Econometric Theory, **14**, 1998, 744–769. Printed in the United States of America.

TESTING FOR EMBEDDABILITY BY STATIONARY REVERSIBLE CONTINUOUS-TIME MARKOV PROCESSES

JEAN-PIERRE FLORENS
Université des Sciences Sociales de Toulouse

ERIC RENAULT
CREST-INSEE

NIZAR TOUZI
Université Paris IX Dauphine

Given an observation of a discrete-time process $\{Y_i, i = 0 \ldots n\}$ assumed to be Markov, stationary, and time reversible, we develop a (conservative) test procedure of embeddability by a continuous-time reversible Markov process. The test statistic is derived from a set of moment inequality restrictions implied by the spectral properties of such continuous-time processes. Most interesting is that the embeddability condition of interest is a direct extension of the well-known embeddability problem by a two-state Markov chain. Empirical experiments show that the embeddability hypothesis is rejected more frequently for exchange rate daily data than for stock indices data.

1. INTRODUCTION

Continuous-time models are extensively used in finance theory and in labor market analysis. Computations with continuous-time models are often easier than in the discrete framework. The treatment of data observed at different frequencies may be done unambiguously, and a continuous-time model is a better representation of a structural phenomenon because its structure does not depend on a specific observation rule. Therefore, faced with a sequence $(y_{t_0}, \ldots, y_{t_m})$ of observations of a real variable measured at discrete times $t_0, \ldots, t_n$, an econometrician will frequently specify a latent continuous-time process Y_t whose realizations at time $t_0, \ldots, t_n$ are assumed to be the observed data.

There is in general an infinite family of continuous-time models such that the probability distribution of Y_{t_i}, $i = 1, 2, \ldots, n$, coincides with the data-generating process, and the set of latent continuous-time candidates is usually restricted to

We are grateful to Guillaume Lezan (GEREMAQ and Crédit Agricole) for computational assistance. We also thank Lars P. Hansen for useful comments. Address correspondence to: Jean-Pierre Florens, Institut d'Economie Industrielle, Université des Sciences Sociales, Place Anatole France, 31042 Toulouse cedex, France; e-mail: florens@cict.fr.

 © 1998 Cambridge University Press 0266-4666/98 $9.50

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

a class characterized by some properties. The main example is given by the Markov processes: if the discrete-time observations are generated by a Markov process, the underlying continuous-time model will be constrained to satisfy the Markov property. More properties may be required for the latent process, such as continuity or regularity of the sample paths, homogeneity or stationarity, and so on.

Given this set of limitations two questions arise naturally: the existence of the process $\{Y_t\}$ satisfying the hypothesis and the unicity of such a continuous-time process. The first question is called the embeddability problem and the second one the aliasing problem. A large literature exists on embeddability in the case of finite state space Markov processes and a presentation of this particular question was given by Geweke, Marshall, and Zarkin (1986). The embeddability problem of discrete-time ARMA models by continuous-time ARMA processes was addressed by Chan and Tong (1987) and He and Wang (1989). The aliasing property will not be discussed in this paper; we refer the interested reader to Bergstrom (1984).

The paper is concerned with the embeddability of discrete-time real observations by a stationary reversible continuous-time Markov process. An important example of such continuous-time processes is provided by the class of stationary scalar diffusions characterized by the stochastic differential equation

$$dY_t = \mu(Y_t)\,dt + \sigma(Y_t)\,dW_t, \tag{1}$$

where $\{W_t, t \geq 0\}$ is a scalar Brownian motion and μ and σ are two real-valued functions called the drift and the diffusion coefficient. Under minor regularity conditions, stationary scalar diffusions are also time reversible in the following sense: for any $s < t$, the distribution of Y_t given $Y_s = y$ coincides with that of Y_s given $Y_t = y$.

A first necessary condition of embeddability follows from the observation that, if the underlying process is stationary continuous-time Markov process, then the discrete-time data are necessarily generated by a stationary Markov process. In this paper we assume this first embeddability condition satisfied, and we focus on restrictions implied by the reversibility of the process.

To illustrate the kind of restrictions that will be imposed by the embeddability we first consider in Section 2 two examples: the embeddability of a finite state space Markov process and the embeddability of an autoregressive Gaussian process by an Ornstein–Uhlenbeck diffusion model. Section 3 briefly recalls the main tools convenient to relate the discrete-time process to the underlying reversible Markov continuous-time process, if such exists. These tools, essentially based on the semigroup of conditional expectation operators and the infinitesimal generator of the continuous-time Markov process, allow us to derive a set of restrictions on the first-order autocorrelations of the process (see Hanson and Scheinkman, 1995). These restrictions are testable and should be tested before using the discrete-time observation to estimate the characteristics of the underlying continuous-time process (the trend and the diffusion coefficients in the case of scalar diffusions).

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

Section 4 presents the test of embeddability we propose and its main properties. Section 5 is devoted to a further analysis of the implicit null hypothesis of our test through two interesting examples and provides further restrictions to be tested for the embeddability by stationary scalar diffusions.

Finally Section 6 presents some empirical results. Whereas the embeddability hypothesis is not rejected for stock index daily data, it turns out to be rejected for many exchange rates' daily data. This is consistent with a fairly well-documented observation that, although one may sometimes accept that stock price indices follow univariate diffusion processes whose volatility depends only on time and the concurrent value of the process, such an assumption is much more heroic for exchange rates[1] because extremely high values of kurtosis coefficients lead one to introduce hidden Markov processes (stochastic volatility processes) to capture this evidence of mixture of distributions. Of course, observable time series of exchange rates with hidden Markov stochastic volatility are no longer Markov, but it appears that testing for embeddability (without directly questioning Markovianity) is a powerful trick to reject the univariate diffusion hypothesis.

2. TWO INTRODUCTORY EXAMPLES

2.1. The Finite State Space Markov Processes

The observations $y_{t_0}, \ldots, y_{t_n}$ are assumed to be valued in a finite state space $E = \{1, \ldots, K\}$, and we suppose for simplicity that $t_i - t_{i-1}$ is constant and equal to 1 by convention. Then we drop the index t and the data are denoted by the sequence $y_0, \ldots, y_n$. The data-generating process is a homogeneous Markov model characterized by the transition matrix

$$T = (T_{jk})_{(j,k)\in E\times E} \qquad T_{jk} = P(Y_i = k | Y_{i-1} = j) \tag{2}$$

and by the initial distribution $Q = (Q_j)_{j\in E}$. The process is assumed to be stationary, and Q satisfies the condition $Q'T = Q'$, where Q' is the transpose matrix of Q. The elements of T (and Q) may be consistently estimated from the data, and the asymptotic distributions of these estimators are well known (see, e.g., Billingsley, 1961). Then one can assume that T and Q are known from the sample.

The embeddability problem raises the question of the existence of a standard Markov stationary continuous-time process whose discrete-time observations are generated by the previous Markov chain. A homogeneous continuous-time Markov process is characterized by the transition probabilities

$$T(t) = (T_{jk}(t))_{(j,k)\in E\times E} \qquad T_{jk}(t) = P(Y_{t+u} = k | Y_u = j) = P(Y_t = k | Y_0 = j) \quad \forall u, t, \tag{3}$$

and the stationary marginal distribution must verify $Q'T(t) = Q'$ for any t. The matrices $T(t)$ satisfy in particular the relation $T_t = T_s T_u$ if $s + u = t$. We restrict our attention to standard Markov processes or equivalently to strongly continuous matrices T_t satisfying

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

$$\lim_{t\downarrow 0} T_{jk}(t) = \begin{cases} 0 & \text{if } k \neq j \\ 1 & \text{if } k = j \end{cases}. \tag{4}$$

This property implies that the sample paths are right continuous.

The discrete-time process is characterized by the transition T, and the embedding question is then the following: is it possible to construct a family $T(t)$ strongly continuous such that $T = T(1)$? Intuitively this question is tantamount to the "divisibility" of T into $T(\alpha)$ and $T(1-\alpha)$ such that $T = T(\alpha)T(1-\alpha)$ for $\alpha \in (0,1)$.

The solution of this problem requires the definition of the matrix A of intensities. This matrix is defined by

$$A_{jk} = \lim_{t\downarrow 0} \frac{1}{t} T_{jk}(t) \quad \text{if } j \neq k,$$

$$A_{jj} = -\sum_{k\neq j} A_{jk}. \tag{5}$$

It turns out that (see Cox and Miller, 1965)

$$T(t) = \exp At = \sum_{r=0}^{\infty} \frac{A^r t^r}{r!}. \tag{6}$$

Then the embeddability issue reduces to the existence of a matrix A with positive elements (except on the diagonal that is defined by (4)) solution of the equation

$$T = \exp A. \tag{7}$$

If A is computed from (7), the family $T(t)$ may be deduced from (6).

The case $K = 2$ can be easily treated (see, e.g., Fougère and Kamionka, 1992). Let

$$T = \begin{pmatrix} T_{11} & 1-T_{11} \\ 1-T_{22} & T_{22} \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} -A_{12} & A_{12} \\ A_{21} & -A_{21} \end{pmatrix}. \tag{8}$$

The marginal probability Q is defined by

$$\left(\frac{1-T_{22}}{(1-T_{11})+(1-T_{22})}, \frac{1-T_{11}}{(1-T_{11})+(1-T_{22})} \right).$$

The joint distribution of (Y_i, Y_{i+1}) is then symmetric. Therefore time reversibility is fulfilled in discrete time. The matrix A has two eigenvalues 0 and $-(A_{12}+A_{21})$ and can be represented in the form $A = D\Lambda D'$ where Λ is the diagonal matrix of eigenvalues and D an orthonormal matrix of eigenvectors. From (6), it follows that

$$T(t) = D \begin{pmatrix} 1 & 0 \\ 0 & e^{-(A_{12}+A_{21})t} \end{pmatrix} D'. \tag{9}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

The eigenvalues of $T(t)$ are 1 and $e^{-(A_{12}+A_{21})t} \in (0,1)$, and the eigenvectors of $T(t)$ are identical to those of A.

The eigenvalues of T are 1 and $(T_{11} + T_{22}) - 1$, and the equation $T = \exp A$ admits a solution if and only if

$$T_{11} + T_{22} - 1 > 0. \tag{10}$$

An elementary computation shows that (10) is equivalent to

$$E[\varphi(Y_i)\varphi(Y_{i+1})] > 0 \quad \text{for any } \varphi : E \to \mathbb{R}. \tag{11}$$

This expectation does not depend on i and involves the joint distribution of two successive elements.

The problem is considerably more complicated if K is greater than 2 because the eigenvalues of T and A are not necessarily real. To the best of our knowledge, general necessary and sufficient conditions for embeddability are not fully available in the literature (some conditions are given by Geweke et al., 1986).

2.2. The Gaussian AR(1) Process

We still assume that the observation times satisfy $t_i - t_{i-1} = 1$ and that the discrete-time sample $(y_0, \ldots, y_n)$ is Markovian homogeneous with state space equal to $\mathbb{R}$. The transition probability is Gaussian and satisfies

$$Y_i \mid Y_{i-1} = y \sim \mathcal{N}(\rho y, \tau^2) \qquad |\rho| < 1. \tag{12}$$

The stationary marginal probability distribution Q is normal with zero mean and variance equal to $\tau^2/(1 - \rho^2)$. The parameters τ^2 and ρ may be estimated from the sample, and they are now assumed to be known.

The purpose is now the search for a continuous-time diffusion process that generates a data-generating process on discrete-time observations satisfying (12). It is well known that the Ornstein–Uhlenbeck process is the only diffusion with Gaussian stationary distribution satisfying the condition $E(Y_i | Y_{i-1} = y) = \rho y$ (see Appendix C), and we therefore restrict our attention to continuous-time processes Y_t such that

$$dY_t = -kY_t\,dt + \sigma\,dW_t \qquad (k > 0, \sigma > 0). \tag{13}$$

By a simple integration we get

$$Y_t \mid Y_{t-1} = y \sim \mathcal{N}\left(e^{-k}y, \frac{\sigma^2}{2k}[1 - e^{-2k}]\right).$$

The model (12) is then compatible with (13) if and only if ρ is positive (see, e.g., Chan and Tong, 1987). If it is the case the knowledge of ρ and τ^2 uniquely determines the knowledge of k and σ.

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

Finally let us remark that the condition $\rho > 0$ is equivalent to the condition

$$E(\varphi(Y_i)\varphi(Y_{i+1})) > 0 \tag{14}$$

for any nonzero φ square integrable with respect to the marginal distribution. The expectation in (14) is taken over the joint distribution of (Y_i, Y_{i+1}), which does not depend on i. Although (14) seems to be an elementary property of the bivariate normal distribution, it is not so easy to check (see Appendix A for a direct proof). Nevertheless, it can be easily understood thanks to the Hilbert basis of Hermite polynomials through the following arguments:

(i) Let $\{\varepsilon_j, j \in \mathbb{N}\}$ be the Hermite polynomials (see Appendix B) relative to the $\mathcal{N}(0 \cdot \tau^2/1 - \rho^2)$ probability. They define an orthonormal basis of $L^2(Q) \cdot (E(\varepsilon_j \varepsilon_\ell) = 0$ if $j \neq \ell$ and $E(\varepsilon_j^2) = 1)$. Then it is easily checked that

$$E(\varepsilon_j(Y_{i+1}) \mid Y_i = y) = \rho^j \varepsilon_j(y),$$

i.e., the ε_j's are the eigenvectors corresponding to eigenvalues equal to ρ^j of the operator

$$\varphi \mapsto E(\varphi(Y_{i+1}) \mid Y_i = y); \tag{15}$$

(ii) For $\varphi = \varepsilon_j$, (14) is immediate; for a general element φ of $L^2(Q)$, the result follows from its decomposition on the basis $\{\varepsilon_j, j \in \mathbb{N}\}$.

As we shall show in the following section, if condition (14) is violated then the discrete-time AR(1) Gaussian process is not embeddable by any stationary reversible continuous-time Markov process.

3. A NECESSARY CONDITION OF EMBEDDABILITY BY STATIONARY REVERSIBLE MARKOV PROCESSES

In this section $\{Y_t, t \geq 0\}$ is an n-dimensional stationary reversible continuous-time Markov process on a probability space $(\Omega, \mathcal{F}, P)$. Let Q be the probability measure induced on $\mathbb{R}^n$ by Y_t and $L^2(Q)$ the Hilbert space of (equivalence classes of) measurable complex-valued functions f defined on $\mathbb{R}^n$ for which $\int_{\mathbb{R}^n} |f(y)|^2 \, dQ(y) < \infty$. The scalar product on $L^2(Q)$ is $\langle f, g \rangle = \int_{\mathbb{R}^n} \overline{f(y)} g(y) \, dQ(y)$, where $\overline{f(y)}$ is the natural conjugate of $f(y)$.

For any $t \geq 0$, the conditional expectation operator $\mathcal{T}_t$ defined by

$$\mathcal{T}_t \varphi(y) = E[\varphi(Y_t) \mid Y_0 = y], \qquad \varphi \in L^2(Q) \tag{16}$$

maps $L^2(Q)$ into itself. It is then well known that $\{\mathcal{T}_t, t \geq 0\}$ is a semigroup of unit norm linear operators that we assume to be strongly continuous, i.e., $\mathcal{T}_t \varphi \to \varphi$ (in $L^2(Q)$) as t goes to zero for any $\varphi \in L^2(Q)$).

Under the strong continuity condition, the semigroup is differentiable at zero for some functions in $L^2(Q)$. Precisely, for $\varphi \in L^2(Q)$, the so-called infinitesimal generator $\mathcal{A}$ is such that $\mathcal{A}\varphi$ is defined as the limit in the sense of $L^2(Q)$ of

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

$(1/t)[\mathcal{T}_t\varphi - \varphi]$ as $t \downarrow 0$, whenever this limit exists. The domain $D(\mathcal{A})$ of this generator is the family of functions $\varphi \in L^2(Q)$ for which $\mathcal{A}\varphi$ is well defined. Under the strong continuity condition, it is then well known that $D(\mathcal{A})$ is dense in $L^2(Q)$ and $\mathcal{A}$ is a closed operator on $L^2(Q)$ (see Pazy, 1983, Corollary 2.5, p. 5).

The reversibility of the continuous-time process $\{Y_t, t \geq 0\}$ implies that the conditional expectation operator is self-adjoint, i.e., for any $t \geq 0$, $\langle \mathcal{T}_t\varphi, \psi\rangle = \langle \varphi, \mathcal{T}_t\psi\rangle$ for all $\varphi, \psi \in L^2(Q)$. This proves that the infinitesimal generator $\mathcal{A}$ is self-adjoint, a crucial property for our set of moments restrictions. Notice that, for scalar diffusions, time reversibility is almost implied by stationarity. See Section 5.2 for more details.

Before presenting the basic result for our testing procedure, we first recall the definition of the spectrum of a linear operator. For some $\lambda \in \mathbb{C}$, consider the linear operator $\lambda\mathcal{I} - \mathcal{A}$, where $\mathcal{I}$ is the identity operator of $L^2(Q)$. The term λ is said to belong to the *resolvent set* of $\mathcal{A}$ if (i) $\lambda\mathcal{I} - \mathcal{A}$ is one-to-one, (ii) the range of $\lambda\mathcal{I} - \mathcal{A}$ is $L^2(Q)$, and (iii) $(\lambda\mathcal{I} - \mathcal{A})^{-1}$ is a bounded linear operator on $L^2(Q)$. The spectrum of $\mathcal{A}$ is the set of those $\lambda \in \mathbb{C}$ that do not belong to the resolvent set and will be denoted by $\mathrm{Sp}(\mathcal{A})$ in the sequel.

Closed self-adjoint operators have closed real spectrum (see Weidmann, 1980, Theorem 5.14, p. 98, and Theorem 5.23, p. 108). Moreover because $\mathcal{A}$ is the infinitesimal generator of a contraction semigroup, by the Hille–Yosida theorem (see Pazy, 1983, Theorem 3.1, p. 8), we have

$$\mathrm{Sp}(\mathcal{A}) \text{ is a closed subset of } (-\infty, 0]. \tag{17}$$

Notice that zero is always an eigenvalue associated with the constant eigenvector.

We first derive the basic moments conditions for our testing procedure by a heuristic argument. As defined, the infinitesimal generator is the derivative at zero of the associated semigroup of linear operators in the sense of $L^2(Q)$. This suggests that $\mathcal{T}_t$ can be obtained from the infinitesimal generator by $\exp(t\mathcal{A})$ as in the Markov chain case (6). However, because $\mathcal{A}$ is an unbounded operator, the last expression can not be defined in the usual way. The correct link between the semigroup and its associated infinitesimal generator is established through the well-known exponential formula

$$\mathcal{T}_t\varphi = \lim_{n\to\infty}\left(\mathcal{I} - \frac{t}{n}\mathcal{A}\right)^{-n}\varphi$$

(see Pazy, 1983, Theorem 8.3, p. 33). Such an exponential formula suggests that the spectrum of $\mathcal{T}_t$ is obtained by exponentiating the spectrum of $t\mathcal{A}$, and therefore we expect to have $\mathrm{Sp}(\mathcal{T}_t) \subset [0,1]$ for all $t \geq 0$, as a consequence of (17). Therefore

$$\langle \varphi, \mathcal{T}_t\varphi\rangle \geq 0 \quad \text{for any } t \geq 0,\ \varphi \in L^2(Q) \tag{18}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

(for the equivalence between $\mathrm{Sp}(\mathcal{T}_t) \subset [0,1]$ and (18), see Weidmann, 1980, Proposition (3), p. 197). For $t = 1$ the last property can be written for real-valued functions:

$$E[\varphi(Y_t)\varphi(Y_{t+1})] \geq 0 \quad \text{for } t \geq 0,\ \varphi \in L^2(Q). \tag{19}$$

The last derivation of the moments condition (18) fails to be rigorous because we do not have $\mathrm{Sp}(\mathcal{T}_t) \subset [0,1]$ as a consequence of (17) together with the approximate exponential formula. Indeed, the spectrum of a linear operator can be decomposed in three nonintersecting parts. The point spectrum $\sigma_p(\mathcal{A})$ is the set of eigenvalues of $\mathcal{A}$, i.e., $\lambda \in \sigma_p(\mathcal{A})$ if $\lambda I - \mathcal{A}$ is not one-to-one. An element of the spectrum λ is said to be in the continuous spectrum $\sigma_c(\mathcal{A})$ if $\lambda I - \mathcal{A}$ is one-to-one, but its range is a proper subset of $L^2(Q)$ dense in $L^2(Q)$. Finally, an element of the spectrum λ is said to be in the residual spectrum $\sigma_r(\mathcal{A})$ if $\lambda I - \mathcal{A}$ is one-to-one and its range is not dense in $L^2(Q)$.

Bounded self-adjoint operators have no residual spectrum, and then $\sigma_r(\mathcal{T}_t)$ is void (see Dunford and Schwartz, 1963, Lemma 3, p. 904). Furthermore, the point spectra of $\mathcal{A}$ and $\mathcal{T}_t$ for all $t > 0$ are related by

$$e^{t\sigma_p(\mathcal{A})} \subset \sigma_p(\mathcal{T}_t) \subset e^{t\sigma_p(\mathcal{A})} \cup \{0\} \tag{20}$$

(see Pazy, 1983, Theorem 2.4, p. 46). However, concerning the continuous spectrum, we only have

$$e^{t\sigma_p(\mathcal{A})} \subset \sigma_p(\mathcal{T}_t),$$

and the converse does not hold (see Pazy, 1983, Theorem 2.6, p. 47). We are nevertheless able to prove the following theorem.

THEOREM. *Let (Y_t) be a univariate continuous-time time-reversible stationary Markov process. Then the discrete-time (Y_{t_i}) process is time reversible, stationary, Markovian and satisfies*

$$E[\varphi(Y_{t_i})\varphi(Y_{t_{i+1}})] \geq 0 \quad \forall \varphi \in L^2(Q).$$

Proof. See Appendix D.

Hence the well-known embeddability condition in the two-state Markov chain framework (reviewed in Section 2.1) holds for stationary reversible continuous-time Markov processes. However, in contrast with the two-state Markov chain case, it is only a necessary condition of embeddability (see Section 5 for further discussion of this issue).

Obviously, the first-order autoregressive process with negative autoregressive coefficient considered in Section 2.2 does not satisfy relation (19) ($\varphi(x) = x$ is in $L^2(Q)$ and violates condition (19)) and therefore can not be embedded by any stationary reversible continuous-time Markov process. Whereas this example is very simple to solve, in general it is not so easy to find an $L^2(Q)$ function that

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

violates condition (19). In the remainder of this paper, we exploit (19) to construct a (conservative) test of embeddability of an observed discrete-time process by a stationary reversible continuous-time Markov process.

4. A CONSERVATIVE TEST OF EMBEDDABILITY BY STATIONARY REVERSIBLE CONTINUOUS-TIME MARKOV PROCESSES

4.1. The Null Hypothesis

As announced in the introduction, we assume in this section that the observation of the discrete-time process Y_i, $i = 0,1,2,\ldots,n$ allows the statistician to check that it is stationary, Markovian, and time reversible. In particular, we denote by Q the common probability distribution of each Y_i that can be estimated from discrete-time observations.

Therefore, we assume in the sequel that Q is given and then $L^2(Q)$ is well defined. The null hypothesis of interest is the infinite set H_0 of moment inequality restrictions:

$$H_0 : \forall \varphi \in L^2(Q) \quad E[\varphi(Y_i)\varphi(Y_{i+1})] \geq 0. \tag{21}$$

Because $L^2(Q)$ is a separable Hilbert space, it admits a countable Hilbert basis $\{\varphi_j, j \geq 0\}$. Moreover, $L^2(Q)$ contains the constant functions, and therefore we can assume w.l.o.g. that $\varphi_0 = 1$. In this paper, we do not address the issue of estimating the Hilbert basis $\{\varphi_j, j \geq 0\}$, and we consider it as given (for a similar approach, see Conley, Hansen, Luttmer, and Scheinkman, 1995).

We denote by K_p the Hilbert space spanned by $\varphi_1,\ldots,\varphi_p$ and by H_{0p} the null hypothesis:

$$H_{0p} : \forall \varphi \in K_p \quad E[\varphi(Y_i)\varphi(Y_{i+1})] \geq 0. \tag{22}$$

The purpose of this section is to construct a consistent test of asymptotic size α for H_{0p} for a given p. Of course such a test is conservative for H_0 because K_p is a proper subset of $L^2(Q)$. The subspace K_p does not contain φ_0 because the constant functions satisfy obviously H_0 for any discrete-time process $\{Y_i, i = 0 \ldots n\}$, and therefore there is no need to include them in the set of test functions.

Notice that if one wants to increase the set of restrictions K_p, the null hypothesis H_{0p} and the additional hypothesis $H_0(p+1) : E[\varphi_{p+1}(Y_i)\varphi_{p+1}(Y_{i+1})] \geq 0$ have to be considered jointly. Indeed if H_{0p+1} is rejected as soon as H_{0p} or $H_0(p+1)$ are rejected at level α, then it is clear[2] that the size of the resulting test of H_0 may be dramatically greater than α.

Moreover the joint test of the set of restrictions implied by H_{0p} has to be performed only for values of p relatively small with respect to the sample size n. As a matter of fact, because K_p is a finite-dimensional Hilbert space, H_{0p} can be written in the following parametric form:

$$H_{0p} : \forall \theta \in \mathbb{R}^p \quad \theta' \Gamma_p \theta \geq 0, \tag{23}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

where Γ_p is the matrix of size p whose (j,k) coefficient is $E[\varphi_j(Y_i)\varphi_k(Y_{i+1})]$, $j,k=1,\ldots,p$. Therefore the test statistic of H_{0p} will be founded on the empirical counterpart of Γ_p. Of course, if $p>n$, such an empirical counterpart is a singular matrix, and then (23) contains an infinite set of redundant restrictions. Besides, if p is not sufficiently small with respect to n, there is a substantial risk of multicolinearity, which also induces a singularity problem.

Thanks to the time reversibility of the process, Γ_p is a symmetric matrix and then can be diagonalized in an orthonormal basis of $\mathbb{R}^p$ with real eigenvalues. Therefore H_{0p} can be rewritten in

$$H_{0p}: \lambda_p \geq 0, \tag{24}$$

where λ_p is the smallest eigenvalues of Γ_p. In the following, the ranked eigenvalues of Γ_p are denoted by

$$\lambda_p = \lambda_p^1 \leq \cdots \leq \lambda_p^p < 1, \tag{25}$$

where the last inequality follows from the contracting feature of the conditional expectation operator for any discrete-time process $\{Y_i, i=0\ldots n\}$.

4.2. Definition of the Testing Procedure

We consider a Wald test procedure based on the asymptotic probability distribution of a consistent estimator of λ_p.

A natural estimator of the moment matrix Γ_p is the associated empirical counterpart

$$\hat{\Gamma}_p^+ = \frac{1}{n}\sum_{i=0}^{n-1} \phi(Y_i)\phi'(Y_{i+1}), \tag{26}$$

where $\phi(x)$ is the column vector $(\varphi_j(x), j=1\ldots p)$. However from the time reversibility of the process,

$$\Gamma_p = E[\phi(Y_i)\phi'(Y_{i+1})] = E[\phi(Y_{i+1})\phi'(Y_i)], \tag{27}$$

and therefore, a more accurate consistent estimator is obtained by

$$\hat{\Gamma}_p = \tfrac{1}{2}[\hat{\Gamma}_p^+ + \hat{\Gamma}_p^-], \tag{28}$$

where

$$\hat{\Gamma}_p^- = \frac{1}{n}\sum_{i=0}^{n-1} \phi(Y_{i+1})\phi'(Y_i). \tag{29}$$

In contrast with $\hat{\Gamma}_p^+$ and $\hat{\Gamma}_p^-$, the estimator $\hat{\Gamma}_p$ fulfills the symmetry condition of true Γ_p, so that it can be diagonalized in an orthonormal basis of $\mathbb{R}^{p+1}$. We denote by

$$\hat{\lambda}_p = \hat{\lambda}_p^1 \leq \cdots \leq \hat{\lambda}_p^p$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

its eigenvalues. The test statistic of H_{0p} is deduced from the asymptotic probability distribution of $\hat{\lambda}_p$. We shall prove later that

$$\sqrt{n}[\hat{\lambda}_p - \lambda_p] \xrightarrow{\mathcal{D}} \mathcal{N}[0, \sigma^2] \quad \text{as } n \to \infty, \tag{30}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in the distribution sense. The one-sided test of $\lambda_p \geq 0$ is then defined by the critical region

$$\frac{\sqrt{n}\hat{\lambda}_p}{\hat{\sigma}} \leq -u_{1-\alpha}, \tag{31}$$

where $u_{1-\alpha}$ is the $(1-\alpha)$ quantile of the $\mathcal{N}(0,1)$ distribution and $\hat{\sigma}$ is a consistent estimator of σ.

The following section is devoted to the proof of (30) and the way to estimate σ for practical application of the test procedure.

4.3. Asymptotic Distribution of $\hat{\lambda}_p$

The term $\hat{\lambda}_p$ is a well-defined function of $\hat{\Gamma}_p$. For sake of simplicity, we shall assume that the true unknown matrix Γ_p has p distinct eigenvalues, and therefore $\hat{\Gamma}_p$ fulfills the same property for n sufficiently large. Notice that for large p the set of eigenvalues of Γ_p converges to the spectrum of the conditional expectation operator whose eigenvalues are known to be simple (see Weidmann, 1987), and therefore the last assumption holds asymptotically. Nevertheless, the procedure described subsequently can be easily adapted to the case of multiple eigenvalues but with cumbersome notations.

Our simplifying assumption is useful for the application of the delta theorem because we have in this case at our disposal an easy expression of the derivative of λ_p with respect to each coefficient $\gamma_{k\ell}$ of Γ_p:

$$\frac{\partial \lambda_p}{\partial \gamma_{k\ell}}(\Gamma_p) = \frac{\Delta_{k\ell}(\lambda_p)}{\prod_{j=2}^{p}(\lambda_p^j - \lambda_p)} \qquad k, \ell = 1, \ldots, p, \tag{32}$$

where $\Delta_{k\ell}(\lambda)$ is the cofactor of the (k,ℓ) coefficient of the matrix $\Gamma_p - \lambda I$, I being the identity matrix of $\mathbb{R}^p$ (see, e.g., Gandolfo, 1981). From the delta theorem and (32), the key ingredient for the asymptotic probability distribution (a.p.d.) of $\hat{\lambda}_p$ is the a.p.d. of $\hat{\Gamma}_p$ or more precisely of vech $\hat{\Gamma}_p$ where vech is the operator that stacks the elements on and below the main diagonal of a symmetric matrix (see, for instance, Lütkepohl, 1991, p. 465).

Assuming some ergodicity and mixing properties of the underlying discrete-time Markov process Y_i (see, e.g., the Doeblin conditions in Doob, 1953, Theorem 7.5, p. 232), the a.p.d. of $\hat{\Gamma}_p$ is given by

$$\sqrt{n}(\text{vech}\,\hat{\Gamma}_p - \text{vech}\,\Gamma_p) \xrightarrow{\mathcal{D}} \mathcal{N}(O, \Sigma) \tag{33}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

and therefore we deduce the a.p.d. of $\hat{\lambda}_p$:

$$\sqrt{n}(\hat{\lambda}_p - \lambda_p) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where

$$\sigma^2 = \frac{\partial \lambda_p}{\partial (\operatorname{vech} \Gamma_p)'} (\operatorname{vech} \Gamma_p) \Sigma \frac{\partial \lambda_p}{\partial (\operatorname{vech} \Gamma_p)} (\operatorname{vech} \Gamma_p) \tag{34}$$

and $\partial \lambda_p / \partial (\operatorname{vech} \Gamma_p)(\operatorname{vech} \Gamma_p)$ is the column matrix of size $p(p+1)/2$ defined by

$$\left(\frac{\partial \lambda_p}{\partial \gamma_{11}}, \frac{\partial \lambda_p}{\partial \gamma_{21}}, \ldots, \frac{\partial \lambda_p}{\partial \gamma_{p,1}}, \frac{\partial \lambda_p}{\partial \gamma_{12}}, \frac{\partial \lambda_p}{\partial \gamma_{22}}, \ldots, \ldots, \frac{\partial \lambda_p}{\partial \gamma_{p,p}} \right)'$$

evaluated at the point Γ_p.

A natural consistent estimator $\hat{\sigma}^2$ of σ^2 is obtained from the empirical counterpart of (34):

$$\hat{\sigma}^2 = \frac{\partial \lambda_p}{\partial (\operatorname{vech} \Gamma_p)'} (\operatorname{vech} \hat{\Gamma}_p) \hat{\Sigma} \frac{\partial \lambda_p}{\partial (\operatorname{vech} \Gamma_p)} (\operatorname{vech} \hat{\Gamma}_p). \tag{35}$$

The variance estimator given in (35) requires a consistent estimation $\hat{\Sigma}$ of Σ, which is not a trivial issue because the order of serial correlation of a nonlinear function $\phi(Y_i)\phi(Y_{i+1})' + \phi(Y_{i+1})\phi(Y_i)'$ of the Markov process $(Y_i)_{i \in \mathbb{N}}$ is generally unknown. Nevertheless this issue is now well documented, particularly in the generalized method of moments (GMM) literature (see, e.g., Ogaki, 1993). As a matter of fact, Σ is the asymptotic covariance matrix of $\sqrt{n} \operatorname{vech} \hat{\Gamma}_p$, where components of $\operatorname{vech} \hat{\Gamma}_p$ are defined by

$$\hat{\gamma}_{k\ell} = \frac{1}{n} \sum_{i=0}^{n-1} \psi_{k\ell}(y_i, y_{i+1}), \tag{36}$$

where $\ell = 1, \ldots, p$ and $k = \ell, \ldots, p$ with

$$\psi_{k\ell}(y_i, y_{i+1}) = \tfrac{1}{2}[\varphi_k(y_i)\varphi_\ell(y_{i+1}) + \varphi_k(y_{i+1})\varphi_\ell(y_i)]. \tag{37}$$

Let us denote by

$$\hat{\gamma} = (\hat{\gamma}_j)_{1 \le j \le p(p+1)/2}$$

the vector $\operatorname{vech} \hat{\Gamma}_p$ whose components are $\hat{\gamma}_{k\ell}$, $\ell = 1, \ldots, p$, $k = \ell, \ldots, p$, and in the same way

$$\psi(y_i, y_{i+1}) = (\psi_j(y_i, y_{i+1}))_{1 \le j \le p(p+1)/2}.$$

Then $\hat{\gamma} = 1/n \sum_{i=0}^{n-1} \psi(y_i, y_{i+1})$ and the matrix Σ is by definition

$$\Sigma = \operatorname{Var}_{as}[\sqrt{n}\hat{\gamma}] = \operatorname{Var}_{as}\left[\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \psi(Y_i, Y_{i+1}) \right] = \lim_{n=\infty} J_n \tag{38}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

with

$$J_n = \frac{1}{n}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\mathrm{Cov}[\psi(Y_i,Y_{i+1}),\psi(Y_j,Y_{j+1})]. \tag{39}$$

Following the notations of Andrews (1991), the estimand J_n can be rewritten as

$$J_n = \sum_{h=-n+1}^{n-1} R_n(h), \tag{40}$$

where

$$R_n(h) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{j=h}^{n-1}\mathrm{Cov}[\psi(Y_j,Y_{j+1}),\psi(Y_{j-h},Y_{j-h+1})] & \text{for } h \geq 0 \\ \dfrac{1}{n}\displaystyle\sum_{j=-h}^{n-1}\mathrm{Cov}[\psi(Y_{j+h},Y_{j+h+1}),\psi(Y_j,Y_{j+1})] & \text{for } h < 0. \end{cases}$$

Indeed, in our stationary framework, the terms of the sums $R_n(h)$ are identical and $\Sigma = \lim_{n=\infty} J_n$ is nothing but 2π times the spectral density matrix of the stationary process $(\psi(Y_j))_{j\in\mathbb{N}}$ evaluated at $\lambda = 0$. Nevertheless, the class of estimators of Σ proposed by Andrews (1991) and suggested by the Parzen (1957) class of kernel estimators of this spectral density can be used for more general alternatives of nonstationary processes; the considered estimators of Σ have the form

$$\hat{J}_n = \sum_{h=-n+1}^{n-1} k\left(\frac{h}{S_n}\right)\hat{R}_n(h), \tag{41}$$

where $k(\cdot)$ is a real-valued kernel, S_n is a bandwidth parameter, and

$$\hat{R}_n(h) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{j=h}^{n-1}[\psi(Y_j,Y_{j+1}) - \bar{\psi}_n]\psi'[Y_{j-h},Y_{j-h+1}] & \text{for } h \geq 0 \\ \dfrac{1}{n}\displaystyle\sum_{j=-h}^{n-1}[\psi(Y_{j+h},Y_{j+h+1}) - \bar{\psi}_n]\psi'[Y_j,Y_{j+1}] & \text{for } h < 0 \end{cases} \tag{42}$$

with

$$\bar{\psi}_n = \frac{1}{n}\sum_{j=0}^{n-1}\psi(Y_j,Y_{j+1}).$$

The estimators corresponding to these kernels generally place zero weights on $\hat{R}_n(h)$ for $h \geq S_n$ ($k(x) = 0$ for $|x| \geq 1$) so that $S_n - 1$ can be interpreted as a lag truncation parameter. We can use for instance the Bartlett's kernel:

$$k(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{43}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

Among the usual kernels, the Bartlett and the more complicated Parzen's kernels have the advantage of generating positive semidefinite estimators of Σ in finite samples. Andrews (1991) advocated an estimator of Σ that uses the so-called quadratic spectral kernel, which fulfills the same positivity requirement but does not place zero weights on any $\hat{R}_n(h)$; Andrews (1991) showed that it is an optimal kernel in the sense that it minimizes asymptotic mean squared error among the estimators of this form which are guaranteed to be positive semidefinite. Moreover Andrews (1991) provided formulas for optimal choice of the bandwidth parameter for a variety of kernels; these formulas include unknown parameters, and Andrews proposed an automatic bandwidth estimator in which these unknown parameters are estimated from the data.

5. IMPLICIT NULL HYPOTHESIS

5.1. Null Eigenvalues

The set of moment restrictions (19) has been derived under the assumption that the semigroup of conditional expectations induced by the underlying continuous-time process is strongly continuous, a necessary condition for the definition of the infinitesimal generator $\mathcal{A}$. The purpose of this section is to prove that, even in the case where the data-generating process is embeddable by a continuous-time process, the moment restrictions condition H_0 does not imply the strong continuity of the underlying continuous-time process.

Of course, a one-sided test such as (31) has little power against the boundary of the null hypothesis. By continuity, the power of this test is exactly equal to the level α when 0 is an eigenvalue of the operator, i.e., when there exists a nondegenerate $\varphi \in L^2(Q)$ satisfying

$$E[\varphi(Y_t)\varphi(Y_{t+1})] = 0. \tag{44}$$

Notice that the existence of such an eigenfunction does not prevent the discrete-time process from being embeddable by a strongly continuous continuous-time process in general (see Pazy, 1983, Theorem 2.4, p. 46). We now exhibit an example of a discrete-time process satisfying the testing hypothesis H_0 and embeddable by a continuous-time process that is not strongly continuous.

Consider the following discrete-time homogeneous Markov process:

The marginal probability distribution Q is the standard $\mathcal{N}(0,1)$ normal probability distribution.

The transition probability is defined by its density function:

$$f(y|x) = 2q(y)\{\pi \mathbb{I}_{[xy>0]} + (1-\pi)\mathbb{I}_{[xy\leq 0]}\}, \tag{45}$$

where q is the density function of $Q = \mathcal{N}(0,1)$, π is a real parameter in $[0,1]$, and

$$\mathbb{I}_{[xy>0]} = \begin{cases} 1 & \text{if } xy > 0 \\ 0 & \text{otherwise.} \end{cases}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

It is clear that this definition is consistent because

$$f(y|x)q(x) = f(x|y)q(y), \tag{46}$$

which proves that

$$\int_{-\infty}^{+\infty} f(y|x)q(x)\,dx = q(y).$$

The time reversibility of the associated discrete-time process follows from the symmetry property (46). Indeed, because $f(\cdot|x)$ is by definition the density function of Y_{i+1} given $Y_i = x$, the density function of Y_i given $Y_{i+1} = x$ is also defined by

$$\begin{aligned} f_{Y_i|Y_{i+1}}(y|x) &= \frac{f_{Y_i,Y_{i+1}}(y,x)}{q(x)} \\ &= \frac{f(y|x)q(x)}{q(x)} \quad \text{thanks to (46)} \\ &= f(y|x). \end{aligned}$$

If $0 < \pi < 1$, the ergodicity of this process follows directly from Proposition 9.3.24 of Florens, Mouchard, and Rolin (1990) because in this case the joint probability distribution of (Y_i, Y_{i+1}) is equivalent to the product of their marginal distributions and then the Doeblin condition is satisfied.

The operator $\mathcal{T}$ is then defined by

$$\begin{aligned} \mathcal{T}(\varphi)(x) &= E[\varphi(Y_{i+1})|Y_i = x] \\ &= \mathbb{I}_{[x>0]}\{\pi E[\varphi(Y_{i+1})|Y_{i+1} > 0] + (1-\pi)E[\varphi(Y_{i+1})|Y_{i+1} \le 0]\} \\ &\quad + \mathbb{I}_{[x\le 0]}\{(1-\pi)E[\varphi(Y_{i+1})|Y_{i+1} > 0] + \pi E[\varphi(Y_{i+1})|Y_{i+1} \le 0]\}. \end{aligned}$$

An eigenvector associated to a nonzero eigenvalue λ is then necessarily of the form

$$\varphi(y) = a\mathbb{I}_{[y>0]} + b\mathbb{I}_{[y\le 0]}. \tag{47}$$

In the following, we show that the discrete-time process Y is embeddable by a continuous-time process. Notice that the discrete-time process Y is not embeddable by any reversible scalar diffusion. To see this we refer to Hansen, Scheinkman, and Touzi (1995), who proved that any function lying in the domain of the infinitesimal generator of a reversible scalar diffusion is continuously differentiable. Because

$$\begin{aligned} \mathcal{T}(\mathbb{I}_{[y>0]})(x) &= \pi\mathbb{I}_{[x>0]} + (1-\pi)\mathbb{I}_{[x\le 0]}, \\ \mathcal{T}(\mathbb{I}_{[y\le 0]})(x) &= (1-\pi)\mathbb{I}_{[x>0]} + \pi\mathbb{I}_{[x\le 0]}, \end{aligned}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

it is clear that any nonzero eigenvalue λ associated to an eigenvector defined by a and b satisfies

$$\begin{cases} \pi a + (1-\pi)b = \lambda a \\ (1-\pi)a + \pi b = \lambda b. \end{cases}$$

By adding these two equations we get

$$a + b = \lambda(a + b).$$

Therefore

Either $\lambda = 1$ and then $b = a$
Or $a + b = 0$ and then $\lambda = 2\pi - 1$.

In other words, $\lambda_0 = 1$ and $\lambda_1 = 2\pi - 1$ are the two nonzero eigenvalues (if $\pi \neq \frac{1}{2}$) with associated normalized eigenvectors $\varepsilon_0(y) = 1$ and $\varepsilon_1(y) = \mathbb{I}_{[y>0]} - \mathbb{I}_{[y\leq 0]}$. Moreover, if φ is orthogonal to both ε_0 and ε_1, φ is orthogonal to both $\mathbb{I}_{[y>0]}$ and $\mathbb{I}_{[y\leq 0]}$, and then $\mathcal{T}\varphi = 0$. This proves that 0 is an eigenvalue associated to the eigenspace orthogonal to ε_0 and ε_1.

Now suppose that $\lambda_1 = 2\pi - 1 > 0$. Then the discrete-time process is clearly embeddable by the following continuous-time process:

$$\mathcal{T}_t(\varphi)(x) = \mathbb{I}_{[x>0]}\left\{\frac{1+\lambda_1^t}{2}E[\varphi(Y_t)|Y_t > 0] + \frac{1-\lambda_1^t}{2}E[\varphi(Y_t)|Y_t < 0]\right\} + \mathbb{I}_{[x\leq 0]}\left\{\frac{1-\lambda_1^t}{2}E[\varphi(Y_t)|Y_t > 0] + \frac{1+\lambda_1^t}{2}E[\varphi(Y_t)|Y_t < 0]\right\}.$$

However, such a continuous-time process is not strongly continuous. Indeed considering any function φ orthogonal to the eigenfunctions ε_0 and ε_1, we see that $\mathcal{T}_t\varphi = 0$ for any $t > 0$ and therefore $\mathcal{T}_t\varphi$ does not converge to φ as t goes to zero.

Hence, in the case $2\pi - 1 > 0$, although our necessary embeddability condition $E[\varphi(Y_i)\varphi(Y_{i+1})] \geq 0$ is satisfied, the underlying continuous-time process does not satisfy the strong continuity condition assumed in Section 3 to define the infinitesimal generator.

5.2. The Case of Stationary Scalar Diffusions

Stationary scalar diffusions are widely used in financial applications. Under mild regularity conditions they are natural candidates in the class of stationary reversible continuous-time Markov processes. Therefore, if H_{0p} is rejected for given data (see Section 6) then such processes can not be used to model the dynamics of the data. This section recalls some known results on stationary scalar diffusions.

Let $\{W_t, t \geq 0\}$ be a scalar Brownian motion on $(\Omega, \mathcal{F}, P)$ and consider the scalar diffusion process $\{Y_t, t \geq 0\}$ satisfying the stochastic differential equation

$$dY_t = \mu(Y_t)\,dt + \sigma(Y_t)dW_t \qquad \textbf{(48)}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

in the domain $\{l, r\}$, which can be bounded or unbounded, closed, semiclosed, or open. In the case where one or both boundaries are in the state space of the process, we shall consider it as a reflecting boundary. There are many results in the literature that establish the existence and uniqueness of a Markov process $\{Y_t, t \geq 0\}$ satisfying (48). One set of sufficient conditions requires that the diffusion coefficient σ^2 be strictly positive with a bounded and continuous second derivative and that the local mean μ have a bounded and continuous first derivative. These conditions also imply that $\{Y_t, t \geq 0\}$ is a Feller process with continuous sample path.

To ensure the existence of a stationary measure, we assume that the speed measure function is integrable, i.e.,

$$\int_l^r m(y)\, dy < \infty, \tag{49}$$

where

$$m(y) = \frac{1}{\sigma^2(y)} \exp\left[\int_{y_0}^{y} \frac{2\mu(x)}{\sigma^2(x)}\, dx\right], \qquad l < y, y_0 < r. \tag{50}$$

Under this assumption, if we use the measure Q with density

$$q(y) = \frac{m(y)}{\int_l^r m(x)\, dx}, \qquad l < y < r \tag{51}$$

to initialize the process, then the Markov process $\{Y_t, t \geq 0\}$ generated via the stochastic differential equation (48) is stationary (see Karlin and Taylor, 1981, p. 241).

Next define the infinitesimal generator of the stationary scalar diffusion $\{Y_t, t \geq 0\}$ in the Hilbert space $L^2(Q)$ as in Section 3. Under mild regularity conditions, the set of twice-continuously differentiable functions with compact support C_K^2 is a core for the infinitesimal generator $\mathcal{A}$ (see Ethier and Kurtz, 1986; Hansen and Scheinkman, 1994). Now considering the C_K^2 functions φ and ψ, the reversibility condition $\langle \mathcal{A}\varphi, \psi \rangle = \langle \varphi, \mathcal{A}\psi \rangle$ is easily verified by a simple integration by parts. Therefore the conditions ensuring that C_K^2 is a core for $\mathcal{A}$ also guarantee that $\mathcal{A}$ is self-adjoint and the process $\{Y_t, t \geq 0\}$ is reversible.

Given a discrete-time process, the set of moment restrictions (19) is a necessary condition of embeddability by stationary scalar diffusions. Thus if H_{0p} is rejected, then embeddability by stationary scalar diffusions is rejected. However such a necessary condition of embeddability by stationary scalar diffusions is far from being sufficient because it is easy to construct discrete-time processes embeddable by a stationary reversible continuous-time Markov process that is not a scalar diffusion (see the following section).

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

Further testable conditions of embeddability by stationary scalar diffusions can be derived from the spectral properties of stationary scalar diffusions. Indeed, if the top of the spectrum of $\mathcal{A}$ is isolated, then it is possible to estimate the associated eigenfunctions. However stationary scalar diffusions are completely identified from the stationary density and an eigenelement whose eigenvalue is the closest to zero (see Hansen et al., 1995). Therefore one further testable embeddability condition is the following. The estimation of the latter eigenelements and the stationary distribution implies a precise (theoretical) spectral decomposition of the infinitesimal generator if the discrete-time process is embeddable by a stationary scalar diffusion. The eigenelements corresponding to the discrete part at the top of the spectrum can be estimated directly from the data as in Hansen et al. (1995). The embeddability test consists in an adequacy test between the estimated eigenelements and the theoretical ones implied by the estimation of the stationary scalar diffusion.

5.3. To Be (a Diffusion) or Not to Be

This section provides an example of discrete-time process embeddable by a strongly continuous stationary ergodic Markov process that is not a diffusion.

Let $Q = \mathcal{N}(0, \tau^2)$ be the marginal probability and $(\varepsilon_j)_{j\in\mathbb{N}}$ the orthonormal basis of Hermite polynomials ($\varepsilon_0 = 1$) of $L^2(Q)$.

Let us consider (ρ_1, σ_1^2) and (ρ_1, σ_2^2) such that

$$0 < \rho_i < 1, \qquad i = 1,2, \quad \text{and} \quad \frac{\sigma_1^2}{1-\rho_1^2} = \frac{\sigma_2^2}{1-\rho_2^2} = \tau^2 \tag{52}$$

and define the conditional probability distribution given x by the mixture

$$\pi\mathcal{N}[\rho_1 x, \sigma_1^2] + (1-\pi)\mathcal{N}[\rho_2 x, \sigma_2^2].$$

Because the operator associated to the normal transition $\mathcal{N}[\rho_i x, \sigma_i^2]$ is

$$\mathcal{T}^i(\varphi) = \sum_{j=0}^{\infty} \rho_i^j E[\varphi\varepsilon_j]\varepsilon_j$$

(notice that, thanks to (52), $\mathcal{T}^1$ and $\mathcal{T}^2$ have the same eigenvectors), the operator $\mathcal{T}$ associated to the mixture is

$$\mathcal{T}(\varphi) = \sum_{j=0}^{\infty} [\pi\rho_1^j + (1-\pi)\rho_2^j] E(\varphi\varepsilon_j)\varepsilon_j.$$

In other words, $(\varepsilon_j)_{j\in\mathbb{N}}$ are also the eigenvectors of $\mathcal{T}$ and $(\pi\rho_1^j + (1-\pi)\rho_2^j)_{j\in\mathbb{N}}$ are the corresponding eigenvalues. They are positive, and the discrete-time process is indeed embeddable by the continuous-time process defined by the semigroup

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

$$\mathcal{T}_t(\varphi) = \sum_{j=0}^{\infty} (\pi\rho_1^j + (1-\pi)\rho_2^j)^t E(\varphi\varepsilon_j)\varepsilon_j.$$

This semigroup is strongly continuous, and the associated infinitesimal generator is

$$\mathcal{A}(\varphi) = \sum_{j=1}^{\infty} \mathrm{Log}[\pi\rho_1^j + (1-\pi)\rho_2^j] E(\varphi\varepsilon_j)\varepsilon_j. \tag{53}$$

Now we can claim that such a continuous-time process is not a scalar diffusion process. As is shown in Appendix B, the only scalar diffusion process whose eigenvectors are the Hermite polynomials associated to the marginal distribution $Q = \mathcal{N}(0, \tau^2)$ is the Ornstein–Uhlenbeck process. But the eigenvalues associated to the Ornstein–Uhlenbeck process are of the form ρ^j, which is inconsistent with (53) when π is not 0 or 1.

6. EMPIRICAL IMPLEMENTATION FOR FINANCIAL DATA

Financial applications provide us with an interesting setup for our testing procedure. In the financial literature there are "basically" two classes of continuous-time models that are considered:

(i) Stationary scalar diffusions are very popular because they produce closed form solutions for pricing and hedging strategies; even if closed form solutions are not available, numerical methods are easy to implement and provide fast solutions,
(ii) Bivariate diffusions are known as stochastic volatility models and can be seen as the continuous-time limit of (EG)ARCH models, which are known to fit financial data very well (Nelson, 1990). However, such models are more difficult to use in practice because closed form solutions are more difficult to derive and the usual numerical methods turn out to be time consuming.

As recalled in Section 5.2, stationary scalar diffusions are time reversible under minor regularity conditions. This result does not hold in general for multivariate diffusions. Therefore, in contrast with scalar diffusions, multivariate diffusions are not in the class of stationary reversible Markov processes, and the set of moments restrictions (19) allows us to discriminate between these two models. Hence our testing procedure provides a very simple tool to check whether given data can be modeled by a stationary scalar diffusion. As argued in Section 5.2, if the embeddability hypothesis (19) is not rejected, more restrictions should be tested to be able to accept the stationary scalar diffusion modeling. However if the embeddability hypothesis is rejected then modeling by stationary scalar diffusions is rejected.

We perform the test with $p = 5$ elements of $L^2(Q)$:[3]

$$\varphi_j = x^j - E(x^j), \qquad j = 1, \ldots, 5,$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

TABLE 1. Stock index data

	$\sqrt{n}\hat{\lambda}_p/\hat{\sigma}$	Decision (5% level)
JP Morgan France	−0.55	—
JP Morgan World	−1.48	—
Morgan Stanley France	−1.11	—
Morgan Stanley World	−3.05	reject
CAC40 (France)	−1.59	—
DAX (Germany)	−1.72	reject
NIKKEI (Japan)	−1.53	—
S&P (USA)	−2.45	reject

and the variance matrix Σ of (33) is estimated as indicated in Section 4 using the Bartlett kernel (43). The rejection limit to the 5% level is 1.65.

The test procedure is implemented for stock index data: 4 international stock indices (4 series of daily observations from October 1989 to November 1994) and 4 stock indices for the major markets (4 series of daily observations from September 1987 to June 1991) as indicated in Table 1. The embeddability hypothesis is only rejected for the Morgan Stanley World, the DAX, and the S&P indices data and is close to rejection for the CAC40 index (to the 5% level).

Next the test procedure is implemented for exchange rate data for the 15 EEC countries vs. the French franc (14 series of daily observations from October 1989 to November 1994) as indicated in Table 2. The embeddability hypothesis is now rejected for all but 2 series (Finland and United Kingdom) (to the 5% level).

TABLE 2. Exchange rate data (vs. French franc)

	$\sqrt{n}\hat{\lambda}_p/\hat{\sigma}$	Decision (5% level)
Europe (ECU)	−2.28	reject
Germany	−2.47	reject
Belgium	−2.45	reject
Italy	−3.77	reject
Switzerland	−1.97	reject
Netherlands	−4.61	reject
Spain	−2.17	reject
United Kingdom	−0.92	—
Ireland	−2.57	reject
Norway	−2.02	reject
Sweden	−4.19	reject
Finland	−1.32	—

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

These findings are consistent with the well-known observation that exchange rate volatility data are more volatile than stock index data and therefore should be modeled by multifactor models.

NOTES

1. In discrete time, this is the debate GARCH vs. stochastic volatility modeling. The evidence of mixture of distributions that supports the stochastic volatility paradigm for exchange rates was already well documented. See, e.g., Ghysels, Harvey, and Renault (1996), Kim, Shephard, and Chib (1996), and Shephard (1996).

2. Let $W_\alpha(\cdot)$ denote the critical region of a given test procedure at the level α and $\overline{W}_\alpha(\cdot)$ the corresponding acceptance region. Then under H_0

$$\begin{aligned} P[W_\alpha(H_{0p}) \cup W_\alpha(H_0(p+1))] &= 1 - P[\overline{W}_\alpha(H_{0p})]P[\overline{W}_\alpha(H_0(p+1))|\overline{W}_\alpha(H_{0p})] \\ &= 1 - (1-\alpha)P[\overline{W}_\alpha(H_0(p+1))|\overline{W}_\alpha(H_{0p})] \\ &\geq \alpha; \end{aligned}$$

the inequality is generally strict because the test statistics of $H_0(p+1)$ and H_{0p} are generally not perfectly correlated.

3. Actually the functions $\varphi_j, j = 1,\ldots,5$ are assumed to be in $L^2(Q)$. A more rigorous approach would be to start by estimating the stationary distribution of the data and then considering a basis of the implied $L^2(Q)$ Hilbert space. Such a basis can be obtained by rescaling the process using the estimated stationary density to resort to the Hilbert space $L^2[0,1]$.

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Bergstrom, A.R. (1984) Continuous-time stochastic models and issues of aggregation over time. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. II, pp. 1145–1212. Amsterdam: Elsevier.

Billingsley, P. (1961) *Statistical Inference for Markov Processes.* Chicago: University of Chicago Press.

Chan, K.S. & H. Tong (1987) A note on embedding a discrete parameter ARMA model in a continuous parameter ARMA model. *Journal of Time Series Analysis* 8, 3.

Conley, T., L.P. Hansen, E. Luttmer, & J. Scheinkman (1995) Estimating Subordinated Diffusions from Discrete-Time Data. Preprint, University of Chicago.

Cox, D.R. & H.D. Miller (1965) *The Theory of Stochastic Processes.* London: Methun.

Doob, J.L. (1953) *Stochastic Processes.* New York: Wiley.

Dunford, N. & J.T. Schwartz (1963) *Linear Operators*, part II: *Spectral Theory.* New York: Wiley.

Ethier, S.N. & T.G. Kurtz (1986) *Markov Processes: Characterization and Convergence*. New York: John Wiley & Sons.

Florens, J.P., M. Mouchart, & J.M. Rolin (1990) *Elements of Bayesian Statistics.* New York: Marcel Dekker.

Fougère, D. & T. Kamionka (1992) Un modèle markovien du marché du travail. *Annales d'Economie et de Statistique* 27, 149–189.

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

Gandolfo, G. (1981) *Qualitative Analysis and Econometric Estimation of Continuous-Time Dynamic Models.* Amsterdam: North-Holland.
Geweke, J., R.C. Marshall, & G.A. Zarkin (1986) Mobility indices in continuous-time Markov chains. *Econometrica* 54, 1407–1423.
Ghysels, E., A.C. Harvey, & E. Renault (1996) Stochastic volatility. In G.S. Maddala & C.R. Rao (eds.), *Handbook of Statistics*, vol. 14: *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.
Hansen, L.P. & J.A. Scheinkman (1995) Back to the future: Generating moment implications for continuous-time Markov processes. *Econometrica* 63 (4), 767–804.
Hansen, L.P., J.A. Scheinkman, & N. Touzi (1995) Spectral methods for identifying scalar diffusions. *Journal of Econometrics* 86, 1–32.
He, S.W. & J.G. Wang (1989) On embedding a discrete-parameter ARMA model in a continuous-parameter ARMA model. *Journal of Time Series Analysis* 10 (4), 315–323.
Karlin, S. & H.M. Taylor (1981) *A Second Course in Stochastic Processes.* New York: Academic Press.
Kim, S., N. Shephard, & S. Chib (1996) Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies.* Forthcoming.
Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis.* Berlin: Springer-Verlag.
Nelson, D.B. (1990) ARCH models as diffusion approximations. *Journal of Econometrics* 45, 7–38.
Ogaki, M. (1993) Generalized method of moments: Econometric applications. In G.S. Maddala, C.R. Rao, & H.D. Vinod (eds.), *Handbook of Statistics*, vol. 11: *Econometrics.* Amsterdam: North-Holland.
Parzen, E. (1957) On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329–348.
Pazy, A. (1983) *Semigroups of Linear Operators and Applications to Partial Differential Equations.* Berlin: Springer-Verlag.
Rudin, W. (1991) *Functional Analysis.* New York: McGraw-Hill.
Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In D.R. Cox, O.E. Barndorff-Nielson, & D.V. Hinkley (eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1–67. London: Chapman and Hall.
Weidmann, J. (1987) *Spectral Theory of Ordinary Differential Operators*. Berlin: Springer-Verlag.

APPENDIX A

It is clearly sufficient to check that

$$E[\varphi(Y_1)\varphi(Y_2)] > 0$$

with a given function φ, $0 < \rho < 1$, and

$$Y_1 \rightsquigarrow \mathcal{N}(0,1)$$

$$Y_2 \mid Y_1 = y \rightsquigarrow \mathcal{N}(\rho y, 1-\rho^2).$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

In other words, we have to prove that the function H defined by

$$H(\rho) = \int_{\mathbb{R}^2} (2\pi)^{-1} \varphi(y_1)\varphi(y_2) \frac{1}{\sqrt{1-\rho^2}} \exp\left[-\frac{y_1^2}{2} - \frac{(y_2 - \rho y_1)^2}{2(1-\rho^2)} \right] dy_1 dy_2$$

is positive for $0 < \rho < 1$. Introducing

$$f_\rho(y) = (2\pi)^{-1/2} \varphi(y) \frac{1}{\sqrt{1-\rho^2}} \exp\left[-\frac{y^2}{2(1-\rho^2)} \right],$$

the function $H(\cdot)$ can be written

$$H(\rho) = \int_{\mathbb{R}^2} f_0(y_1) f_\rho(y_2) \exp\left[\frac{\rho y_1 y_2}{1-\rho^2} \right] \exp\left[-\frac{\rho^2 y_1^2}{2(1-\rho^2)} \right] dy_1 dy_2.$$

Now notice that H is clearly an analytical function on $[0, r)$ for any r, $0 < r < 1$ and therefore, for $0 < \rho < 1$ we have

$$H(\rho) = \sum_{p=0}^{\infty} H^{(p)}(0) \frac{\rho^p}{p!}.$$

To prove the required result, it is sufficient to check that

$$\begin{cases} H^{(p)}(0) \geq 0 & \forall p \in \mathbb{N}, \\ H^{(p)}(0) > 0 & \text{for some } p \in \mathbb{N}. \end{cases}$$

But,

$$\exp\left[\frac{\rho y_1 y_2}{1-\rho^2} \right] = \sum_{p=0}^{\infty} \frac{\rho^p}{(1-\rho^2)^p} y_1^p y_2^p,$$

and, thanks to Lebesgue's theorem of dominated convergence,

$$\begin{aligned} H(\rho) &= \sum_{p=0}^{\infty} \frac{\rho^p}{p!} \int_{\mathbb{R}^2} f_0(y_1) y_1^p \exp\left[-\frac{\rho^2 y_1^2}{2(1-\rho^2)} \right] \frac{f_\rho(y_2) y_2^p}{(1-\rho^2)^p} dy_1 dy_2 \\ &= \sum_{p=0}^{\infty} \frac{\rho^p}{p!} \left[\int_{\mathbb{R}} f_0(y_1) y_1^p \exp\left[-\frac{\rho^2 y_1^2}{2(1-\rho^2)} \right] dy_1 \right] \left[\int_{\mathbb{R}} \frac{f_\rho(y_2) y_2^p}{(1-\rho^2)^p} dy_2 \right]. \end{aligned} \tag{A.1}$$

Differentiating with respect to ρ, we get

$$H^{(p)}(\rho) = \left[\int_{\mathbb{R}} f_0(y_1) y_1^p \exp\left[-\frac{\rho^2 y_1^2}{2(1-\rho^2)} \right] dy_1 \right] \left[\int_{\mathbb{R}} \frac{f_\rho(y_2) y_2^p}{(1-\rho^2)^p} dy_2 \right] + \rho G_p(\rho).$$

It then follows that

$$\begin{aligned} H^{(p)}(0) &= \left[\int_{\mathbb{R}} f_0(y_1) y_1^p dy_1 \right] \left[\int_{\mathbb{R}} f_0(y_2) y_2^p dy_2 \right] \\ &= \left[\int_{\mathbb{R}} f_0(y_1) y_1^p \, dy_1 \right]^2 \geq 0. \end{aligned} \tag{A.2}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

Finally suppose that $H^{(p)}(0) = 0\ \forall p \in \mathbb{N}$. Then φ is orthogonal to any $y^p, p \in \mathbb{N}$ in the sense of $\mathcal{L}^2(Q)$, which clearly implies $\varphi = 0$.

APPENDIX B

In this appendix we prove that the only scalar diffusion that admits the Hermite polynomials as eigenvectors is the Ornstein–Uhlenbeck process. The first (nondegenerate) Hermite polynomials associated to the normal distribution $\mathcal{N}(0, \tau^2)$ are proportional to the functions

$$\varphi_1(x) = x,$$

$$\varphi_2(x) = -x^2 + \tau^2,$$

$$\varphi_3(x) = -x^3 + 3\tau^2 x.$$

Consider a scalar diffusion with local mean $\mu(\cdot)$ and local variance $\sigma(\cdot)$ and suppose that $(\varphi_i, \lambda_i), i = 1 \ldots 3$ are eigenelements of the associated infinitesimal generator, i.e.,

$$\mathcal{A}\varphi_i(x) = \mu(x)\varphi'(x) + \tfrac{1}{2}\sigma^2(x)\varphi''(x) = \lambda_i \varphi(x), \qquad i = 1 \ldots 3, \quad \forall x.$$

Then, it is easily seen that the two first eigenelements provide μ and σ:

$$m(x) = \lambda_1 x \quad \text{and} \quad \sigma^2(x) = (\lambda_2 - 2\lambda_1)x^2 - \lambda_2 \tau^2.$$

It remains to check that (ϕ_3, λ_3) is an eigenelement of $\mathcal{A}$. Direct computation shows that this is possible if and only if

$$\lambda_2 = 2\lambda_1,$$

which proves the required result.

APPENDIX C

In this appendix we prove that the only scalar diffusion embedding a Gaussian AR(1) process is the Ornstein–Uhlenbeck process. In the framework (11), the stationary distribution of the Gaussian AR(1) process is given by

$$q(x) = \frac{1}{\gamma\sqrt{2\pi}} \exp - \frac{x^2}{2\gamma^2} \quad \text{with } \gamma^2 = \frac{\tau^2}{1 - \rho^2.}$$

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

As recalled in Section 5.2, the infinitesimal generator of any stationary scalar diffusion can be expressed in terms of the stationary density

$$\mathcal{A}\varphi(x) = \frac{1}{q(x)}\frac{\partial}{\partial x}\left\{\frac{1}{2}\sigma^2(x)q(x)\varphi'(x)\right\}, \quad \text{for any } \varphi \in D(\mathcal{A}).$$

Now because the Gaussian AR(1) process satisfies $E(Y_{t+1}|Y_t = y) = \rho y$, the function $\varphi_1(x) = x$ is an eigenvector of the conditional expectation operator $\mathcal{T}_1$ associated to the eigenvalue ρ. As recalled in Section 3, this shows that the function φ_1 is an eigenvector of the infinitesimal generator $\mathcal{A}$ associated to the eigenvalue $-k = \ln \rho$. Using the preceding expression of the infinitesimal generator, we get

$$\frac{\partial}{\partial x}\left\{\frac{1}{2}\sigma^2(x)q(x)\right\} = -kxq(x).$$

Now from Hansen et al. (1998) we have that $\lim_{x\to\infty}\sigma^2(x)q(x) = 0$, and therefore by integrating the last equation from $-\infty$ to some $x \in \mathbb{R}$ we have

$$\begin{aligned}\frac{1}{2}\sigma^2(x)q(x) &= -k\int_{-\infty}^{x} yq(y)\,dy \\ &= k\gamma^2 q(x), \end{aligned} \tag{C.1}$$

which proves that the local variance function is a constant:

$$\sigma^2(x) = 2k\gamma^2 = \frac{-2\tau^2 \ln \rho}{1-\rho^2}.$$

Direct computation shows that the local mean coefficient is a linear function.

APPENDIX D

In this appendix, we provide a rigorous derivation of the moments conditions (18).

(i) We first prove the required result for $\varphi \in \mathcal{D}(\mathcal{A})$. As a self-adjoint operator, the infinitesimal generator $\mathcal{A}$ admits a unique spectral representation in terms of a (unique) spectral family $\{E(\lambda), \lambda \in \mathbb{R}\}$. Therefore, for any $\varphi \in \mathcal{D}(\mathcal{A})$, we have

$$\varphi = \int_{-\infty}^{0} dE(\lambda)\varphi,$$

where the integration is confined to the nonpositive real line by virtue of (17). Furthermore, because $\{\mathcal{T}_t, t \geq 0\}$ is a strongly continuous semigroup of self-adjoint (and then normal) bounded operators, each $\mathcal{T}_t$ admits the spectral representation on $\mathcal{D}(\mathcal{A})$:

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press

$$\mathcal{T}_t\varphi = \int_{-\infty}^{0} e^{t\lambda}\, dE(\lambda)\varphi$$

(see Rudin, 1991, Theorem 13.38, p. 382). We then have

$$\langle \mathcal{T}_t\varphi, \varphi\rangle = \int_{-\infty}^{0} e^{t\lambda}\, \|dE(\lambda)\varphi\|^2 \geq 0.$$

(ii) To extend the result to any $\varphi \in L^2(Q)$, we use the fact that $\mathcal{D}(\mathcal{A})$ is dense in $L^2(Q)$ (see Pazy, 1983, Corollary 2.5, p. 5). Then let $(\varphi_n)_{n\in\mathbb{N}}$ be a sequence of elements of $\mathcal{D}(\mathcal{A})$ converging (in the sense of $L^2(Q)$) to some $\varphi \in L^2(Q)$. Then, by the Cauchy–Schwartz inequality and the fact that $\mathcal{T}_t$ is a contraction operator ($\|T_t\varphi\| \leq \|\varphi\|$), we see that

$$\begin{aligned}|\langle \mathcal{T}_t\varphi, \varphi\rangle - \langle \mathcal{T}_t\varphi_n, \varphi_n\rangle| &= |\langle T_t\varphi, \varphi - \varphi_n\rangle + \langle T_t(\varphi - \varphi_n), \varphi_n\rangle| \\ &\leq \|\varphi - \varphi_n\|(\|\varphi\| + \|\varphi_n\|).\end{aligned}$$

By sending n to infinity, we see that the right-hand-side term of the last inequality converges to zero because $\varphi_n \to \varphi$ in the sense of $L^2(Q)$ and $(\varphi_n)_{n\in\mathbb{N}}$ is bounded in $L^2(Q)$ as a converging sequence. This provides the required result.

https://doi.org/10.1017/S0266466698146029 Published online by Cambridge University Press