time heterogeneity of the underlying Markov process. As before, we estimate all density functions using the product Gaussian

kernel and bandwidths as in (7) and compute the bootstrap-based tests using $B = 499$ artificial Markov samples.

Table 2 reports mixed results in the sense that the Markov hypothesis seems to suit only some of the bid-ask spreads under consideration. We clearly reject the Markov property for the Boeing, Coca-Cola 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 Disney and Exxon bid-ask spreads. Interestingly, the results are quite robust in the sense that they do not depend on whether the spread durations are adjusted or not for the time-of-day effect.[3] 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. Further, it is also comforting that there is no palpable difference between the asymptotic and bootstrap-based test results.

5. CONCLUSION

This paper develops a test for Markovian dynamics that is particularly tailored to ultra-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

---

[3] Further analyses show that the results are not very sensitive to reasonable changes in the bandwidths, as well.

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, Coca-Cola 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 Disney and Exxon bid-ask spreads well, suggesting low adverse selection costs. We thus conclude that market microstructure models rooted in Markov processes, such as in Amaro de Matos and Rosário (2002), may deserve more attention.

APPENDIX: PROOFS

LEMMA 1: Consider the functional

$$I_n = \int \varphi(a_1, x, a_2) \left[\hat{f}(a_1, x, a_2) - f(a_1, x, a_2) \right]^2 \, \mathrm{d}(a_1, x, a_2).$$

Under assumptions A1 to A4,

$$n \, b_n^{1/2} I_n - b_n^{-1/2} \, e_K \, E\left[\varphi(a_1, x, a_2)\right] \xrightarrow{d} N\left(0, v_K \, E\left[\varphi^2(a_1, x, a_2) f(a_1, x, a_2)\right]\right),$$

provided that the above expectations are finite.

PROOF: Let $z = (a_1, x, a_2)$, $K_{b_n}(z) = b_n^{-1} K(z/b_n)$, $r_n(z, Z) = \varphi(z)^{1/2} K_{b_n}(z - Z)$, and $\breve{r}_n(z, Z) = r_n(z, Z) - E_Z[r_n(z, Z)]$. Consider then the following decomposition

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

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

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

We show in the sequel 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. In turn, assumption A4 ensures that the third and fourth terms are negligible. To begin, observe that the first moment of $r_n(z, Z)$ reads

$$
\begin{aligned}
E_Z[r_n(z, Z)] &= \varphi^{1/2}(z) \int K_{b_n}(z - Z) f(Z) \, \mathrm{d}Z \\
&= \varphi^{1/2}(z) \int K(u) f(z + u b_n) \, \mathrm{d}u \\
&= \varphi^{1/2}(z) \int K(z) \left[f(z) + \frac{1}{2} f'(z) u b_n + f''(z^*) u^2 b_n^2 \right] \mathrm{d}u \\
&= \varphi^{1/2}(z) f(z) + O\left(b_n^2\right),
\end{aligned}
$$

where $f^{(i)}(\cdot)$ denotes the ith derivative of $f(\cdot)$ and $z^* \in [z, z + u b_n]$. Applying similar algebra to the second moment yields $E_Z[r_n^2(z, Z)] = b_n^{-1} e_K \varphi(z) f(z) + O(1)$. This means that

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

whereas $\mathrm{Var}(I_{2n}) = O\left(n^{-3} b_n^{-2}\right)$. It follows from Chebyshev's inequality that $n \, b_n^{1/2} I_{2n} - b_n^{-1/2} e_K E[\varphi(z)] = o_p(1)$. In turn, the deterministic term I_{3n} is proportional to the integrated squared bias of the fixed kernel density estimation, hence it is of order $O\left(b_n^4\right)$. Assumption A4 then implies that $n \, b_n^{1/2} I_{3n} = o(1)$. Further,

$$
E(I_{4n}) = 2 \int \varphi(z) E_Z \left[\hat{f}(z) - E \hat{f}(z) \right] \left[E \hat{f}(z) - f(z) \right] \mathrm{d}z = 0,
$$

whereas $E(I_{4n}^2) = O\left(n^{-1} b_n^4\right)$ as in Hall (1984, Lemma 1). It then suffices to impose assumption A4 to ensure, by Chebyshev's inequality, that $n \, b_n^{1/2} I_{4n} = o_p(1)$. Lastly, recall that $I_{1n} = \sum_{i<j} H_n(Z_i, Z_j)$, where

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

17

Because $H_n(Z_i, Z_j)$ is symmetric, centered and such that $E[H_n(Z_i, Z_j)|Z_j] = 0$ almost surely, I_{1n} is a degenerate U-statistic. Fan and Li's (1999) central limit theorem for degenerate U-statistics implies that, under assumptions A1 to A4, $n\, b_n^{1/2} I_{1n} \xrightarrow{d} N(0, \Omega)$, where

$$
\begin{aligned}
\Omega &= \frac{n^4 b_n}{2} E_{Z_1, Z_2}[H_n^2(Z_1, Z_2)] \\
&= 2b_n \int_{Z_1, Z_2} \left[\int \breve{r}_n(z, Z_1) \breve{r}_n(z, Z_2)\, \mathrm{d}z \right]^2 f(Z_1, Z_2)\, \mathrm{d}(Z_1, Z_2) \\
&= 2b_n \int \left[\int \breve{r}_n(z, Z) \breve{r}_n(z', Z) f(Z)\, \mathrm{d}Z \right]^2 \mathrm{d}(z, z') \\
&= 2 \int \varphi^2(z) \left[\int K(u) K(u + v) f(z - u b_n)\, \mathrm{d}u \right. \\
&\qquad\qquad \left. - b_n \int K(u) f(z - u b_n)\, \mathrm{d}u \int K(u) f(z + v b_n - u b_n)\, \mathrm{d}u \right]^2 \mathrm{d}(z, v) \\
&\cong 2 \int \varphi^2(z) \left[\int K(u) K(u + v) f(z - u b_n) \right]^2 \mathrm{d}(z, v) \\
&\cong 2 v_K \int \varphi^2(z)\, f(z)\, \mathrm{d}F(z),
\end{aligned}
$$

which completes the proof.

PROOF OF PROPOSITION 1: Consider the second-order functional Taylor expansion

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

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

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

18

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

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

As is apparent, imposing the null hypothesis induces singularity in the first functional derivative $\mathrm{D}\Lambda_f$. To complete the proof, it then suffices to appreciate that, under the null, the second-order derivative reads

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

given that all other terms will depend on $f_{iXj} - g_{iXj}$. Observe, however, that $\mathrm{D}g_{iXj}$ converges at a faster rate than does h_{iXj} due to its lower dimensionality. The result then follows from a straightforward application of Lemma 1 with $\varphi(a_1, x, a_2) = f_{iXj}(a_1, x, a_2)$.

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 to A4,

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

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

$$\begin{aligned}
\Lambda_{f^{[n]}} &= E\left[f^{[n]}(d_{i,n}, X_{i,n}, d_{j,n}) - g^{[n]}(d_{i,n}, X_{i,n}, d_{j,n})\right]^2 + O_p\left(n^{-1/2}\right) \\
&= n^{-1}b_n^{-1/2}\ell_2 + o_p\left(n^{-1}b_n^{-1/2}\right).
\end{aligned}$$

References

Aït-Sahalia, Y., 1994, The delta method for nonparametric kernel functionals, Graduate School of Business, University of Chicago.

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

Amaro de Matos, J., Rosário, J. S., 2002, Market power and feedback effects from hedging derivatives, International Journal of Theoretical and Applied Finance 5, 845–875.

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.

Dawid, A. P., 1979, Conditional independence in statistical theory, Journal of the Royal Statistical Society B 41, 1–31.

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

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

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., 1999, Nonparametric tests for the Markov property, Getulio Vargas Foundation.

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.

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.

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.

Meitz, M., Saikkonen, P., 2004, Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models, Stockholm School of Economics and University of Helsinki.

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., 1994, The stationary bootstrap, Journal of the American Statistical Association 89, 1303–1313.

Robinson, P. M., 1991, Consistent nonparametric entropy-based testing, Review of Economic Studies 58, 437–453.

Rosemblatt, M., 1975, A quadratic measure of deviation of two-dimensional density estimates and a test of independence, Annals of Statistics 3, 1–14.

Silverman, B. W., 1986, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.

Skaug, H. J., Tjøstheim, D., 1993, Nonparametric tests of serial independence, in: T. S. Rao (ed.), Developments in Time Series Analysis, Chapman and 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.
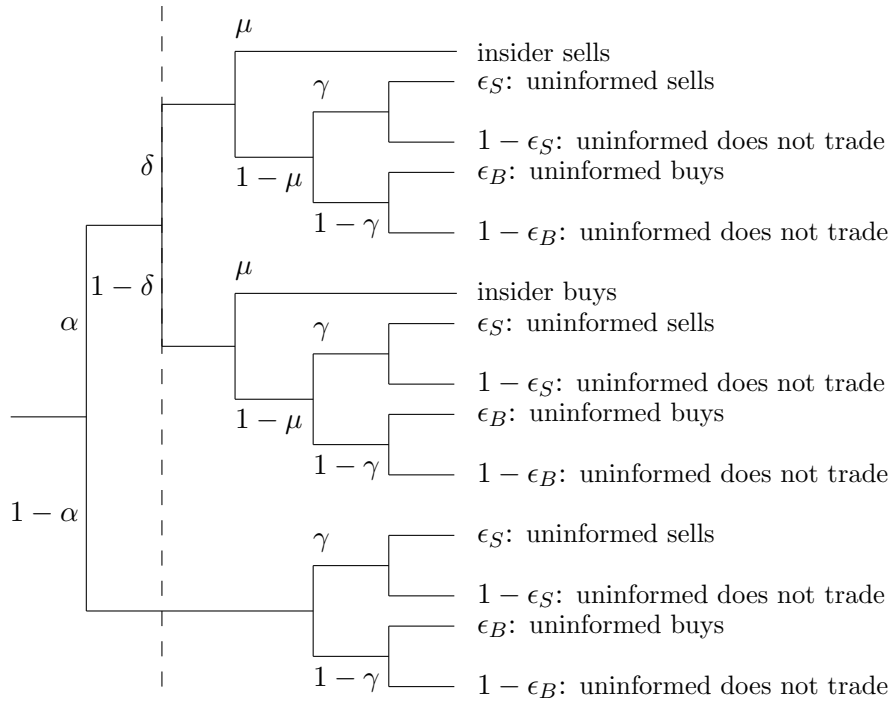
FIGURE 1

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.

TABLE 1

Phillips and Perron's (1988) unit root tests

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 $\ell = \lfloor 4(T/100)^{2/9} \rfloor$, where $\lfloor z \rfloor$ denotes the integer part of z.

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
Coca-Cola	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

TABLE 2

NONPARAMETRIC TESTS OF THE MARKOV PROPERTY

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 499 Markov bootstrap samples.

stock	duration		adjusted duration	
	$\hat{\lambda}_n$	p-value	$\hat{\lambda}_n$	p-value
Boeing	2.8979	(0.0019)	4.0143	(0.0000)
		[0.0002]		[0.0000]
Coca-Cola	19.4297	(0.0000)	18.6433	(0.0000)
		[0.0000]		[0.0000]
Disney	-3.2095	(0.9993)	-2.6822	(0.9963)
		[0.9940]		[0.7820]
Exxon	-1.0120	(0.8442)	0.4234	(0.3360)
		[0.6680]		[0.1480]
IBM	20.0711	(0.0000)	14.1883	(0.0000)
		[0.0002]		[0.0000]