TECHNICAL WORKING PAPER SERIES

BACK TO THE FUTURE: GENERATING MOMENT IMPLICATIONS FOR CONTINUOUS-TIME MARKOV PROCESSES

Lars Peter Hansen
José Alexandre Scheinkman

Technical Working Paper No. 141

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1993

Conversations with Buz Brock, Henri Berestycki, Darrell Duffie, Ivar Ekeland, Hedi Kallal, Carlos Kening, Jean Michel Lasry, Pierre Louis Lions, Andy Lo, Erzo Luttmer, Jesus Santos and Arnold Zellner are gratefully acknowledged. We received helpful comments from Peter Robinson and three anonymous referees on an earlier version of this paper. Thanks to Andy Lo and Steven Spielberg for suggesting the title. Portions of this research were funded by grants from the National Science Foundation. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #141
September 1993

BACK TO THE FUTURE: GENERATING MOMENT IMPLICATIONS FOR CONTINUOUS-TIME MARKOV PROCESSES

ABSTRACT

Continuous-time Markov processes can be characterized conveniently by their infinitesimal generators. For such processes there exist forward and reverse-time generators. We show how to use these generators to construct moment conditions implied by stationary Markov processes. Generalized method of moments estimators and tests can be constructed using these moment conditions. The resulting econometric methods are designed to be applied to discrete-time data obtained by sampling continuous-time Markov processes.

Lars Peter Hansen
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NORC and NBER

José Alexandre Scheinkman
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637

1. Introduction

In this paper we derive moment conditions for estimating and testing continuous-time Markov models using discrete-time data. An extensive literature exists on estimating continuous time *linear* models from discrete-time data emanating from the work of A. W. Phillips (1959). This literature includes treatments of identification (*e.g.*, see P.C.B. Phillips 1973; Hansen and Sargent 1983) as well as estimation (*e.g.*, see P.C.B. Phillips 1973; Robinson 1976; Harvey and Stock 1985). Our aim is to develop new methods for estimation and inference that can be applied to continuous time *nonlinear* Markov models, again from the vantage point of discrete-time sampling. Recently there has been a considerable interest among economists in understanding the role nonlinearities in dynamic models (see Scheinkman 1990 for a survey of this literature). Furthermore, several particular nonlinear continuous-time models have been proposed for the term structure of interest rates (*e.g.*, see Cox, Ingersoll and Ross 1985; Heath, Jarrow and Morton 1990) and for exchange rates (*e.g.*, see Froot and Obstfeld 1991; and Krugman 1991). Among other things, we develop tools for assessing the empirical plausibility of these models.

Likelihood-based methods of estimation and inference for nonlinear continuous time models can be very difficult to implement due to the computational costs associated with evaluating the likelihood function (*e.g.*, see Lo 1988). This is true even when the Markov state vector is completely observable at any point in time, as we assume here. The reason for this difficulty is that a continuous-time Markov process is typically specified in terms of its local evolution. Evaluating the discrete-time transition density can be quite costly because it may require solving numerically a partial differential equation as in the case of a diffusion. The computational costs can become excessive because these calculations must be repeated for each hypothetical parameter value and each observed state. [1]

[1] Exceptions to this is the nonlinear Markov process assumed by Cox, Ingersoll and Ross (1985) in their analysis of the term structure of interest rates and the reflecting barrier model of exchange

In this paper we adopt a more pragmatic approach. We begin by considering a Markov process specified in terms of its *infinitesimal generator*. Formally, this generator is defined as an operator on a function space, and, in effect, this operator stipulates the local evolution of the process. For instance, for a diffusion model specified as a solution to a stochastic differential equation, the generator can be constructed from the coefficients of the differential equation and the associated boundary conditions.

Given the infinitesimal generator, we show how to construct two sets of moment conditions that are often easy to compute in practice. As a consequence, both sets of moment conditions can be used to construct generalized-method-of-moments estimators of an unknown parameter vector and diagnostic tests. One set involves only the contemporaneous Markov state vector and hence uses only the marginal distribution of that vector. The second set includes functions of the state vector in two adjacent time periods, and hence, like the score vector from a likelihood function, exploits properties of the conditional distribution of the current state vector given the past.

Moment conditions in the second set are most conveniently represented in terms of the original generator as well as the *reverse-time generator* for a process running *backwards* in time. Although there exist general characterizations of reverse-time generators, the second set of moment conditions may be easiest to apply when the underlying Markov processes under consideration are (time) reversible. As a consequence, we use results in the probability theory literature to show that a potentially rich collection of models are reversible, including many multi-factor models of the term structure of interest rates that have been suggested in the literature.

Another strategy that has been proposed for estimating continuous time Markov processes is to use numerical methods to approximate moments. Duffie and Singleton (1993) suggested the use of simulation while He (1990) proposed the use of binomial ap-

rates assumed by Krugman (1991). The transition densities for these process have been fully characterized (*e.g.*, see Feller 1951, Wong 1964, and Levy 1993) However, small departures in the form of the nonlinearities make likelihood-based methods much more numerically intensive to implement.

proximations. An attractive feature of the Duffie-Singleton approach is that the Markov state vector does not have to be fully observed. However, for both of these approaches it may be difficult to account for the magnitude of the approximation error, and it may be numerically costly to ensure that the approximation error is small.

While an aim of our analysis is to reduce substantially the computational costs *vis-a-vis* the method of maximum likelihood for models indexed by a finite-dimensional parameter vector, we also study nonparametric identification via these moment conditions. It is well known that identifying continuous-time models from discrete-time data can be problematic because of what is known as the *aliasing phenomenon*: distinct continuous-time processes may look identical when sampled at regular time intervals. Using spectral representation theory for self-adjoint operators, we show that there is no aliasing problem when it is known that the continuous-time process is reversible. Although the generalized-method-of-moments approach described in our paper is designed to be computationally tractable, complete (nonparametric) identification of reversible processes is not possible using the moment conditions we derive. However, we do show that in the context of scalar diffusion models, the local mean (drift) and local variance (diffusion coefficient) can be identified up to a common scale factor using our approach.

The focus of this paper is on deriving moment conditions implied by infinitesimal generators and on characterizing the extent to which these moment conditions can discriminate among the members of a class of infinitesimal generators. In addition, we provide restrictions on infinitesimal generators that ensure that the Law of Large Numbers and Central Limit Theorem apply to a discrete-sampled process. Armed with these approximation results, we can apply the results in Hansen (1982) to justify formally estimation and inference using generalized method of moments. On the other hand, we do not address formally issues of statistical efficiency and nonparametric estimation and inference using the moment conditions we derive. Such issues are deferred to future work.

This paper is organized as follows. In Section 2 we review the mathematical construction of an infinitesimal generator for a continuous time Markov process and describe properties of the generator that are important for our analysis. In Section 3 we show how to use the infinitesimal generator to construct two families of moment conditions expressed in terms of the Markov state vector. Familiar examples of Markov processes together with their infinitesimal generators are presented in Section 4. We study the observable implications of each family of moment conditions in Section 5. For instance, we show that our first family of moment conditions can be used to distinguish alternative candidate generators that imply distinct marginal distribution for the Markov state vector. We also provide a characterization of the additional informational content provided by the second family of moment conditions obtained by reversing calendar time. Since the domains of the generators are sometimes difficult to characterize fully, in Section 6 we show how to reduce the family of test functions used in the moment conditions to include only ones for which the generator can be represented in a convenient manner. For the moment conditions to be of use in practice, we must be able to approximate expectations of functions of the state vector using a discrete-time moment analog. In Section 7 we present sufficient conditions for these approximations to be valid. To facilitate verification, these conditions are expressed as restrictions on the infinitesimal generator. These large-sample approximations can also be used to justify other estimation methods than the one described in this paper.

2. Infinitesimal Generators

In this section we give the mathematical basis for our analysis. The focal point is on the construction of the *infinitesimal generator* of a strictly stationary, continuous-time, n-dimensional, vector Markov process $\{x_t\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P}r)$.

Let Q be the probability measure induced on $\mathbf{R}^n$ by x_t (for any t), $\mathcal{L}^2(Q)$ be the space of all Borel measurable functions $\phi : \mathbf{R}^n \rightarrow \mathbf{R}$ such that

$$\int_{\mathbf{R}^n} \phi^2 dQ < \infty,$$

and $< \cdot \mid \cdot >$ and $\| \cdot \|$ be the usual inner product and norm on $\mathcal{L}^2(Q)$. Associated with the Markov process is a family of operators $\{\mathcal{T}_t : t \geq 0\}$ where for each $t \geq 0, \mathcal{T}_t$ is defined by:[2]

$$\mathcal{T}_t\phi(y) \equiv E[\phi(x_t) \mid x_0 = y]. \tag{2.1}$$

This family is known to satisfy several properties. For instance, it follows from Nelson (1958, Theorem 3.1) that for each $t \geq 0$,

Property P1: $\mathcal{T}_t : \mathcal{L}^2(Q) \rightarrow \mathcal{L}^2(Q)$ is well defined, *i.e.* if $\phi = \psi$ with Q probability one, then $\mathcal{T}_t\phi = \mathcal{T}_t\psi$ with Q probability one and for each $\phi \in \mathcal{L}^2(Q), \mathcal{T}_t\phi \in \mathcal{L}^2(Q)$.

Property P2: $\| \mathcal{T}_t\phi \| \leq \| \phi \|$ for all $\phi \in \mathcal{L}^2(Q)$, *i.e.* $\mathcal{T}_t$ is a (weak) contraction.

Furthermore,

Property P3: For any $s, t \geq 0$, $\mathcal{T}_{t+s} = \mathcal{T}_t\mathcal{T}_s$, *i.e.* $\{\mathcal{T}_t : t \geq 0\}$ is a (one-parameter) semigroup .

Property *P2* is the familiar result that the conditional expectation operator can only reduce the second moment of a random variable. Property *P3* is implied by the Law of Iterated Expectations since the expectation of $\phi(x_{t+s})$ given x_0 can be computed by first conditioning on information available at time s.

Our approach to exploring the implications of continuous-time Markov models for discrete time data is to study limits of expectations over small increments of time. In order for this approach to work, we impose a *mild* restriction on the smoothness properties of $\{\mathcal{T}_t\}$.

[2]Throughout this paper we follow the usual convention of not distinguishing between an equivalence class and the functions in the equivalence class. Moreover, in (2.1) we are abusing notation in a familiar way, *e.g.* see Chung (1974, pages 299 and 230).

Assumption A1: For each $\phi \in \mathcal{L}^2(Q)$, $\{\mathcal{T}_t\phi : t \geq 0\}$ converges [in $\mathcal{L}^2(Q)$] to ϕ for all $\phi \in \mathcal{L}^2(Q)$ as $t \downarrow 0$.

Assumption *A1* is weak in the sense that it is implied by measurability properties of the underlying stochastic process $\{x_t\}$. Recall that $\{x_t\}$ can always be viewed as a mapping from $\mathbf{R} \times \Omega$ into $\mathbf{R}^n$. Form the product sigma algebra using the Borelians of R and the events of Ω. A sufficient condition for Assumption *A1* is that the mapping defined by the stochastic process be Borel measurable with respect to the product sigma algebra. [3]

For some choices of $\phi \in \mathcal{L}^2(Q)$, the family of operators is *differentiable* at zero, *i.e.* $\{[\mathcal{T}_t\phi - \phi]/t : t > 0\}$ has an $\mathcal{L}^2(Q)$ limit as t goes to 0. Whenever this limit exists, we denote the limit point $\mathcal{A}\phi$. We refer to $\mathcal{A}$ as the *infinitesimal generator*. The domain $\mathcal{D}$ of this generator is the family of functions ϕ in $\mathcal{L}^2(Q)$ for which $\mathcal{A}\phi$ is well defined. Typically, $\mathcal{D}$ is a proper subset of $\mathcal{L}^2(Q)$. Since $\mathcal{A}$ is the *derivative* of $\{\mathcal{T}_t : t \geq 0\}$ at $t = 0$, and $\{\mathcal{T}_t : t \geq 0\}$ is a semigroup, $\mathcal{A}$ and $\mathcal{T}_t$ commute on $\mathcal{D}$. Moreover the following properties are satisfied (*e.g.* see Pazy 1983, Theorem 2.4, page 4):

Property P4: For any $\phi \in \mathcal{L}^2(Q)$, $\int_0^t \mathcal{T}_s\phi ds \in \mathcal{D}$ and $\mathcal{A}\int_0^t \mathcal{T}_s\phi ds = \mathcal{T}_t\phi - \phi$.

Property P5: For any $\phi \in \mathcal{D}$, $\mathcal{T}_t\phi - \phi = \int_0^t \mathcal{A}[\mathcal{T}_s\phi]ds = \int_0^t \mathcal{T}_s[\mathcal{A}\phi]ds$.

Property *P4* gives the operator counterpart to the familiar relation between derivatives and integrals. Property *P5* extends the formula by interchanging the order of integration and applications of the operators $\mathcal{T}_s$ and $\mathcal{A}$.

There are three additional well known properties of infinitesimal generators of continuous-time Markov operators that we will use in our analysis:

Property P6: $\mathcal{D}$ is dense in $\mathcal{L}^2(Q)$.

[3] It follows from Halmos (1974, page 148) that $\mathcal{T}_t\phi : t \geq 0$ is weakly measurable for any $\phi \in \mathcal{L}^2(Q)$ and hence from Theorem 1.5 of Dynkin (1965, page 35) that *A1* is satisfied.

Property P7: $\mathcal{A}$ is a closed linear operator, *i.e.* if $\{\phi_n\}$ in $\mathcal{D}$ converges to ϕ_o and $\{\mathcal{A}\phi_n\}$ converges to ψ_o, then ϕ_o is in $\mathcal{D}$ and $\mathcal{A}\phi_o = \psi_o$.

Property P8: For every $\lambda > 0$, $\lambda I - \mathcal{A}$ is onto.

(A reference for *P6* and *P7* and *P8* is Pazy 1983, Chapter 1, Corollary 2.5 and Theorem 4.3).

A final property of $\mathcal{A}$ that will be of value to us is:

Property P9: $< \phi \mid \mathcal{A}\phi > \leq 0$ for all $\phi \in \mathcal{D}$, *i.e.* $\mathcal{A}$ is quasi negative semidefinite.[4]

This property follows from the (weak) contraction property of $\{\mathcal{T}_t\}$ and the Cauchy-Schwarz Inequality since for any $t > 0$,

$$< \phi \mid \mathcal{T}_t\phi - \phi > \leq \| \phi \| (\| \mathcal{T}_t\phi \| - \| \phi \|) \leq 0. \tag{2.2}$$

In modeling Markov processes one may start with a candidate infinitesimal generator satisfying a particular set of properties, and then show that there exists an associated Markov process (*e.g.*, see Corollary 2.8 of Ethier and Kurtz, 1986, page 170). Furthermore, combinations of the above properties are sufficient for $\mathcal{A}$ to be the infinitesimal generator of a semi-group of contractions satisfying *A1*. For instance, it suffices that *P6* and *P9* holds as well as *P8* for *some* $\lambda > 0$. This is part of the Lumer-Phillips Theorem (*e.g.*, see Pazy 1983, Chapter 1, Theorem 4.3).

3. Moment Conditions

In this section we characterize two sets of moment conditions that will be of central interest for our analysis. These moment conditions are derived from two important

[4] When a generator of a semigroup defined on an arbitrary Banach space satisfies the appropriate generalization of *P9*, it is referred to as being *dissipative* (*e.g.*, see Pazy 1983, page 13).

(and well known) relations. The first relation links the stationary distribution Q and the infinitesimal generator $\mathcal{A}$, and the second one exploits the fact that $\mathcal{A}$ and $\mathcal{T}_t$ commute. Much of our analysis will focus on a fixed sampling interval which we take to be one. From now on we write $\mathcal{T}$ instead of $\mathcal{T}_1$.

Since the process $\{x_t\}$ is stationary, $E[\phi(x_t)]$ is independent of calendar time t implying that its derivative respect to t is zero. To see how this logic can be translated into a set of moment conditions, note that by the Law of Iterated Expectations, the expectation of $\phi(x_t)$ can either be computed directly or by first conditioning on x_0:

$$\int_{\mathbf{R}^n} \phi dQ = \int_{\mathbf{R}^n} \mathcal{T}_t \phi dQ \text{ for all} \phi \in \mathcal{L}^2(Q). \tag{3.1}$$

Hence for any $\phi \in \mathcal{D}$, we have that

$$\int_{\mathbf{R}^n} \mathcal{A}\phi dQ = \lim_{t \downarrow 0}(1/t) \int_{\mathbf{R}^n} [\mathcal{T}_t\phi - \phi] dQ = 0 \tag{3.2}$$

since (3.1) holds for all positive t and $\{(1/t)[\mathcal{T}_t\phi - \phi] : t > 0\}$ converges in $\mathcal{L}^2(Q)$ to $\mathcal{A}\phi$. Relation (3.2) shows the well known link between the generator $\mathcal{A}$ and the stationary distribution Q (*e.g.*, see Ethier and Kurtz 1986, Proposition 9.2, page 239). Hence our first set of moment conditions is

$C1 : E[\mathcal{A}\phi(x_t)] = 0$ for all $\phi \in \mathcal{D}$.

The stationarity of $\{x_t\}$ also implies that $E[\phi(x_{t+1})\psi(x_t)]$ does not depend on calendar time. Rather than exploiting this invariance directly and differentiating, a more convenient derivation is to start by using the fact that $\mathcal{A}$ and $\mathcal{T}$ commute, that is:

$$E[\mathcal{A}\phi(x_{t+1}) \mid x_t = y] = \mathcal{A}\{E[\phi(x_{t+1}) \mid x_t = y]\} \text{ for all } \phi \in \dot{\mathcal{D}}. \tag{3.3}$$

It may be difficult to evaluate the right side of (3.3) in practice because it entails computing the conditional expectation of $\phi(x_{t+1})$ prior to the application of $\mathcal{A}$. For this reason, we also derive an equivalent set of unconditional moment restrictions which are often easier to use. These moment restrictions are representable using the semigroup and generator associated with the reverse-time Markov process.

The semigroup $\{\mathcal{T}_t^*\}$ associated with the reverse-time process is defined via:

$$\mathcal{T}_t^* \phi^*(y) \equiv E[\phi^*(x_0) \mid x_t = y]. \tag{3.4}$$

The family $\{\mathcal{T}_t^* : t \geq 0\}$ is also a contraction semigroup of operators satisfying continuity restriction A1. Let $\mathcal{A}^*$ denote the infinitesimal generator for this semigroup with domain $\mathcal{D}^*$. The operator $\mathcal{T}_t^*$ is the adjoint of $\mathcal{T}_t$ and $\mathcal{A}^*$ is the operator adjoint of $\mathcal{A}$. To verify these results, note that it follows from (2.1), (3.4) and the Law of Iterated Expectations that

$$\begin{aligned} < \phi^* \mid \mathcal{T}_t\phi > &= E\{\phi^*(x_0)E[\phi(x_t) \mid x_0]\} \\ &= E[\phi^*(x_0)\phi(x_t)] \\ &= E\{\phi(x_t)E[\phi^*(x_0) \mid x_t]\} \\ &= < \phi \mid \mathcal{T}_t^*\phi^* > . \end{aligned} \tag{3.5}$$

We can now use Corollary 10.6 of Pazy (1983, page 41) to show that $\mathcal{A}^*$ is the operator adjoint of $\mathcal{A}$ and *vice versa*.

For any ϕ in the domain $\mathcal{D}$ of $\mathcal{A}$ and ϕ^* in $\mathcal{L}^2(Q)$, the fact that $\mathcal{A}$ and $\mathcal{T}$ commute implies that

$$< \mathcal{T}\mathcal{A}\phi \mid \phi^* > - < \mathcal{A}\mathcal{T}\phi \mid \phi^* > = 0. \tag{3.6}$$

When ϕ^* is restricted to be in the domain $\mathcal{D}^*$ of $\mathcal{A}^*$,

$$< \mathcal{A}\mathcal{T}\phi \mid \phi^* > = < \phi \mid \mathcal{T}^*\mathcal{A}^*\phi^* > . \tag{3.7}$$

Substituting (3.6) into (3.7) and applying the Law of Iterated Expectations we find that

$C2 : E[\mathcal{A}\phi(x_{t+1})\phi^*(x_t) - \phi(x_{t+1})\mathcal{A}^*\phi^*(x_t)] = 0$ for all $\phi \in \mathcal{D}$ and $\phi^* \in \mathcal{D}^*$.

Since $\mathcal{A}^*$ enters $C2$, these moment conditions exploit both the forward- and reverse-time characterization of the Markov process. As we hinted above, these moment conditions can be interpreted as resulting from equating the time derivative of $E[\phi(x_{t+1})\phi^*(x_t)]$ to zero.

It turns out that $C2$ implies $C1$. To see this let ψ be a constant function and note that $\mathcal{A}^*\psi$ is identically zero. However, the set of moment conditions $C1$ is sometimes easier to exploit in practice because it does not require computation of the adjoint $\mathcal{A}^*$. As we will see in the next section, calculating the adjoint $\mathcal{A}^*$ may require computing the stationary distribution Q implied by $\mathcal{A}$. For this reason, portions of our analysis focus on the role of $C1$ in identifying and estimating the infinitesimal generator $\mathcal{A}$. The more extensive set of moment conditions $C2$ are still of interest for two reasons. First, for many scalar diffusion processes scalar processes, $\mathcal{A}$ and $\mathcal{A}^*$ are the same (see Section 4), and thus empirical implementation of $C2$ is no more difficult than implementation of $C1$. Second, given the generator $\mathcal{A}$, it is typically easier to compute the adjoint $\mathcal{A}^*$ than the distribution of x_{t+1} conditioned on x_t as is required for evaluating the likelihood function.

One strategy for using these moment conditions for estimation and inference is as follows. Suppose the problem confronting an econometrician is to determine which, if any, among a parameterized set of infinitesimal generators is compatible with a discrete-time sample of the process $\{x_t\}$. For instance, imagine that the aim is to estimate the "true parameter value", say β_o, associated with a parameterized family of generators $\mathcal{A}_\beta$ for β in some admissible parameter space. Vehicles for accomplishing this task are the sample counterparts to moment conditions $C1$ and $C2$. By selecting a finite number of test functions ϕ, the unknown parameter vector β_o can be estimated using generalized method of moments (*e.g.*, see Hansen 1982) and the remaining over-identifying moment conditions can be tested. In Section 5 of this paper we characterize the information content of each of these two sets of moment conditions for discriminating among alternative sets of infinitesimal generators; and in Section 7 we supply some supporting analysis for generalized method of moments estimation and inference by deriving some sufficient conditions on the infinitesimal generators for the Law of Large Numbers and

Central Limit Theorem to apply.

4. Examples

In this section we give several illustrations of infinitesimal generators for continuous-time Markov processes.

Example 4.1 (Markov Jump Process): Let η be a nonnegative bounded function mapping $\mathbf{R}^n$ into $\mathbf{R}$ and $\pi(y,\Gamma)$ denote a transition function in the Cartesian product of $\mathbf{R}^n$ and the Borelians of $\mathbf{R}^n$. Imagine the following stochastic process $\{x_t\}$. Dates at which changes in states occur are determined by a Poisson process with parameter $\eta(y)$ if the current state is y. Given that a change occurs, the transition probabilities are given by $\pi(y,\cdot)$.

Additional restrictions must be imposed for this process to be stationary. First, suppose there exists a nonzero Borel measure $\tilde{Q}$ satisfying the equation:

$$\tilde{Q}(\Gamma) = \int \pi(y,\Gamma)d\tilde{Q}(y) \text{ for any Borelian } \Gamma.$$

Next suppose that

$$\int [1/\eta(y)]d\tilde{Q}(y) < \infty,$$

and construct the probability measure Q to be

$$dQ = \frac{\tilde{Q}}{\eta \int (1/\eta)d\tilde{Q}} \tag{4.1}$$

Under these restrictions on π and η, the Markov jump process $\{x_t\}$ will be stationary so long as it is initialized at Q.

Define the conditional expectation operator $\tilde{T}$ associated with the underlying Markov chain:

$$\tilde{T}\phi \equiv \int \phi(y')\pi(y,dy').$$

Analogous to the operator $\mathcal{T}, \tilde{\mathcal{T}}$ maps $\mathcal{L}^2(\tilde{Q})$ into itself. Using the fact that

$$\mathcal{T}_t\phi - \phi = t\eta\tilde{\mathcal{T}} - t\eta\phi + o(t)$$

one can show that the generator for the continuous time jump process can be represented as

$$\mathcal{A}\phi = \eta[\tilde{\mathcal{T}}\phi - \phi] \tag{4.2}$$

(*e.g.*, see Ethier and Kurtz 1986, pages 162-163). It is easy to verify that the generator $\mathcal{A}$ is a bounded operator on all of $\mathcal{L}^2(Q)$. Since η is bounded, any function ϕ in $\mathcal{L}^2(Q)$ must also be in $\mathcal{L}^2(\tilde{Q})$.

It is also of interest to characterize the adjoint $\mathcal{A}^*$ of $\mathcal{A}$. To do this we study the transition probabilities for the reverse-time process. Our candidate for $\mathcal{A}^*$ uses the adjoint $\tilde{\mathcal{T}}^*$ in place of $\tilde{\mathcal{T}}$:

$$\mathcal{A}^*\phi^* = \eta[\tilde{\mathcal{T}}^*\phi^* - \phi^*].$$

To verify that $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$, first note that

$$< \mathcal{A}\phi \mid \phi^* > = \int \eta[\tilde{\mathcal{T}}\phi - \phi]\phi^* dQ = \int \tilde{\mathcal{T}}\phi\phi^*\eta dQ - \int \eta\phi\phi^* dQ.$$

By construction $\tilde{\mathcal{T}}^*$ is the $\mathcal{L}^2(\tilde{Q})$ adjoint of $\tilde{\mathcal{T}}$. It follows from (4.1) that ηdQ is proportional to $d\tilde{Q}$. Consequently,

$$\begin{aligned} \int \tilde{\mathcal{T}}\phi\phi^*\eta dQ - \int \eta\phi\phi^* dQ &= \int \phi\tilde{\mathcal{T}}^*\phi^*\eta dQ - \int \eta\phi\phi^* dQ \\ &= < \phi \mid \mathcal{A}^*\phi^*) > . \end{aligned}$$

To use moment conditions $C1$ and $C2$ for this Markov jump process requires that we compute $\tilde{\mathcal{T}}\phi$ and $\tilde{\mathcal{T}}^*\phi$ for test functions ϕ. Suppose that the Markov chain is a discrete time Gaussian process. It is then straightforward to evaluate $\tilde{\mathcal{T}}\phi$ and $\tilde{\mathcal{T}}^*\phi$ for polynomial test functions. Nonlinearities in the continuous-time process could still be captured by nonlinear specification of the function η. On the other hand, when nonlinearities are introduced into the specification of the Markov chain, it may be difficult to compute

$\tilde{T}\phi$ and $\tilde{T}^*\phi$. In these cases our approach may not be any more tractable than, say, the method of maximum likelihood.

Recall that the moment conditions $C1$ and $C2$ can be used in situations in which the sampling interval is fixed and hence where the econometrician does not know the number of jumps that occurred between observations. This should be contrasted with econometric methods designed to exploit the duration time in each state.

Example 4.2 (Scalar Diffusion Process): Instead of specifying the infinitesimal generator directly, it is more common to start with a stochastic differential equation. As we will discuss below, there are well known connections between the coefficients of the stochastic differential equation and the infinitesimal generator.

Suppose that $\{x_t\}$ satisfies the stochastic differential equation:

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t \tag{4.3}$$

where $\{W_t : t \geq 0\}$ is a scalar Brownian motion. There are many results in the literature that establish the existence and uniqueness of a Markov process $\{x_t\}$ satisfying (4.3). 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 has a bounded and continuous first derivative. These conditions also imply that $\{x_t\}$ is a *Feller* process, that is, for any continuous function ϕ, $T\phi$ is continuous.[5]

We follow Karlin and Taylor (1981) and others by introducing a *scale function* S and its derivative:

$$s(y) \equiv \exp\{-\int^y 2[\mu(z)/\sigma^2(z)]dz\};$$

[5] For alternative Lipschitz and growth conditions see the hypotheses in Theorem 3.2, page 79 or the weaker local hypotheses of Theorem 4.1, page 84 of Has'minskii (1980). Alternatively, the Yamada-Watanabe Theorem in Karatzas and Shreve (1988, page 291) can be applied.

and a *speed density* $1/s\sigma^2$, which we assume to be integrable.[6] If we use the measure Q with density:

$$q(y) = \frac{1}{s(y)\sigma^2(y)\int[s(z)\sigma^2(z)]^{-1}dz} \tag{4.4}$$

to initialize the process, then the Markov process $\{x_t\}$ generated via the stochastic differential equation (4.3) will be stationary. In fact, under these hypothesis Q is the unique stationary distribution that can be associated with a solution of (4.3).

The infinitesimal generator is defined on a subspace of $\mathcal{L}^2(Q)$ that contains at least the subset of functions ϕ for which ϕ' and ϕ'' are continuous, $\sigma\phi' \in \mathcal{L}^2(Q)$, and such that the second-order differential operator:

$$L\phi(y) \equiv \mu(y)\phi'(y) + (1/2)\sigma^2(y)\phi''(y) \tag{4.5}$$

yields an $\mathcal{L}^2(Q)$ function. Ito's Lemma implies that $\{\phi(x_t) - \phi(x_0) - \int_0^t L\phi(x_s)ds\}$ is a continuous martingale. Take expectations conditional on x_0, to obtain:

$$E[\phi(x_t) \mid x_o] - \phi(x_0) = E[\int_0^t L\phi(x_s)ds \mid x_0]. \tag{4.6}$$

Or by Fubini's Theorem,

$$[\mathcal{T}_t\phi(x_0) - \phi(x_0)]/t = (1/t)\int_0^t \mathcal{T}_s L\phi(x_0)ds.$$

Using the continuity property of $\mathcal{T}_s$ and applying the Triangle Inequality we conclude that $\mathcal{A}\phi = L\phi$.

Under the conditions just given, stationary diffusions on the line are reversible. Following the usual approach of introducing the integrating factor $1/q$, we write:

$$\mathcal{A}\phi \;=\; (1/q)(\sigma^2 q\phi'/2)'. \tag{4.7}$$

[6] It suffices for integrability that for some strictly positive C and K and $|y| \geq K$, $\mu(y)y + \sigma^2/2 \leq -C$. This can be verified by using this inequality to construct an integrable upper bound for the speed density or by appealing to more general results in Has'minsky (1980).

Let C^2_{K}, denote the space of functions that are twice continuously differentiable with a compact support. Integration by parts implies that for functions ψ and ϕ in C^2_K,

$$< \mathcal{A}\phi \mid \psi > = < \phi \mid \mathcal{A}\psi > . \tag{4.8}$$

As we will discuss in Section 6, it is sufficient to verify (4.8) for ψ and ϕ in C^2_K. Consequently, $\mathcal{A}$ is symmetric; and since it is the generator of a contraction semigroup in a Hilbert space, it is self adjoint and $\{x_t\}$ is reversible (see Proposition VII.6 on page 113 of Brezis 1983).

Many economic examples deal with processes that are restricted to a finite interval or to the nonnegative reals. The reasoning above extends immediately to processes defined on an interval when both end points are entrance boundaries. Moreover, our analysis can be extended to processes with reflecting boundaries as the ones assumed in the literature on exchange rate bands. Typically, a process with reflecting boundaries in an interval (ℓ, u) is constructed by changing equation (4.3) to include an additional term that is "activated" at the boundary points:

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t + \theta(x_t)d\kappa_t$$

where $\{\kappa_t\}$ is a nondecreasing process that increases only when x_t is at the boundary, and

$$\theta(\ell) = 1 \text{ and } \theta(u) = -1.$$

In addition the original μ and σ^2 must define a *regular* diffusion, that is the functions s and $1/\sigma^2 s$ are integrable. We will also assume that s is bounded away from zero.[7] Ito's Lemma applies to such processes with an additional term $\phi'(x_t)\theta(x_t)d\kappa_t$. This term vanishes if $\phi'(\ell) = \phi'(u) = 0$, and $\mathcal{A}\phi$ is again given as (4.5) or (4.7) for ϕ's that satisfy these additional restrictions. To check that $\mathcal{A}$ is self adjoint, it is now sufficient to verify (4.8) for ψ and ϕ that are twice continuously differentiable with first derivatives that

[7] If σ^2 is bounded away from zero, all these assumptions hold but they also hold for some processes where σ^2 vanishes at the boundary.

vanish at the boundaries. When the drift coefficient has continuous first derivatives and diffusion coefficient continuous second derivatives, the right-hand side of (4.4) again defines the unique stationary distribution associated with the reflexive barrier process. The case with one reflecting barrier and the other one nonattracting can be handled in an analogous fashion.

For particular families of scalar diffusions and test functions, moment conditions in the class $C1$ have been used previously, albeit in other guises. For instance Wong (1964) has shown that first-order polynomial specifications of μ, and second-order polynomial specifications of σ^2 are sufficient to generate processes with stationary densities in the Pearson class. Pearson's method of moment estimation of these densities can be interpreted, except for its assumption that the data generation is *i.i.d.*, as appropriately parameterizing the polynomials defining the drift and diffusion coefficients and using polynomial test functions in $C1$. This approach has been extended to a broader class of densities by Cobb, Koopstein and Chen (1983). These authors suggested estimating higher-order polynomial specifications of μ for a pre-specified (second-order) polynomial specification of σ^2. Their estimation can again be interpreted as using moment conditions $C1$ with polynomial test functions. Of course moment conditions $C1$ and $C2$ are easy to apply for other test functions ϕ and ϕ^* whose first and second derivatives can be computed explicitly. Although Cobb, Koopstein and Chen (1983) prespecify the local variance to facilitate identification, this assumption is not convenient for many applications in economics and finance. As we will see in Section 5, moment conditions in the class $C2$ can be used to help identify and estimate unknown parameters of the local variance.[8]

Example 4.3: (*Multivariate Factor Models*) We have just seen that many scalar diffusions are reversible. Reversibility carries over to multivariate diffusions built up as time

[8]Since Cobb, Koopstein and Chen assume that the data generation is *i.i.d.* , moment conditions $C2$ are not informative in their framework. Recall, however, that we are interested in the case where the data originates from a single realization of $\{x_t\}$.

invariant functions of a vector of independent scalar diffusions. More precisely, let $\{f_t\}$ be a vector diffusion with component processes that are independent, stationary and reversible. Think of the components of $\{f_t\}$ as independent unobservable "factors." Suppose that the observed process is a time invariant function of the factors:

$$x_t = F(f_t)$$

for some function F. Factor models of bond prices like those proposed Cox, Ingersoll and Ross (1985), Longstaff and Schwartz (1992), Frachot, Janci and Lacoste (1992) and Duffie and Kan (1993) all have this representation. The resulting $\{x_t\}$ process will clearly be stationary as long as the factors are stationary. Moreover, since $\{f_t\}$ is reversible, so is $\{x_t\}$. To ensure that $\{x_t\}$ is a Markov process, we require that the factors can be recovered from the observed process. That is, F must be one-to-one. Finally, to guarantee that $\{x_t\}$ satisfies a stochastic differential equation, we restrict F to have continuous second derivatives. For a more general characterization of multivariable reversible diffusions, see Kent (1978).

Example 4.4: (*Multivariate Diffusion Models*) More generally, suppose that $\{x_t\}$ satisfies the stochastic differential equation:

$$dx_t = \mu(x_t)dt + \Sigma(x_t)^{1/2}dW_t$$

where $\{W_t : t \geq 0\}$ is an n-dimensional Brownian motion. Entry i of the *local mean* μ, denoted μ_i, and entry (i, j) of the *diffusion matrix* Σ, denoted σ_{ij}, are functions from $\mathbf{R}^n$ into $\mathbf{R}$. The functions μ_i and the entries of $\Sigma(x_t)^{1/2}$ are assumed to satisfy Lipschitz conditions. (See Theorem 3.2, page 79; the weaker local hypothesis of Theorem 4.1, page 84 of Has'minsky, 1980; or Theorem 3.7, page 297, Theorem 3.11, page 300 and Remark 2.7 on page 374 of Ethier and Kurz, 1986.) The existence of a unique stationary distribution is assured if there exists a K such that $\Sigma(y)$ is positive definite for $|\, y \,| \leq K$ and for $|\, y \,| \geq K$,

$$\mu(y) \cdot Vy + (1/2)\text{trace}[\Sigma(y)V] < -1$$

for some positive definite matrix V. Moreover this guarantees that the process is mean recurrent. (See Corollary 1, page 99; Example 1, page 103; and Corollary 2, page 123 of Has'minsky, 1980). If $\Sigma(y)$ is positive definite for all $y \in \mathbf{R}^n$, then the density q is given by the unique nonnegative bounded solution q to the partial differential equation:

$$-\sum_{j=1}^{n} \partial/\partial y_j[\mu_j(y)q(y)] + (1/2)\sum_{i=1}^{n}\sum_{j=1}^{n}(\partial^2/\partial y_i \partial y_j)[\sigma_{ij}(y)q(y)] = 0$$

that integrates to one. (See Has'minsky 1980, Lemma 9.4, page 138). Again by initializing the process using the measure Q with density q, we construct a stationary Markov process $\{x_t\}$.

The infinitesimal generator is defined on a subspace of $\mathcal{L}^2(Q)$ that contains at least the space of functions ϕ for which $\partial\phi/\partial y$ and $\partial^2\phi/\partial y \partial y'$ are continuous, the entries of $\Sigma^{1/2}\partial\phi/\partial y \in \mathcal{L}^2(Q)$ via:

$$\mathcal{A}\phi(y) = \mu(y) \cdot \partial\phi(y)/\partial y + (1/2)\text{trace}[\Sigma(y)\partial^2\phi(y)/\partial y \partial y'], \tag{4.9}$$

if the right-hand side of (4.9) is in $\mathcal{L}^2(Q)$. Under suitable regularity conditions, a time-reversed diffusion is still a diffusion and the adjoint $\mathcal{A}^*$ can be represented as

$$\mathcal{A}^*\phi^*(y) = \mu^*(y) \cdot \partial\phi^*(y)/\partial y + (1/2)\text{trace}[\Sigma^*(y)\partial^2\phi^*(y)/\partial y \partial y'] \tag{4.10}$$

on a subset of its domain.[9]

In (4.10) the diffusion matrix Σ^* turns out to be equal to Σ; however, the local mean μ^* may be distinct from μ. Let Σ_j denote column j of Σ. It follows from Nelson (1958), Anderson (1982), Haussmann and Pardoux (1986) or Millet, Nualart and Sanz (1989) that

$$\mu^*(y) = -\mu(y) + [1/q(y)]\sum_j \partial/\partial y_j[q(y)\Sigma_j(y)].$$

[9] For example, Millet, Nualart and Sanz (1989) showed that it suffices for coefficients μ and Σ of the diffusion to be twice continuously differentiable, to satisfy Lipschitz conditions, and for the matrix Σ to uniformly nonsingular (see Theorem 2.3 and Proposition 4.2).

Therefore, on a dense subset of its domain the adjoint $\mathcal{A}^*$ can be constructed from knowledge of μ, Σ and the stationary density q.

Moment conditions in the class $C1$ remain easy to apply to test functions whose two derivatives can be readily computed. For multivariate diffusions that are not time reversible, it is, in general, much more difficult to calculate the reverse-time generator used in $C2$. An alternative approach is to approximate the reverse-time generator using a nonparametric estimator of the logarithmic derivative of the density. With this approach, nonparametric estimators of the density and its derivative appear in the constructed moment conditions even though the underlying estimation problem is fully parametric. The nonparametric estimator is used only as a device for simplifying calculations. Rosenblatt (1969) and Roussas (1969) described and justified nonparametric estimators of densities for stationary Markov processes. Moreover, estimation using moment conditions constructed with nonparametric estimates of functions such as $\partial[\log q(y)]/\partial y$ has been studied in the econometrics literature (*e.g.*, see Gallant and Nychka 1987; Powell, Stock and Stoker 1989; Robinson 1989; Chamberlain 1992; Newey 1993; Lewbel 1991). Presumably results from these literatures could be extended to apply to our estimation problem.

5. Observable Implications

Recall that in Section 3 we derived two sets of moment conditions to be used in discriminating among a family of candidate generators. In this section we study the informational content of these two sets of moment conditions. Formally, there is a true generator $\mathcal{A}$ underlying the discrete-time observations. We then characterize the class of observationally equivalent generators from the vantage point of each of the sets of moment conditions. In the subsequent discussion, when we refer to a *candidate generator* we will presume that it is a generator for a Markov process.

We will establish the following identification results. First we will show that if a candidate generator $\hat{\mathcal{A}}$ satisfies the moment conditions in the set $C1$, it has a stationary distribution in common with the true process. Second we will show that if $\hat{\mathcal{A}}$ also satisfies the moment conditions in the set $C2$, it must commute with the true conditional expectation operator $\mathcal{T}$. If in addition $\hat{\mathcal{A}}$ is self adjoint, and the true process is reversible (*i.e.* $\mathcal{A}$ is self adjoint), then $\hat{\mathcal{A}}$ and $\mathcal{A}$ commute. One implication of this last result is that the drift and diffusion coefficients of stationary scalar diffusions can be identified up to a common scale factor using $C1$ and $C2$. A byproduct of our analysis is that using the conditional expectation operator allows one to identify fully the generator of reversible processes.

Consider first moment conditions $C1$. Since one of the goals of the econometric analysis is to ascertain whether a candidate generator $\hat{\mathcal{A}}$ has Q as a stationary distribution, it is preferable to begin with a specification of $\hat{\mathcal{A}}$ without reference to $\mathcal{L}^2(Q)$. Instead let $\mathcal{B}$ denote the space of bounded functions on $\mathbf{R}^n$ endowed with the sup norm. Suppose that $\hat{\mathcal{A}}$ is the infinitesimal generator for a strongly continuous contraction semigroup $\{\hat{\mathcal{T}}_t : t \geq 0\}$ defined on a closed subspace $\mathcal{L}$ of $\mathcal{B}$ containing at least all of the continuous functions with compact support. In this setting, strong continuity is the sup-norm counterpart to Assumption *A1*, *i.e.* $\{\hat{\mathcal{T}}_t\phi : t \geq 0\}$ converges uniformly to ϕ as t declines to zero for all ϕ in $\mathcal{L}$. Let $\hat{\mathcal{D}}$ denote the domain of $\hat{\mathcal{A}}$. We say that a candidate $\hat{\mathcal{A}}$ has Q as its stationary distribution if Q is the stationary distribution for a Markov process associated with this candidate. The following result is very similar to part of Proposition 9.2 of Ethier and Kurtz (1986, page 239).

Proposition 5.1. *Let $\hat{\mathcal{A}}$ be a candidate generator defined on $\hat{\mathcal{D}} \subseteq \mathcal{L}$. Then $\hat{\mathcal{A}}$ satisfies $C1$ for all $\phi \in \hat{\mathcal{D}}$ if and only if Q is a stationary distribution of $\hat{\mathcal{A}}$.*

Proof: Since convergence in the sup norm implies $\mathcal{L}^2(Q)$ convergence, our original derivation of $C1$ still applies. Conversely, note that the analog of Property *P4* implies that for any $\phi \in \mathcal{L} : \int_0^t \hat{T}_s(\phi)ds \in \hat{\mathcal{D}}$ and

$$\hat{\mathcal{A}} \int_0^t \hat{T}_s \phi ds = \hat{T}_t(\phi) - \phi. \tag{5.1}$$

Hence integrating both sides of (5.1) with respect to Q and using the fact that $\hat{\mathcal{A}}$ satisfies $C1$, we have that for all $\phi \in \mathcal{L}$:

$$\int_{\mathbf{R}^n} (\hat{T}_t \phi - \phi) dQ = 0. \tag{5.2}$$

Relation (5.2) can be shown to hold for all indicator functions of Borelians of $\mathbf{R}^n$ because $\mathcal{L}$ contains all continuous functions with a compact support. *Q.E.D.*

In light of Proposition 5.1, any infinitesimal generator $\hat{\mathcal{A}}$ satisfying $C1$ has Q as a stationary distribution. In other words, moment conditions in the family $C1$ cannot be used to discriminate among models with stationary distribution Q. On the other hand, if $\hat{\mathcal{A}}$ does not have Q as its stationary distribution, then there exists a test function ϕ in $\hat{\mathcal{D}}$ such that $E\hat{\mathcal{A}}\phi$ is different from zero.

Example 5.1: As in Example 4.2 suppose $\{x_t\}$ is a scalar diffusion that satisfies equation (4.3). If the stationary density q is given by the right hand side of (4.4) then

$$\mu = (1/2)[(\sigma^2)' + \sigma^2 q'/q]. \tag{5.3}$$

For a fixed q, this equation relates the diffusion coefficient σ^2 with the local drift μ. This equation gives us sets of observationally equivalent pairs (μ, σ^2) from the vantage point of $C1$. In fact Banon (1978) and Cobb, Koopstein and Chen (1983) used equation (5.3) as a basis to construct flexible (non-parametric) estimators of μ for prespecified σ^2. As is evident from (5.3) parameterizing (μ, σ^2) is equivalent, modulo some invertibility and regularity conditions, to parameterizing $(q'/q, \sigma^2)$. For many purposes the latter parameterization is simpler and more natural. If we start by describing our candidate models using this parameterization, moment conditions $C1$ yield no information about the diffusion coefficient. On the other hand, as we will see later in this section, moment conditions $C2$ provide a considerable amount of information about the diffusion coefficient.

To illustrate these points, as in Cobb, Koopstein and Chen (1983), consider a family of diffusions defined on the nonnegative reals parameterized by a "truncated" Laurent series:

$$(q'/q)(y,\alpha) = \sum_{j=-k}^{l} \alpha_j y^j .$$

where $\alpha = (\alpha_{-k}, \ldots, \alpha_l)$ is a vector of unknown parameters that must satisfy certain restrictions for q to be nonnegative and integrable. We also assume that $\sigma^2 = \rho y^\gamma$ where $\rho > 0$ and $\gamma > 0$. This parameterization is sufficiently rich to encompass the familiar "square-root" process used in the bond pricing literature as well as other processes that exhibit other volatility elasticities. The implicit parameterization of μ can be deduced from (5.3).[10] Moment conditions $C1$ will suffice for the identification of α. Since for fixed α, variations in γ leave invariant the stationary distribution q, γ cannot be inferred from moment conditions $C1$. However, as we will see in our subsequent analysis, moment conditions $C2$ will allow us to identify γ, but not ρ.

To assess the incremental informational content of the set of moment condition $C2$, we focus only on generators that satisfy $C1$. In light of Proposition 5.1, all of these candidates have Q as a stationary distribution. Strong continuity of the semigroup $\{\hat{\mathcal{T}}_t : t \geq 0\}$ in $\mathcal{L}$ implies Assumption *A1*. Thus we are now free to use $\mathcal{L}^2(Q)$ (instead of the more restrictive domain $\mathcal{L}$) as the common domain of the semigroups associated with the candidate generators. To avoid introducing new notation, for a candidate generator $\hat{\mathcal{A}}$ satisfying $C1$, we will still denote by $\hat{\mathcal{A}}$ the generator of the semigroup defined on $\mathcal{L}^2(Q)$ and by $\hat{\mathcal{D}}$ its domain.

Recall that $C2$ was derived using the fact that $\mathcal{A}$ and $\mathcal{T}$ commute. In fact, if a candidate $\hat{\mathcal{A}}$ satisfies $C2$, then $\hat{\mathcal{A}}$ must commute with $\mathcal{T}$, *i.e.* for any ϕ in $\hat{\mathcal{D}}$, $\mathcal{T}\phi$ is in $\hat{\mathcal{D}}$ and $\hat{\mathcal{A}}\mathcal{T}\phi = \mathcal{T}\hat{\mathcal{A}}\phi$.

[10] Additional restrictions must be imposed on the parameters to guarantee that the there is a solution to the associated stochastic differential equation. As mentioned in footnote 5, there are a variety of alternative sufficient conditions that can be employed to ensure that a solution exists. Alternatively, we can work directly with the implied infinitesimal generators and verify that there exist associated Markov processes for the admissible parameter values.

Proposition 5.2. *Suppose $\hat{\mathcal{A}}$ satisfies $C1$. Then $\hat{\mathcal{A}}$ satisfies $C2$ for $\phi \in \hat{\mathcal{D}}$ and $\phi^* \in \hat{\mathcal{D}}^*$ if, and only if $\hat{\mathcal{A}}\mathcal{T} = \mathcal{T}\hat{\mathcal{A}}$.*

Proof: By mimicking the reasoning in Section 3 one shows that $\hat{\mathcal{A}}\mathcal{T} = \mathcal{T}\hat{\mathcal{A}}$ is sufficient for $C2$. To prove necessity, note that for any ϕ in $\hat{\mathcal{D}}$ and any ϕ^* in $\hat{\mathcal{D}}^*$, it follows from $C2$ and the Law of Iterated Expectations that

$$< \mathcal{T}\hat{\mathcal{A}}\phi \mid \phi^* > = < \phi \mid \mathcal{T}^*\hat{\mathcal{A}}^*\phi^* > . \tag{5.4}$$

Since $\mathcal{T}$ is the adjoint of $\mathcal{T}^*$,

$$< \phi \mid \mathcal{T}^*\hat{\mathcal{A}}^*\phi^* > = < \mathcal{T}\phi \mid \hat{\mathcal{A}}^*\phi^* > . \tag{5.5}$$

It follows that $\hat{\mathcal{A}}\mathcal{T}\phi = \mathcal{T}\hat{\mathcal{A}}\phi$ for all ϕ in $\hat{\mathcal{D}}$, because the adjoint of $\hat{\mathcal{A}}^*$ is $\hat{\mathcal{A}}$,. *Q.E.D.*

Since it is typically hard to compute $\mathcal{T}\phi$ for an arbitrary function ϕ in $\mathcal{L}^2(Q)$, it may be difficult to establish directly that $\hat{\mathcal{A}}$ commutes with $\mathcal{T}$. As an alternative, it is often informative to check whether $\hat{\mathcal{A}}$ commutes with $\mathcal{A}$. To motivate this exercise, we investigate moment conditions $C2$ for arbitrarily small sampling intervals. By Proposition 5.2, this is equivalent to studying whether $\mathcal{T}_t$ and $\hat{\mathcal{A}}$ commute for arbitrarily small t.

Proposition 5.3. *Suppose $\hat{\mathcal{A}}$ commutes with $\mathcal{T}_t$ for all sufficiently small t. Then $\hat{\mathcal{A}}\mathcal{A}\phi = \mathcal{A}\hat{\mathcal{A}}\phi$ for all ϕ in $\hat{\mathcal{D}}$ with $\hat{\mathcal{A}}\phi$ in $\mathcal{D}$.*

Proof: Note that

$$\mathcal{A}\hat{\mathcal{A}}\phi = \lim_{t\downarrow 0}[(\mathcal{T}_t - \mathcal{I})/t]\hat{\mathcal{A}}\phi = \lim_{t\downarrow 0}\hat{\mathcal{A}}[(\mathcal{T}_t - \mathcal{I})/t]\phi. \tag{5.6}$$

Since $\hat{\mathcal{A}}$ has a closed graph, the right side of (5.6) must converge to $\hat{\mathcal{A}}\mathcal{A}\phi$. *Q.E.D.*

In light of this result, we know that any $\hat{\mathcal{A}}$ satisfying the *small interval* counterpart to $C2$ must commute with $\mathcal{A}$. Thus from Propositions 5.2- 5.3, if there exists an admissible test function ϕ such that

$$\hat{\mathcal{A}}\mathcal{A}\phi \neq \mathcal{A}\hat{\mathcal{A}}\phi,$$

then there is a sampling interval for which $\hat{\mathcal{A}}$ fails to satisfy some of the moment conditions in the collection $C2$. While we know there exists such a sampling interval, the conclusion is not necessarily applicable to all sampling intervals and hence may not be applicable to one corresponding to the observed data. As we will now see, this limitation can be overcome when additional restrictions are placed on the candidate and true generators. Conversely, suppose that $\hat{\mathcal{A}}$ commutes with $\mathcal{A}$ in a subspace of $\mathcal{L}^2(Q)$. Bequillard (1989) gave sufficient conditions on that subspace to ensure that $\hat{\mathcal{A}}$ commutes with $\{\mathcal{T}_t : t \geq 0\}$. Hence such a candidate generator can never be distinguished from $\mathcal{A}$ using moment conditions $C2$.

When both the candidate and true processes are reversible, we can show that requiring $\hat{\mathcal{A}}$ to commute with $\mathcal{T}$ is equivalent to requiring $\hat{\mathcal{A}}$ to commute with $\mathcal{A}$. Recall that reversible Markov processes have infinitesimal generators that are self adjoint. Such operators have unique spectral representations of the following form:

$$\mathcal{A}\phi = \int_{(-\infty,0]} \lambda d\mathcal{E}(\lambda)\phi \tag{5.7}$$

where $\mathcal{E}$ is a "resolution of the identity", *i.e.*

Definition 5.1: $\mathcal{E}$ is a resolution of the identity if

(i) $\mathcal{E}(\Lambda)$ is a self-adjoint projection operator on $\mathcal{L}^2(Q)$ for any Borelian $\Lambda \subset \mathbf{R}$;

(ii) $\mathcal{E}(\Phi) = 0, \mathcal{E}\{\mathbf{R}\} = \mathcal{I}$;

(iii) for any two Borelians Λ_1 and Λ_2, $\mathcal{E}(\Lambda_1 \cap \Lambda_2) = \mathcal{E}(\Lambda_1)\mathcal{E}(\Lambda_2)$;

(iv) for any two disjoint Borelians Λ_1 and Λ_2, $\mathcal{E}(\Lambda_1 \cup \Lambda_2) = \mathcal{E}(\Lambda_1) + \mathcal{E}(\Lambda_2)$;

(v) for $\phi \in \mathcal{D}$ and $\psi \in \mathcal{L}^2(Q)$, $< \mathcal{E}\phi \mid \psi >$ defines a measure on the Borelians.

Spectral representation (5.7) is the operator counterpart to the spectral representation of symmetric matrices. It gives an orthogonal decomposition of the operator $\mathcal{A}$ in the sense that if Λ_1 and Λ_2 are disjoint, $\mathcal{E}(\Lambda_1)$ and $\mathcal{E}(\Lambda_2)$ project onto orthogonal subspaces (see condition (iv) characterizing the resolution of the identity.) When the spectral measure of $\mathcal{E}$ has a mass point at a particular point λ, then λ is an eigenvalue of the operator and $\mathcal{E}(\lambda)$ projects onto the linear space of eigenfunctions associated with that eigenvalue. In light of condition (v), the integration in (5.7) can be defined formally in terms of inner products. The integration can be confined to the to the interval $(-\infty, 0]$ instead of all of $\mathbf{R}$ because $\mathcal{A}$ is negative semidefinite (*e.g.*, see Theorems 13.30 and 13.31 on page 349 of Rudin 1973). Finally, spectral decomposition (5.7) of $\mathcal{A}$ permits us to represent the semigroup $\{\mathcal{T}_t : t \geq 0\}$ via the exponential formula:

$$\mathcal{T}_t\phi = \int_{(-\infty,0]} \exp(\lambda t) d\mathcal{E}(\lambda)\phi \tag{5.8}$$

(*e.g.*, see Theorem 13.37 of Rudin 1973, page 360).

Consider now a candidate generator $\hat{\mathcal{A}}$ that is reversible and satisfies $C1$. Then

$$\hat{\mathcal{A}}\phi = \int_{(-\infty,0]} \lambda d\hat{\mathcal{E}}(\lambda)\phi, \tag{5.9}$$

where $\hat{\mathcal{E}}$ is also resolution of the identity. Suppose that $\hat{\mathcal{A}}$ commutes with $\mathcal{T}$. Then $\mathcal{T}$ must commute with $\hat{\mathcal{E}}(\Lambda)$ for every Borelian Λ (see Theorem 13.33 of Rudin 1973, page 351). Let $\{\hat{\mathcal{T}}_t\}$ be the semigroup generated by $\hat{\mathcal{A}}$. Since $\hat{\mathcal{T}}_\tau$ is a bounded operator and can be constructed from $\hat{\mathcal{A}}$ via the exponential formula analogous to (5.8), $\hat{\mathcal{T}}_\tau$ must commute with $\mathcal{T}$ for every nonnegative τ. Moreover, $\hat{\mathcal{T}}_\tau$ must commute with $\mathcal{E}(\Lambda)$ for any Borelian Λ and hence with $\mathcal{T}_t$ for any nonnegative t. We have thus established:

Proposition 5.4. *Suppose $\hat{\mathcal{A}}$ satisfies $C1$ and $\mathcal{A}$ and $\hat{\mathcal{A}}$ are self-adjoint. If $\hat{\mathcal{A}}\mathcal{T} = \mathcal{T}\hat{\mathcal{A}}$, then $\hat{\mathcal{A}}\mathcal{T}_t = \mathcal{T}_t\hat{\mathcal{A}}$ for all $t \geq 0$.*

Propositions 5.1-5.4 support the following approach to identification. Suppose one begins with a parameterization of a family of candidate generators. First partition this family into collections of generators with the same stationary distributions. Then partition further the family into groups of generators that commute (on a sufficiently rich collection of test functions). Two elements in the same subset of the original partition cannot be distinguished on the basis of moment condition in the set $C1$, and two elements in the same subset of the finest partition cannot be distinguished on the basis of $C2$.

Moment conditions $C1$ and $C2$ do not capture all of the information from discrete-time data pertinent to discriminating among generators. To see this, we consider identification results based instead on knowledge of the discrete-time conditional expectation operator $\mathcal{T}$. In the case of reversible Markov processes, any candidate $\hat{\mathcal{A}}$ that implies $\mathcal{T}$ as a conditional expectation operator, must satisfy the exponential formula:

$$\mathcal{T}\phi = \int_{-\infty}^{0} \exp(\lambda t) d\hat{\mathcal{E}}(\lambda)\phi. \tag{5.10}$$

Moreover, the spectral decomposition for self-adjoint operators is unique (see Theorem 13.30, Rudin 1973, page 348). Since the exponential function is one-to-one, it follows that $\hat{\mathcal{E}}$ and $\mathcal{E}$ and hence $\hat{\mathcal{A}}$ and $\mathcal{A}$ must coincide. Thus we have shown:

Proposition 5.5. *Suppose that the generators $\mathcal{A}$ and $\hat{\mathcal{A}}$ are self-adjoint and imply the same conditional expectation operator $\mathcal{T}$. Then $\mathcal{A} = \hat{\mathcal{A}}$.*

Proposition 5.5 implies that there is no *aliasing* problem when the Markov processes is known to be reversible. Aliasing problems in Markov processes arise because of the presence of complex eigenvalues of the generators. For reversible processes all of the eigenvalues are real and negative (the corresponding resolutions of the identity are concentrated on the nonpositive real numbers). Similarly, the eigenvalues of $\mathcal{T}$ must be in the interval (0,1]. As we will see in our examples, moment conditions $C1$ and $C2$ fail to achieve complete identification of $\mathcal{A}$ for reversible processes.

More generally, even if we do not impose reversibility, $\hat{\mathcal{A}}$ and $\mathcal{T}$ are connected through the following alternative exponential formula:

$$\lim_{n\to\infty}(I-\hat{\mathcal{A}}/n)^{-n}\phi = \mathcal{T}\phi \text{ for all } \phi \in \mathcal{L}^2(Q) \tag{5.11}$$

(*e.g.*, see Pazy 1983, page 33). This exponential formula is the generalization of a formula used to study *aliasing* and *embeddability* for finite state Markov chains (*e.g*, see Johansen 1973; Singer and Spillerman 1976). Since $\hat{\mathcal{A}}$ is a continuous operator in this case, the exponential formulas simplify to $\mathcal{T} = \exp(\hat{\mathcal{A}})$. Relation (5.11) also encompasses the exponential formulas derived by Phillips (1973) and Hansen and Sargent (1983) in their analysis of aliasing in the class of multivariate Gaussian diffusion models.

We now apply and illustrate our results in the context of three examples.

Example 5.1 (continued): As we saw earlier, using moment conditions in the set $C1$ may still leave a large class of observationally equivalent Markov processes. We now show that using moment conditions $C2$ we may narrow this class to an easily interpretable one-dimensional family.

Suppose that both the true and the candidate processes are reversible scalar diffusions, that share a stationary distribution Q. Let μ and σ^2 denote the local mean and diffusion coefficients associated with the true process, and let $\hat{\mu}$ and $\hat{\sigma}^2$ denote their counterparts for the candidate process. We maintain the assumptions made in Example 4.2 of Section 4. Write L for the second-order differential operator associated with the pair (μ, σ^2) and $\hat{L}$ for the corresponding operator associated with $(\hat{\mu}, \hat{\sigma}^2)$.

We start by examining the case of diffusions on a compact interval $[\ell, u]$ with two reflective barriers and a strictly positive diffusion coefficient, to which the standard Sturm-Liouville theory of second-order differential equations applies. Consider the following eigenvalue problem associated with $\hat{L}$:

$$\hat{L}\phi = \hat{\lambda}\phi, \quad \phi'(\ell) = \phi'(u) = 0. \tag{5.12}$$

In light of the boundary conditions imposed in (5.12), we know that a twice continuously differentiable (C^2) solution to this eigenvalue problem will result in an eigenvector for $\hat{\mathcal{A}}$. From Sturm-Liouville theory there exists an infinite sequence of negative numbers $\hat{\lambda}_0 > \hat{\lambda}_1 > \hat{\lambda}_2 > \ldots$ with $\lim_{n\to\infty} \hat{\lambda}_n = -\infty$ and corresponding unique, up to constant factors, C^2 functions ϕ_n such that the pair $(\hat{\lambda}_n, \phi_n)$ solves the eigenvalue problem (5.12). Choose a negative $\hat{\lambda}_n$. Suppose that all of the moment conditions in the class $C2$ are satisfied by $\hat{\mathcal{A}}$. By Proposition 5.4, $\mathcal{T}_t$ and $\hat{\mathcal{A}}$ commute, and hence $\mathcal{T}\phi_n$ is also an eigenvector of $\hat{\mathcal{A}}$ associated with $\hat{\lambda}_n$. Consequently, $\mathcal{T}$ and $\mathcal{A}$ must have ϕ_n as an eigenvector associated with a perhaps different eigenvalue λ_n. Since ϕ_n is C^2 and satisfies the appropriate boundary conditions, $\mathcal{A}\phi_n$ coincides with $L\phi_n$ and hence ϕ_n satisfies the counterpart to (5.12):

$$L\phi_n = \lambda_n \phi_n. \tag{5.13}$$

Multiply this equation by q and substitute $(1/2)(\sigma^2 q)'$ for μq, to obtain:

$$(1/2)(\sigma^2 q)'\phi_n' + (1/2)(\sigma^2 q)\phi_n'' = \lambda_n \phi_n q,$$

or,

$$(1/2)(\sigma^2 q \phi_n')' = \lambda \phi_n q.$$

Therefore the eigenvector ϕ_n satisfies:

$$\sigma^2(y)\phi_n'(y)q(y) = 2\lambda_n \int_\ell^y \phi_n(x)q(x)dx + C. \tag{5.14}$$

The boundary conditions on ϕ_n and condition $C1$ assure us that the constant C in (5.14) is in fact zero. Similarly, we conclude that

$$\hat{\sigma}^2(y)\phi_n'(y)q(y) = 2\hat{\lambda} \int_\ell^y \phi_n(x)q(x)du. \tag{5.15}$$

It follows from (5.14) and (5.15) that σ^2 and $\hat{\sigma}^2$ are proportional and hence from formula (5.3) the μ and $\hat{\mu}$ are also proportional with the same proportionality factor given by the ratio of the eigenvalues. In other words, moment condition $C2$ permits us to identify the infinitesimal generator $\mathcal{A}$ up to scale.

We now show how this identification result can be extended to processes defined on the whole real line, even though in this case the generator may fail to have any non-zero eigenvalue. We will proceed by considering reflexive barrier processes that approximate the original process. We assume that both μ and σ^2 are C^2 functions and with $\sigma^2 > 0$. When $\hat{\mathcal{A}}$ satisfies moment conditions $C1$, by Proposition 5.1 the candidate Markov process shares a stationary density q with the true process. Suppose $\hat{\mathcal{A}}$ also satisfies moment conditions $C2$. For a given $k > 0$ consider the processes created by adding to the original candidate and actual processes reflexive barriers at $-k$ and k. The reflexive barrier processes will share a common stationary distribution Q_k with a density q_k that, in the interval $[-k, k]$, is proportional to q. We write $\mathcal{A}_k$ and $\hat{\mathcal{A}}_k$ for the infinitesimal generators associated with the reflexive barrier processes in $\mathcal{L}^2(Q_k)$. As before we can use Sturm-Liouville theory to establish the existence of a negative $\hat{\lambda}$ and a C^2 function ϕ such that

$$\hat{L}\phi = \hat{\lambda}\phi, \ \ \phi'(-k) = \phi'(k) = 0. \tag{5.16}$$

Since μ and σ^2 are C^2 functions, and $\sigma^2 > 0$, the eigenvector $\phi \in C^4$ on $(-k, k)$ and all derivatives up to fourth order have well define limits as $y \to \pm k$. We write $f^{(n)}$ for the n-th derivative of f. Construct ψ_+ to be a C^4 function defined on (k, ∞) with $\psi_+(y) = 0$ when $y \geq k+1$ and $\lim_{y \to k} \psi_+^{(n)}(y) = \lim_{y\to k} \phi^{(n)}(y)$ for $0 \leq n \leq 4$. Similarly construct ψ_- to be a C^4 function defined on $(-\infty, -k)$ with $\psi_-(y) = 0$ when $y \leq -k-1$ and $\lim_{y \to -k} \psi_-^{(n)}(y) = \lim_{y \to -k} \phi^{(n)}(y)$ for $0 \leq n \leq 4$. Finally let $\psi(y) \equiv \phi(y)$ if $-k \leq y \leq k$, and $\psi(y) \equiv \psi_+$ (ψ_-) if $y > k$ (resp. $y < -k$).

Notice ψ is a C^4 function with support in $[-k-1, k+1]$ and that $\hat{L}\psi$ is C^2 and has a compact support. Hence $\hat{\mathcal{A}}\psi = \hat{L}\psi$ and, $\mathcal{A}\hat{\mathcal{A}}\psi = L\hat{L}\psi$. Also, $L\psi$ is a C^2 function with compact support and hence $\hat{\mathcal{A}}\mathcal{A}\psi = \hat{L}L\psi$. Since by Propositions 5.3 and 5.4 $\mathcal{A}$ and $\hat{\mathcal{A}}$ must commute, $\hat{L}L\psi = L\hat{L}\psi$, and hence $\hat{\mathcal{A}}_k\mathcal{A}_k\phi = \mathcal{A}_k\hat{\mathcal{A}}_k\phi$. The result for the compact support case with reflexive barriers implies that the drift and diffusion coefficient are identified up to scale in an interval $(-k, k)$ for arbitrary k.

Finally, we observe that if an eigenvector ϕ of $\mathcal{A}$ can be explicitly calculated, then the corresponding (real) eigenvalue and hence scale constant can also be identified by using the fact that $\mathcal{T}\phi = e^{\lambda}\phi$.

Example 5.2: In Section 4 we introduced a class of factor models in which the process $\{x_t\}$ is a time invariant function of a vector independent scalar diffusions. Suppose that both the candidate process and the true process satisfy these factor restrictions. Then it follows from Propositions 5.1-5.4 that $\mathcal{A}$ and $\hat{\mathcal{A}}$ can be distinguished on the basis of our moment conditions if they imply different stationary distributions or if they fail to commute. By extending the discussion in Example 5.1 we can construct $\hat{\mathcal{A}}$'s that are not distinguishable from $\mathcal{A}$ as follows. Use the same function F mapping the factors into the observable processes as is used for $\mathcal{A}$. Then form an $\hat{\mathcal{A}}$ by multiplying the coefficients of each scalar factor diffusions by a possibly distinct scale factor. While such an $\hat{\mathcal{A}}$ can not be distinguished using moment conditions $C1$ and $C2$ they can be distinguished based on knowledge of the conditional expectation operator $\mathcal{T}$ (see Proposition 5.5).

Example 5.3: For our last example, we start from an arbitrary infinitesimal generator $\mathcal{A}$ and construct a two-parameter family of candidate generators that always satisfy $C1$ and $C2$. Let $\{x_t\}$ be a continuous time stationary Markov process with infinitesimal generator $\mathcal{A}$ on a domain $\mathcal{D}$. For any $\phi \in \mathcal{D}$, construct $\hat{\mathcal{A}}\phi = \eta_1\mathcal{A}\phi + \eta_2[\int \phi dQ - \phi]$ where η_1 and η_2 are two positive real numbers. Notice that we formed our candidate $\hat{\mathcal{A}}$ by changing the *speed* of the original process by multiplying $\mathcal{A}$ by η_1 and adding to that the generator for a particular Markov jump process [see (4.2)]. It is easy to check that $\hat{\mathcal{A}}$ has Q as its stationary distribution and commutes with $\mathcal{A}$. Therefore $\hat{\mathcal{A}}$ cannot be distinguished from $\mathcal{A}$ using $C1$ and $C2$.

When $\mathcal{A}$ is self adjoint so is $\hat{\mathcal{A}}$. Consequently it follows from Proposition 5.5 that in this case the generators can be distinguished based on knowledge of the conditional expectations operator.

6. Cores

So far, we have analyzed observable implications by assuming all of the moment conditions in $C1$ or $C2$ could be checked. To perform such a check would require both knowledge of the domain $\hat{\mathcal{D}}$ and the ability to compute $\hat{\mathcal{A}}$. It is often difficult to characterize the domain $\hat{\mathcal{D}}$ and to evaluate a candidate generator applied to an arbitrary element of that domain. For instance, in the scalar diffusion example (Example 4.2) we only characterized the infinitesimal generator on a subset of the domain. Furthermore, it is desirable to have a *common* set of test functions to use for a parameterized family of generators.

Since the operator $\hat{\mathcal{A}}$ is not necessarily continuous, a dense subset of $\hat{\mathcal{D}}$ does not need to have a dense image under $\hat{\mathcal{A}}$. Consequently, in examining moment conditions $C1$ and $C2$, if we replace $\hat{\mathcal{D}}$ and $\hat{\mathcal{D}}^*$ by arbitrary dense subsets, we may weaken their implications. In addition, recall that the moment conditions in $C2$ require looking at means of random variables of the form $\hat{\mathcal{A}}\phi(x_{t+1})\phi^*(x_t) - \phi(x_{t+1})\hat{\mathcal{A}}^*\phi^*(x_t)$, which are differences of products of random variables with finite second moments. To ensure that standard central limit approximations work, it is convenient to restrict ϕ and ϕ^* so that $\hat{\mathcal{A}}\phi(x_{t+1})\phi^*(x_t) - \phi(x_{t+1})\hat{\mathcal{A}}^*\phi^*(x_t)$ has a finite second moment. For instance, this will be true when both ϕ and ϕ^* are bounded.

To deal with these matters, we now describe a strategy for reducing the sets $\hat{\mathcal{D}}$ and $\hat{\mathcal{D}}^*$ that avoids loosing information and results in random variables with finite second moments. The approach is based on the concept of a *core* for a generator $\mathcal{A}$. Recall that the graph of $\mathcal{A}$ restricted to a set $\mathcal{N}$ is $\{(\phi, \mathcal{A}\phi) : \phi \in \mathcal{N}\}$.

Definition 5.1: A subspace $\mathcal{N}$ of $\mathcal{D}$ is a *core* for $\mathcal{A}$ if the graph of $\mathcal{A}$ restricted to $\mathcal{N}$ is dense in the graph of $\mathcal{A}$.

Clearly, if $\mathcal{N}$ is a core for $\mathcal{A}$, $\mathcal{A}(\mathcal{N})$ is dense in $\mathcal{A}(\mathcal{D})$. As we will argue below, in checking moment conditions $C1$ and $C2$ it suffices to look at sets whose linear spans are cores for $\hat{\mathcal{A}}$ and $\hat{\mathcal{A}}^*$.

For a Markov jump model (Example 4.1), a candidate generator $\hat{\mathcal{A}}$ is a bounded operator. In this case, it suffices to look at a countable collection of bounded functions whose linear span is dense in $\mathcal{L}^2(\hat{Q})$ for all probability measures $\hat{Q}$.

For a Markov diffusion model (Examples 4.2 to 4.4), $\hat{\mathcal{A}}$ is no longer a bounded operator. In order to apply Proposition 5.1, it is important to characterize a core for the candidate generator defined on $\mathcal{L}$. Recall that we assumed that $\mathcal{L}$ contained all the continuous functions with a compact support. For concreteness we now assume that $\mathcal{L} = \{\phi : \mathbf{R}^n \to \mathbf{R}$ s.t. ϕ is continuous, and $\lim_{|x|\to\infty} \phi(x) = 0\}$ with the sup norm. Notice that since convergence in the sup norm implies convergence of the mean, it suffices to verify moment conditions $C1$ on a core. Ethier and Kurtz (1986) (Theorem 2.1 page 371) showed that under the conditions stated in Example 4.2, the space of all infinitely differentiable functions with a compact support (C_K^∞) forms a core for the infinitesimal generator associated with the scalar stochastic differential equation. In this case since $C_K^2 \subset \mathcal{D}$, it is also a core for $\mathcal{A}$. Since, in analogy to the result we presented in Example 4.2, for these functions the infinitesimal generator is given by the second-order differential operator L, we can easily perform the calculations needed to apply moment condition $C1$. The Ethier and Kurtz result also covers certain cases with finite support and inaccessible or reflexive boundary conditions. Extensions to the multi-dimensional case, that typically require stronger smoothness conditions, are given in Theorem 2.6 and Remark 2.7, page 374 of Ethier and Kurtz (1986).

To apply Propositions 5.2-5.4 we need to consider candidate generators defined on $\mathcal{L}^2(Q)$. If $\mathcal{N}$ is a core for both $\mathcal{A}$ and $\mathcal{A}^*$ and our second moment condition holds for any $\phi \in \mathcal{N}$ and $\phi^* \in \mathcal{N}$ then it must hold for any pair $(\phi, \phi^*) \in \mathcal{D} \times \mathcal{D}^*$. This follows since if $\phi_n \to \phi$, $\hat{\mathcal{A}}\phi_n \to \hat{\mathcal{A}}\phi$, $\phi_n^* \to \phi*$ and, $\hat{\mathcal{A}}^*\phi_n^* \to \hat{\mathcal{A}}^*\phi^*$ then, since $\mathcal{T}$ is continuous,

$$< \mathcal{T}\hat{\mathcal{A}}\phi_n \mid \phi_n^* > - < \hat{\mathcal{A}}^*\phi_n^* \mid \mathcal{T}\phi_n > \to < \mathcal{T}\hat{\mathcal{A}}\phi \mid \phi^* > - < \hat{\mathcal{A}}^*\phi^* \mid \mathcal{T}\phi > .$$

Hence it suffices to characterize cores for $\mathcal{A}$ and $\mathcal{A}^*$. We can readily extend the results available in the literature on cores of infinitesimal generators for semigroups defined on subspaces of the space of continuous bounded functions with the sup norm. For instance consider an extension of Ethier and Kurtz's Theorem 2.1 cited above. We use the fact that $\mathcal{N}$ is a core for $\mathcal{A}$ if and only if both $\mathcal{N}$ and the image of $\mathcal{N}$ under $\lambda I - \mathcal{A}$ for some $\lambda > 0$ are dense on the domain of the contraction semigroup.[11] For diffusions with continuous coefficients the infinitesimal generator coincides on C_K^2 with the second-order differential operator. This is true whether the semigroup is defined on $\mathcal{L}$ or on $\mathcal{L}^2(Q)$. Since C_K^∞ forms a core for the candidate infinitesimal generator $\hat{\mathcal{A}}_\mathcal{L}$ of the semigroup defined on $\mathcal{L}$, there must exist a $\lambda > 0$ such that the image of C_K^∞ under $\lambda I - \hat{L}$ is dense in $\mathcal{L}$. It follows that the image of C_K^∞ under $\lambda I - \hat{L}$ is dense in $\mathcal{L}^2(Q)$, because C_K^∞ is dense in $\mathcal{L}^2(Q)$ and sup-norm convergence implies $\mathcal{L}^2(Q)$ convergence. Hence C_K^∞ is a core for $\hat{\mathcal{A}}$, the infinitesimal generator for the semigroup defined in $\mathcal{L}^2(Q)$. Notice that, since sup-norm convergence implies $\mathcal{L}^2(Q)$ convergence, we may apply exactly the same reasoning whenever we have a core $\mathcal{N}$ for $\hat{\mathcal{A}}_\mathcal{L}$ and we know that $\hat{\mathcal{A}}_{|\mathcal{N}} = (\hat{\mathcal{A}}_\mathcal{L})_{|\mathcal{N}}$.[12]

Even when reduced to a core the observable implications of conditions $C1$ and $C2$ pressupose the use of large set of test functions. Of course, for a finite data set, only a small (relative to the sample size) number of test functions will be used. We now obtain a reduction that can be used to support theoretical investigations in which the number of test functions can increase with the sample size such as Bierens (1990), who suggested a way of testing an infinite number of moment conditions using penalty functions, and Newey (1990) who derived results for efficient estimation by expanding the number of moment conditions as an explicit function of the sample size. Their analyses could be potentially adapted to the framework of this paper once we construct a countable collection of test functions whose *span* is a core. For the Markov jump process this

[11] See Proposition 3.1, page 17 of Ethier and Kurtz (1986).

[12] If the assumptions concerning the bounds on the derivatives of the coefficients made in Example 4.2 are replaced by weaker polynomial growth conditions, it is still possible to show directly that C_K^2 is a core for $\hat{\mathcal{A}}$ if the stationary distribution possesses moments of sufficiently high order.

reduction is easy since it suffices to choose any collection of functions with a dense span in $\mathcal{L}^2(Q)$. For Markov diffusions that have C_K^2 as a core we may proceed as follows. Fix a positive integer N and consider the subspace of C_K^2 of functions with support on $\{y : | y | \leq N\}$. This subspace is separable if we use the norm given by the maximum of the sup norm of a function and of its first two derivatives. Choose a countable dense collection for each N and take the union over positive integers N. Since C_K^2 is a core, it is straightforward to show that the linear span of this union is also a core for $\mathcal{A}$.

7. Ergodicity and Martingale Approximation

To use the moment conditions derived in Section II in econometric analyses, we must have some way of approximating expectations of functions of the Markov state vector x_t in the case of $C1$ and of functions of both x_{t+1} and x_t in the case of $C2$. As usual, we approximate these expectations by calculating the corresponding time-series averages from a discrete-time sample of finite length. To justify these approximations via a Law of Large Numbers, we need some form of ergodicity of the discrete-sampled process. In the first subsection, we investigate properties of the infinitesimal generator that are sufficient for ergodicity of both the continuous time and the discrete-time processes.

It is also of interest to assess the magnitude of the sampling error induced by making such approximations. This assessment is important for determining the statistical efficiency of the resulting econometric estimators and in making statistical inferences about the plausibility of candidate infinitesimal generators. The vehicle for making this assessment is a Central Limit Theorem. In the second subsection we derive central limit approximations via the usual martingale approach. Again we study this problem from the vantage point of both continuous and discrete records, and we derive sufficient conditions for these martingale approximations to apply that are based directly on properties of infinitesimal generators.

7.1. Law of Large Numbers

Stationary processes in either discrete time or continuous time obey a Law of Large Numbers. However, the limit points of time series averages may not equal the corresponding expectations under the measure Q. Instead these limit points are expectations conditioned on an appropriately constructed set of *invariant events* for the Markov process. Furthermore, the conditioning set for the continuous-time process may be a proper subset of the conditioning set for a fixed interval discrete-sampled process. Therefore, the limit points for the discrete record Law of Large Numbers may be different than the limit points for the continuous record Law of Large Numbers.

Invariant events for Markov processes (in continuous or discrete time) turn out to have a very simple structure. They are measurable functions of the initial state vector x_0. Hence, associated with the invariant events for the continuous-time process $\{x_t\}$, there is a sigma algebra $\mathcal{G}$, contained in the Borelians of $\mathbf{R}^n$, such that any invariant event is of the form $A = \{x_0 \in B\}$ for some $B \in \mathcal{G}$. We can construct a conditional probability distribution Q_y indexed by the initial state $x_0 = y$ such that expectations conditioned on $\mathcal{G}$ (as a function of y) can be evaluated by integrating with respect to Q_y. Imagine initializing the Markov process using Q_y in place of Q in our analysis where y is selected to be the observed value of x_0. Then under this new initial distribution the process $\{x_t\}$ remains stationary, but it is now ergodic for (Q) almost all y. Therefore, by an appropriate initialization we can convert any stationary process into one that is also ergodic. Alternatively, it can be verified that the moment conditions $C1$ and $C2$ hold conditioned on the invariant events for the continuous-time process $\{x_t\}$. Consequently, imposition of ergodicity for the continuous time process is made as matter of convenience.

Since ergodicity of $\{x_t\}$ is connected directly to the initial distribution imposed on x_0, it is of interest to have a criterion for checking whether an appropriate initial distribution has been selected. As we will see, such a criterion can be obtained by examining the zeros of the operator $\mathcal{A}$. Note that constant functions are always zeros of the operator $\mathcal{A}$.

We omit all such functions from consideration except for the zero function by focusing attention on the following closed linear subspace:

$$\mathcal{Z}(\mathcal{Q}) \equiv \{\phi \in \mathcal{L}^2(\mathcal{Q}) : \int_{\mathbf{R}^n} \phi d\mathcal{Q} = 0\}.$$

For any $t, \mathcal{T}_t$ maps $\mathcal{Z}(\mathcal{Q})$ into itself. Hence we may consider $\{\mathcal{T}_t : t \geq 0\}$ as having the domain $\mathcal{Z}(\mathcal{Q})$. Furthermore, since $\mathcal{D}$ is a linear subspace of $\mathcal{L}^2(\mathcal{Q})$, it follows from *P6* that $\mathcal{D} \cap \mathcal{Z}(\mathcal{Q})$ is a linear subspace and is dense in $\mathcal{Z}(\mathcal{Q})$.

Our first result relates the ergodicity of the continuous-time process $\{x_t\}$ to the uniqueness of the zeros of $\mathcal{A}$ on $\mathcal{Z}(\mathcal{Q})$. In the proof of this proposition and in what follows, $\mathcal{L}^2(\mathcal{P}r)$ denotes the space of all random variables on $(\Omega, \mathcal{F}, \mathcal{P}r)$ that have finite second moments with the usual norm. Note that if $\phi \in \mathcal{L}^2(\mathcal{Q})$, then the $\mathcal{L}^2(\mathcal{Q})$ norm of ϕ coincides with the $\mathcal{L}^2(\mathcal{P}r)$ norm of the random variable $\phi(x_t)$.

Proposition 7.1. *The process* $\{x_t\}$ *is ergodic if, and only if* $\mathcal{A}\phi = 0$ *for* $\phi \in \mathcal{D} \cap \mathcal{Z}(\mathcal{Q})$ *implies that* $\phi = 0$.

Proof: First suppose that $\{x_t\}$ is ergodic. Let $\phi \in \mathcal{D} \cap \mathcal{Z}(\mathcal{Q})$ be a zero of $\mathcal{A}$. By the Mean Ergodic Theorem (see Dunford and Schwartz 1988, Corollary 3, page 689) $\{(1/T)\int_0^T \phi(x_t)dt : T > 0\}$ converges in $\mathcal{L}^2(\mathcal{P}r)$. This limit random variable is invariant to a shift in the starting time and thus it is measurable with respect to the sigma algebra of invariant events. By assumption, $\{x_t\}$ is ergodic implying that the limit random variable must be constant and hence equal to zero. Since

$$E\{|\,(1/T)\int_0^T \phi(x_t)dt\,|^2\} \geq E[|\,E\{(1/T)\int_0^T \phi(x_t)dt \mid x_0\}\,|^2],$$

it follows that $\{(1/T)\int_0^T \mathcal{T}_t\phi dt : T > 0\}$ converges in $\mathcal{L}^2(\mathcal{Q})$ to zero. Recall that the semigroup $\{\mathcal{T}_t : t \geq 1\}$ satisfies *P5*. Consequently,

$$\mathcal{T}_t\phi = \phi + \int_0^t \mathcal{T}_s\mathcal{A}\phi ds = \phi.$$

Therefore, $(1/T)\int_0^T \mathcal{T}_t\phi dt = \phi$ implying that $\phi = 0$ with probability one. Next suppose the only zero of $\mathcal{A}$ on $\mathcal{D} \cap \mathcal{Z}(Q)$ is the zero function. Let $\phi \in \mathcal{Z}(Q)$ be $\mathcal{G}$ measurable. Then $\phi(x_t) = \phi(x_0)$ almost surely ($\mathcal{P}r$), and hence $\mathcal{T}_t\phi = \phi$ almost surely (Q). Consequently, $\phi \in \mathcal{D}$ and $\mathcal{A}\phi = 0$. Hence ϕ must be the zero function and $\mathcal{G}$ must contain only sets with Q measure zero and one. Similarly, invariant events of $\{x_t\}$ must have $\mathcal{P}r$ measure zero or one. *Q.E.D.*

Next we consider ergodicity of the discrete-sampled process. For notational simplicity, we set the sampling interval to one. The discrete time counterpart to Proposition 7.1 is:

Proposition 7.2. *The sampled process $\{x_t : t = 0, 1, \ldots\}$ is ergodic if and only if $\mathcal{T}\phi = \phi$ and $\phi \in \mathcal{Z}(Q)$ imply $\phi = 0$.*

The proof of this result is very similar to the proof of Proposition 7.1 and will be omitted.

Ergodicity of the continuous-time process $\{x_t\}$ does not necessarily imply ergodicity of this process sampled at integer points in time. From equation (5.11) $\mathcal{T}_t$ can be interpreted as the operator exponential of $t\mathcal{A}$ even though, strictly speaking, the exponential of $t\mathcal{A}$ may not be well defined. The Spectral Mapping Theorem for infinitesimal generators (*e.g.*, see Pazy 1983, Theorem 2.4, page 46) states that the nonzero eigenvalues of $\mathcal{T}_t$ are exponentials of the eigenvalues of $t\mathcal{A}$. Therefore, the only way in which $\mathcal{T}$ can have a unit eigenvalue on $\mathcal{Z}(Q)$ is for $\mathcal{A}$ to have $2\pi ki$ as an eigenvalue on $\mathcal{Z}(Q)$. In other words, there must exist a pair of functions ϕ_r and ϕ_i in $\mathcal{Z}(Q)$, at least one of which is different from zero, such that

$$\mathcal{A}(\phi_r + i\phi_i) = 2\pi ki(\phi_r + i\phi_i).$$

When $\mathcal{A}$ has a purely imaginary eigenvalue there will always be a nondegenerate function ϕ in $\mathcal{Z}(Q)$ such that $\phi(x_t)$ is perfectly predictable given x_0. In other words, there exists a nondegenerate function ϕ such that $\phi(x_t) = \mathcal{T}_t\phi(x_0)$ almost surely $(\mathcal{P}r)$ for all $t \geq 0$.[13] To ensure ergodicity for *all* sampling intervals, we must rule out *all* purely imaginary eigenvalues.

Generators of stationary, ergodic (in continuous time) Markov jump processes (Example 4.1) do not have purely imaginary eigenvalues. Suppose to the contrary that the pair (ϕ_r, ϕ_i) solves the eigenvector problem for eigenvalue $i\theta$ different from zero. Then

$$\tilde{\mathcal{T}}(\phi_r + i\phi_i) = [(i\theta/\lambda) + 1](\phi_r + i\phi_i).$$

This leads to a contradiction because $|\, (i\theta/\lambda) + 1 \,| > 1$ and $\tilde{\mathcal{T}}$ is a weak contraction.

Recall that generators of reversible process, including those discussed in Examples 4.2 and 4.3, have only real eigenvalues. Many other multivariate stationary, ergodic (in continuous time) Markov diffusion processes (Example 4.4) do not have purely imaginary eigenvalues. For instance, under the conditions given in Section 4 for the existence of a unique stationary distribution, for sufficiently large $t, \phi(x_t) = \mathcal{T}_t\phi(x_0)$ almost surely $(\mathcal{P}r)$ implies $\phi(x_0) = 0$ almost surely $(\mathcal{P}r\ )$ (see Has'minskii 1980, Lemma 6.5, pages 128 and 129). Hence $\mathcal{A}$ cannot have a purely imaginary eigenvalue in this case.

We now reconsider the moment conditions derived in Section 3. Recall that moment conditions $C1$ imply that $\mathcal{A}(\mathcal{D}) \subseteq \mathcal{Z}(\mathcal{Q})$. Consequently, the approximation results we obtained for functions in $\mathcal{Z}(\mathcal{Q})$ can be applied directly to justify forming finite sample approximations to the moment condition $C1$.

Consider next moment conditions in the set $C2$. Recall from Section 3 that these conditions are of the form:

$$E[\nu(x_{t+1}, x_t)] = 0 \tag{7.1}$$

for functions $\nu : \mathrm{R}^{2n} \rightarrow \mathrm{R}$. The function ν was not necessarily restricted so that the

[13] Let the pair of functions (ϕ_r, ϕ_i) in $\mathcal{D} \cap \mathcal{Z}(\mathcal{Q})$ satisfy the eigenvalue problem for purely imaginary eigenvalue $i\theta$. Thus (ϕ_r, ϕ_i) solves the analogous eigenvalue problem with eigenvalue $exp(i\theta t)$ for $\mathcal{T}_t$ and consequently, $\| \mathcal{T}_t\phi_r \|^2 + \| \mathcal{T}_t\phi_i \|^2 = \| \phi_r \|^2 + \| \phi_i \|^2$. Since $\| \phi_r \|^2 = E[|\, \phi_r(x_t) - \mathcal{T}_t\phi_r(x_0) \,|^2] + \| \mathcal{T}_t\phi_r \|^2 = 0$, by the contraction property $\| \phi_r \|^2 = \| \mathcal{T}_t\phi_r \|^2 = 0$, and similarly for ϕ_i.

random variable $\nu(x_{t+1}, x_t)$ has a finite second moment, although this random variable will always have finite first moment. Ergodicity of the sampled process is sufficient for the sample averages to converge to zero almost surely ($\mathcal{P}r$). Under the additional restriction that the random variable $\nu(x_{t+1}, x_t)$ has a finite second moment, it follows from the Mean Ergodic Theorem that the sample averages will also converge to zero in $\mathcal{L}^2(\mathcal{P}r)$.

7.2. Central Limit Theorem

To obtain central limit approximations for discrete time Markov processes, Rosenblatt (1971) suggested restricting the operator $\mathcal{T}$ to be a strong contraction on $\mathcal{Z}(\mathcal{Q})$. Among other things, this limits the temporal dependence sufficiently for discrete-sampled process to be strongly mixing (see Rosenblatt 1971, Lemma 3, page 200). The way in which the central limit approximations are typically obtained is through martingale approximations. Explicit characterization of these martingale approximations is of independent value for investigating the statistical efficiency of classes of generalized method of moments estimators constructed from infinite-dimensional families of moment conditions (*e.g.*, see Hansen 1985) such as those derived in Section 3.

Since our goal in this subsection is to deduce restrictions on $\mathcal{A}$ that are sufficient for martingale approximations to apply, we begin by investigating martingale approximations for the original continuous time process. This will help to motivate restrictions on $\mathcal{A}$ that are sufficient for $\mathcal{T}$ to be a strong contraction. In studying moment conditions $C1$ using a continuous record, we use the standard argument of approximating the integral $\int_0^T \mathcal{A}\phi(x_t)dt$ by a martingale m_T, and applying a Central Limit Theorem for martingales.[14] For each $T \geq 0$, define:

$$m_T = -\phi(x_T) + \phi(x_0) + \int_0^T \mathcal{A}\phi(x_t)dt. \tag{7.2}$$

Then $\{m_T : T \geq 0\}$ is a martingale, relative to the filtration generated by the continuous-time process $\{x_t\}$, (*e.g.* , see Ethier and Kurtz 1986, Proposition 1.7, page 162). The error in approximating $\int_0^T \mathcal{A}\phi(x_t)dt$ by m_T is just $-\phi(x_T) + \phi(x_0)$ which is bounded by $2 \parallel \phi \parallel$. When scaled by $(1/\sqrt{T})$, this error clearly converges in $\mathcal{L}^2(\mathcal{P}r\ )$ to zero. Consequently, a Central Limit Theorem for $\{(1/\sqrt{T})\int_0^T \mathcal{A}\phi(x_t)dt\}$ can be deduced from the Central Limit Theorem for scaled sequence of martingales $\{(1/\sqrt{T})m_T\}$ (see Billingsley 1961).

The random variable m_T has mean zero and variance:

$$\begin{aligned} E[(m_T)^2] &= \sum_{j=1}^{J} E[(m_{T(J-j+1)/J} - m_{T(J-j)/J})^2] \\ &= JE[(m_{T/J})^2] \\ &= TE[(J/T)(-\phi(x_{T/J}) + \phi(x_0) + \int_0^{T/J} \mathcal{A}\phi(x_t)dt)^2] \end{aligned} \tag{7.3}$$

for any positive integer J since the increments of the martingale are stationary and orthogonal. Thus we are led to investigate the limit

$$\lim_{\epsilon \to 0} E[(1/\epsilon)(-\phi(x_\epsilon) + \phi(x_0) + \int_0^\epsilon \mathcal{A}\phi(x_t)dt)^2]. \tag{7.4}$$

By the Triangle Inequality

$$\begin{aligned} \{E[(1/\epsilon)^{1/2} \int_0^\epsilon \mathcal{A}\phi(x_t)dt]^2\}^{1/2} &\leq \int_0^\epsilon E[(1/\epsilon)\mathcal{A}\phi(x_t)^2]^{1/2}dt \\ &= \epsilon^{1/2} \parallel \mathcal{A}\phi \parallel \end{aligned} \tag{7.5}$$

Thus the limit in (7.4) can be written as

$$\begin{aligned} \lim_{\epsilon \to 0}(1/\epsilon)E\{[\phi(x_\epsilon) - \phi(x_0)]^2\} &= \lim_{\epsilon \to 0}(1/\epsilon)2[< \phi \mid \phi > - < \mathcal{T}_\epsilon\phi \mid \phi >] \\ &= -2 < \phi \mid \mathcal{A}\phi > . \end{aligned} \tag{7.6}$$

[14] The Central Limit Theorem for diffusions on the line appears in Mandl (1968). Our argument generalizes in a straightforward way the reasoning in Florens-Zmirou (1984), who gave an alternative proof to Mandl's and derived a counterpart for the discretized version of some diffusions on the line.

Taking limits on right side of (7.3) as J gets large and substituting from (7.6), we see find that

$$E[(m_T)^2] = -2T < \phi \mid \mathcal{A}\phi > . \tag{7.7}$$

Therefore, the asymptotic variance for the central limit approximation is given by $-2 < \phi \mid \mathcal{A}\phi >$ which is always nonnegative due to the fact that $\mathcal{A}$ is quasi negative semidefinite (Property $P9$).

The fact that the right side of (7.2) is a martingale guarantees that the continuous-time Central Limit Theorem can always be applied to functions $\psi = \mathcal{A}\phi$. Note, however, that the limiting distribution is nondegenerate only when $< \phi \mid \mathcal{A}\phi > < 0$.

We now investigate the discrete-time counterpart to this martingale approximation under the restriction that $\mathcal{T}$ is a ***strong contraction*** on $\mathcal{Z}(Q)$, *i.e.* there exists constant $C < 1$ such that $\| \mathcal{T}\phi \| \leq C \| \phi \|$ for each $\phi \in \mathcal{Z}(Q)$. More precisely, we will show that a martingale M_N approximates $\sum_{t=1}^{N} \psi(x_t)$ for ψ in $\mathcal{Z}(Q)$. Of course, the ψ*'s* we are interested in are the ones constructed by applying $\mathcal{A}$ to an element of its domain $\mathcal{D}$.

The strong contraction property of $\mathcal{T}$ guarantees that $(\mathcal{I} - \mathcal{T})$ has a bounded inverse on $\mathcal{Z}(Q)$. Note that since

$$E[\psi(x_{t+1}) - \mathcal{T}\psi(x_t) \mid x_t] = 0,$$

the discrete-time process $\{M_N : N = 1, 2, \ldots\}$, defined by:

$$M_N \equiv \sum_{t=1}^{N} [(\mathcal{I} - \mathcal{T})^{-1}\psi(x_t) - \mathcal{T}(\mathcal{I} - \mathcal{T})^{-1}\psi(x_{t-1})]. \tag{7.8}$$

is a martingale adapted to the filtration generated by the discrete-sampled Markov process. Equivalently, we may write

$$M_N = \mathcal{T}(\mathcal{I} - \mathcal{T})^{-1}[\psi(x_N) - \psi(x_0)] + \sum_{t=1}^{N} \psi(x_t).$$

which is the discrete time counterpart to (7.2) and agrees with martingale approximation suggested by Gordin (1969). Note that the $\mathcal{L}^2(\mathcal{P}r)$ norm of the the error in approximating the partial sum $\sum_{t=1}^{N} \psi(x_t)$ by the martingale M_N has a bound independent of N. This bound is uniform for ψ in the unit ball of $\mathcal{Z}(Q)$ since

$$(E\{|\mathcal{T}(\mathcal{I}-\mathcal{T})^{-1}[\psi(x_N)-\psi(x_0)]|^2\})^{1/2} \leq 2 \| \mathcal{T}(\mathcal{I}-\mathcal{T})^{-1}(\psi) \|$$

and $\mathcal{T}$ and $(\mathcal{I}-\mathcal{T})^{-1}$ are bounded operators on $\mathcal{Z}(Q)$. Scaling by $(1/\sqrt{N})$ makes the approximation error arbitrarily small as N goes to infinity implying that central limit approximations for $\{(1/\sqrt{N})\sum_{t=1}^{N} \psi(x_t)\}$ can be deduced from central limit approximations for scaled sequence of martingales $\{(1/\sqrt{N})M_N\}$ (see Billingsley 1961). Finally, it follows from (7.8) that

$$\begin{aligned} (1/N)E(M_N^2) &= \\ <(\mathcal{I}-\mathcal{T})^{-1}\psi \mid (\mathcal{I}-\mathcal{T})^{-1}\psi> &- <\mathcal{T}(\mathcal{I}-\mathcal{T})^{-1}\psi \mid \mathcal{T}(\mathcal{I}-\mathcal{T})^{-1}\psi> \\ &= <\psi \mid \psi> + 2 <\psi \mid \mathcal{T}(I-\mathcal{T})^{-1}\psi> . \quad (7.9) \end{aligned}$$

which gives the asymptotic variance for the discrete-time Central Limit Theorem. It can be shown that for $\psi = \mathcal{A}\phi$, the expression on the right side of (7.9) is greater than or equal to the corresponding expression $-2 < \phi \mid \mathcal{A}\phi >$ for the continuous-time martingale approximation. This reflects the loss of information due to sampling in discrete time.

The discrete-time martingale approximation given by equation (7.8) can also be applied to moment conditions in the class $C2$. As we argued previously, these moment conditions can be represented as in (7.1). Suppose that the random variable $\nu(x_{t+1}, x_t)$ has a finite second moment. Then there exists a ψ in $\mathcal{Z}(Q)$ such that

$$E[\nu(x_{t+1}, x_t) \mid x_t] = \psi(x_t).$$

Then a martingale approximator for $\sum_{t=1}^{N} \nu(x_{t+1}, x_t)$ is given by the sum of the martingale $\sum_{t=1}^{N}[\nu(x_{t+1}, x_t) - \psi(x_t)]$ and the martingale approximator for $\sum_{t=1}^{N} \psi(x_t)$.

We now consider restrictions on $\mathcal{A}$ that are sufficient for $\mathcal{T}$ to be a strong contraction. We write var(ϕ) for the variance of the random variable $\phi(x_t)$.

Condition G: There exists a subspace $\mathcal{N}$, a core for $\mathcal{A}$, and a $\delta > 0$ such that $- < \mathcal{A}\phi \mid \phi > \geq \delta \text{ var}(\phi)$ for all $\phi \in \mathcal{N}$.

Notice that since $\mathcal{N}$ is a core, Condition G can be extended to any $\phi \in \mathcal{D}$. Also, since $\phi - \int \phi dQ \in \mathcal{Z}(Q)$, Condition G is equivalent to requiring

$$- < \mathcal{A}\phi \mid \phi > \geq \delta < \phi \mid \phi > \text{ for all } \phi \text{ with mean zero.}$$

Proposition 7.3. *The operator $\mathcal{T}$ is a strong contraction if and only if the generator $\mathcal{A}$ satisfies Condition G.*

Proof: As shown by Banon (1977, Lemma 3.11, page 79), if $\mathcal{T}_t$ is a strong contraction for any fixed $t > 0$, then the semigroup $\{\mathcal{T}_t\}$ must satisfy the *exponential* inequality for some strictly positive δ:

$$\| \mathcal{T}_t(\phi) \| \leq \exp(-\delta t) \| \phi \| \text{ for all } \phi \in \mathcal{Z}(Q).$$

Hence $\mathcal{S}_t \equiv e^{\delta t}\mathcal{T}_t$ defines a contraction semigroup in $\mathcal{Z}(Q)$ with a generator $\mathcal{A} + \delta I$. By *P9*, Condition G holds. Conversely if Condition G holds, $\mathcal{A} + \delta I$ is a quasi negative semidefinite operator with domain $\mathcal{D} \cap \mathcal{Z}(Q)$, and such that for any $\lambda > 0$, $\lambda I - (\mathcal{A} + \delta I)$ is onto. Hence, by the Lumer- Phillips Theorem, $\mathcal{A} + \delta I$ is the generator of a contraction semigroup, $\{\mathcal{S}_t\}$, in $\mathcal{Z}(Q)$ and since $\mathcal{T}_t = e^{-\delta t}\mathcal{S}_t$, it satisfies the exponential inequality. *Q.E.D.*

For reversible generators, Condition G is equivalent to zero being an isolated point in the support of $\mathcal{E}$, the resolution of the identity used in Section 5. In particular for diffusions with reflecting boundaries on a compact interval with strictly positive diffusion coefficient, condition G holds. Condition G requires that the variances of the continuous-time martingale approximators (and hence the discrete-time approximators) be bounded away from zero for test functions ϕ with unit variances. In particular, when Condition G is satisfied, the central limit approximation will be nondegenerate whenever

ϕ in $\mathcal{D}$ is not constant. Also, Condition G ensures that $\| \mathcal{A}(\phi) \|$ is bounded away from zero on the unit sphere and hence $\mathcal{A}^{-1}$ is a bounded operator on $\mathcal{A}(\mathcal{D})$.

We now study restrictions on $\mathcal{A}$ that imply Condition G for the examples given in Section 4. For the Markov jump process (Example 4.1) sufficient condition is that λ is bounded away from zero and the conditional expectation operator $\tilde{\mathcal{T}}$ on the associated chain is a strong contraction on $\mathcal{Z}(\tilde{Q})$. To see this, first note that for $K_1 \equiv \int \phi d\tilde{Q}$ and $K_2 \equiv \int (1/\lambda) d\tilde{Q}$,

$$\begin{aligned} < \phi \mid \mathcal{A}\phi > &= \int \phi[\tilde{\mathcal{T}}(\phi) - \phi]\lambda dQ \\ &= (1/K_2) \int \phi[\tilde{\mathcal{T}}(\phi) - \phi] d\tilde{Q} \\ &= (1/K_2) \int (\phi - K_1)[\tilde{\mathcal{T}}(\phi - K_1) - (\phi - K_1)] d\tilde{Q} \\ &\leq (1/K_2)(\gamma - 1) \int (\phi - K_1)^2 d\tilde{Q} \end{aligned} \tag{7.10}$$

for some $0 < \gamma < 1$ since $\tilde{\mathcal{T}}$ is a strong contraction on $\mathcal{Z}(\tilde{Q})$. Next observe that for $\lambda_\ell \equiv \inf \lambda > 0$:

$$\begin{aligned} (1/K_2)(\gamma - 1) \int (\phi - K_1)^2 d\tilde{Q} &\leq (\gamma - 1)\lambda \int \int (\phi - K_1)^2 dQ \\ &\leq (\gamma - 1)\lambda_\ell < \phi \mid \phi >, \end{aligned} \tag{7.11}$$

as long as $\phi \in \mathcal{Z}(Q)$. The operator $\mathcal{A}$ in conjunction with any positive δ less than $(1 - \gamma)\lambda_\ell$ will satisfy Condition G.

We now turn to Markov diffusion processes (Examples 4.2-4.4). Bouc and Pardoux (1984) provided sufficient conditions for G that include uniformly bounding above and below the diffusion matrix Σ as well as a pointing inward condition for the drift μ. To accommodate certain examples in finance where the diffusion coefficient may vanish at the boundary, it is necessary to relax the assumptions on the diffusion coefficient. For this reason we derive here an alternative set of sufficient conditions on the line.

Consider a scalar diffusion on an interval $[\ell, u]$, allowing for $\ell = -\infty$ and $u = +\infty$. We restrict σ^2 to be positive in the open interval (ℓ, u) and require the speed density to be integrable. If a finite boundary is attainable, we assume that it is reflexive. We start by presenting an alternative characterization of Condition G for diffusions. Let $\mathcal{M}$ be the space of twice continuously differentiable functions $\phi \in \mathcal{L}^2(Q)$ such that $\sigma\phi'$ and $L\phi$ are also in $\mathcal{L}^2(Q)$. In the case of a reflexive boundary we also add the restriction that ϕ' vanishes at the boundary.

For $\phi \in \mathcal{M}$ it follows from the characterization of the infinitesimal generator established in Example 4.2 and equation (7.2), that the martingale approximator $\{m_T\}$ satisfies:

$$m_T = -\int_0^T \phi'(x_t)\sigma(x_t)dW_t$$

Since $\{x_t\}$ is stationary, (7.7) implies that

$$-2 < \phi - \int \phi dQ \mid \mathcal{A}(\phi) > = \int (\phi')^2\sigma^2 dQ.$$

Therefore, Condition G requires for some $\eta > 0$:

$$\int (\phi')^2\sigma^2 dQ \geq \eta \text{ var}(\phi) \text{ for all } \phi \in \mathcal{M}. \tag{7.12}$$

When $\mathcal{M}$ is a core condition (7.12) is equivalent to Condition G. Recall that in Section 6 we discussed sufficient conditions for C_K^2 and hence $\mathcal{M}$ to be a core.[15]

We now derive sufficient conditions for inequality (7.12). Let $z \in (\ell, u)$ and, notice that, for any ϕ in $\mathcal{M}$,

$$\int [\phi - \phi(z)]^2 dQ \geq \text{ var}(\phi).$$

It follows from an inequality in Muckenhoupt (1972, Theorem 2) or Talenti (1969, page 174) that there is a finite K_1 such that

$$\int_z^u [\phi - \phi(z)]^2 dQ \leq K_1 \int_z^u \mid \phi' \mid^2 \sigma^2 dQ$$

[15] Inequality (7.12) has the obvious multivariate extension.

if, and only if

$$\sup_{x<r<u} Q\{[r,u)\} \int_x^r (\sigma^2 q)^{-1}\, dy < \infty. \tag{7.13}$$

Hence (7.12) will hold provided (7.13) and the analogous condition

$$\sup_{\ell<r<x} Q\{(\ell,r]\} \int_r^x (\sigma^2 q)^{-1}\, dy < \infty. \tag{7.14}$$

are satisfied.

Applying L'Hospital's Rule, if:

$$\lim_{r\to u} Q\{[r,u)\}/\sigma(r)q(r) < \infty, \tag{7.15}$$

then (7.13) holds. In particular inequality (7.15) holds if $\liminf_{r\to u} \sigma(r)q(r) > 0$. If $\lim_{r\to u} \sigma(r)q(r) = 0$, we may apply L'Hospital's Rule again to obtain as a sufficient condition:

$$\lim_{r\to u} \frac{1}{\sigma'(r) + \sigma(r)q'(r)/q(r)} \text{ exists and is finite.} \tag{7.16}$$

Using equation (5.3) that relates the logarithmic derivative of the stationary density to the drift and diffusion coefficients we may rewrite (7.16) as:

$$\lim_{r\to u} \frac{\sigma(r)}{2\mu(r) - \sigma(r)\sigma'(r)} \text{ exists and is finite.} \tag{7.17}$$

The conditions at the lower boundary are exactly analogous. We summarize this discussion in a proposition:

Proposition 7.4. *Suppose $\{x_t\}$ solves $dx_t = \mu(x_t)dt + \sigma(x_t)dW_t$ in an interval (ℓ, u) with possibly $\ell = -\infty$ and/or $u = +\infty$; the diffusion coefficient is C^1 and positive on (ℓ, u); and the speed density $1/s\sigma^2$ is integrable. Then the following conditions are sufficient for (7.12) to hold.*

(a) The right boundary satisfies either $\liminf_{r\to u} \sigma(r)q(r) > 0$; or $\liminf_{r\to u} \sigma(r)q(r) = 0$ and $\lim_{r\to u} \sigma(r)/[2\mu(r) - \sigma(r)\sigma'(r)]$ exists and is finite.

(b) The left boundary satisfies either $\liminf_{r\to\ell}\sigma(r)q(r) > 0$*; or* $\liminf_{r\to\ell}\sigma(r)q(r) = 0$ *and* $\lim_{r\to\ell}\sigma(r)/[2\mu(r) - \sigma(r)\sigma'(r)]$ *exists and is finite.*

Notice that for models that are parameterized in terms q and σ^2, condition (7.16) is easier to verify than the equivalent condition (7.17) mentioned in the proposition. Also one may, in some cases, verify directly (7.15).

The square root process $dx_t = \kappa(x_t - \bar{x}) + \sqrt{\bar{x}}dW_t$, with $\kappa < 0$ and $\bar{x} > 0$, is an example of a process that satisfies the sufficient conditions of Proposition 7.4 even though σ^2 is not bounded away from zero. A process $\{x_t\}$ that solves a stochastic differential equation in $\mathbf{R}_+$ with $\sigma(y) \equiv 1$ and, $\mu(y) = -(\sqrt{y})^{-1}$ for large y, is an example where the sufficient conditions of Proposition 7.4 do not hold. However in this case every nonnegative real number is in the spectrum of $\mathcal{A}$ and hence, Condition G fails, even though $\{x_t\}$ is mean recurrent (Bouc.and Pardoux, 1984, page 378.)

The conclusion of Proposition 7.4 also holds for the multivariate factor models described in Example 4.3 when the individual factor processes satisfy the specified conditions.

8. Conclusion

The analysis in this paper is intended as support of empirical work aimed at assessing the empirical plausibility of particular continuous-time Markov models that arise in a variety of areas of economics. In many instances, such models are attractive because of conceptual and computational simplifications obtained by taking continuous time limits. The approach advanced in this paper can, by design, be used to study these models empirically even when it is not possible for an econometrician to approximate a continuous data record. For this reason, we focused our analysis on fixed interval sampling although our moment conditions are also applicable more generally. For instance, we could accommodate systematic patterns to the sampling, or the sampling procedure itself could be modeled as an exogenous stationary process. Both moment conditions are still satisfied, with moment condition $C2$ applied to adjacent observations. The observable implications we obtained extend in the obvious way. This means that our moment conditions can easily handle missing observations that occur in financial data sets due to weekends and holidays. On the other hand, such sampling schemes may alter the central limit approximations reported in Section VII.

One of the many questions left unanswered here is that of the selection of test functions in practice. For finite-dimensional parameter models one could compare the asymptotic efficiency of estimators constructed with alternative configurations of test functions along the lines of Hansen (1985). Finite sample comparisons are likely to require Monte Carlo investigations.

9. References

Anderson, B.D.O. (1982), "Reverse-Time Diffusion Equation-Models," *Stochastic Processes and Their Applications* 12, 313-326.

Banon, G. (1977), "Estimation Non Parametrique de Densite de Probabilite Pour les Processus de Markov," These, Universite Paul Sabatier de Toulouse, France.

Banon, G. (1978), "Nonparametric Identification for Diffusion Processes," *SIAM Journal of Control and Optimization* 16, 380-395.

Bequillard, A.L. (1989), "A Sufficient Condition for Two Markov Semigroups to Commute," *The Annals of Probability* 17, 1478-1482.

Bierens, H.J. (1990), "A Consistent Conditional Moment Test of Functional Form," *Econometrica* 58, 1443-1458.

Billingsley, P. (1961), "The Lindeberg-Levy Theorem for Martingales," *American Mathematical Society* 12, 788-792.

Bouc, R., and E. Pardoux (1984), "Asymptotic Analysis of P.D.E.s with Wide-Band Noise Disturbances, and Expansion of the Moments" *Stochastic Analysis and Applications*, 2(4), 369-422.

Brezis, H. (1983), *Analyse Fonctionnelle, Theorie et applications*, Masson, Paris.

Chamberlain, G. (1992), "Efficiency Bounds for Semiparametric Regression," *Econometrica* 60, 567-596.

Chung, K.L. (1974), *A Course in Probability Theory*, 2nd ed., Academic Press, New York.

Cobb, L., P. Koopstein and N.Y. Chen (1983), "Estimation and Moment Recursion Relations for Multimodal Distributions of the Esponential Family," *Journal of the American Statistical Association* 78, 124-130.

Cox, J., J. Ingersoll, and S. Ross (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica* 53, 385-408.

Duffie, D. and R. Kan (1993),"A Yield-Factor Model of Interest Rates," manuscript.

Duffie, D. and K.J. Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica* 61, 929-952.

Dunford, N. and J.T. Schwartz (1988), *Linear Operators Part One: General Theory*, New York: John Wiley & Sons.

Dynkin, E.B. (1965), *Markov Processes*, Berlin: Springer-Verlag.

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

Feller, W. (1951), "Two Singular Diffusion Problems," *Annals of Mathematics* 54, 173-182.

Florens-Zmirou, D. (1984), "Théoréme de Limite Central pour une Diffusion et pour sa Discrètisée," *C.R. Acad. Sc. Paris* t. 299, Serie I, n 19.

Frachot, A., D. Janci, and V. Lacoste (1992), "Factor Analysis of the Term Structure: A Probabilistic Approach," Banque de France NER # 21.

Froot, K.A., and M. Obstfeld (1991), "Exchange-Rate Dynamics Under Stochastic Regime Shifts: A Unified Approach," *Journal of International Economics* 31, 203-229.

Gallant, A.R., and D.W. Nychka (1987), "Semi-nonparametric Maximum Likelihood Estimation," *Econometrica* 55, 363-390.

Gordin, M.I. (1969), "The Central Limit Theorem for Stationary Processes," *Soviet Mathematics Doklady* 10, 1174-1176.

Halmos, P.R. (1974), *Measure Theory*, New York: Springer-Verlag.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.

Hansen, L.P. (1985), "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics* 30, 203-238.

Hansen, L.P. and T.J. Sargent (1983), "The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities," *Econometrica* 51, 377-387.

Harvey, A.C. and J.H. Stock (1985), "The Estimation of Higher Order Continuous Time Autoregressive Models," *Journal of Econometric Theory* 1, 97-112.

Has'minskii, R.Z. (1980), *Stochastic Stability of Differential Equations* , The Netherlands: Sijthoff & Noordhoff.

Hausmann, U. and E. Pardoux (1985), "Time Reversal of Diffusion Processes," *Stochastic Differential Systems: Filtering and Control*, ed. by M. Metivier and E. Pardoux, Berlin: Springer-Verlag, 176-183.

He, H. (1990), "Moment Approximation and Estimation of Diffusion Models of Asset Prices," manuscript.

Heath, D., R. Jarrow, and A. Morton (1990), "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation," forthcoming in *Econometrica*.

Johansen, S. (1973), "A Central Limit Theorem for Finite Semigroups and Its Application to the Imbedding Problem for Finite State Markov Chains," *Zeitschrift fur Wahrscheinlichkeitsetheorie und Verwandte Gebiete* 26, 171-190

Karatzas, I. and S. Shreve (1988), *Brownian Motion and Stochastic Calculus*, New York, Springer-Verlag.

Karlin, S. and H.M. Taylor (1981), *A Second Course in Stochastic Processes*, New York: Academic Press.

Kent, J. (1978), "Time-Reversible Diffusions," *Advance Applications Problability*, 10, 819-835.

Krugman, P.R. (1991), "Target Zones and Exchange Rate Dynamics," *The Quarterly Journal of Economics* 106, 669-682.

Lewbel, A. (1991), "Consistent Tests of Nonparametric Hypotheses: Theory," Brandeis University manuscript.

Levy, J. (1993), "Two Empirical Tests of Exchange Rate Target Zones," Ph.D. Thesis, University of Chicago.

Lo, A. (1988), "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data," *Journal of Econometric Theory* 4, 231-247.

Longstaff, F.A., and E.S. Schwartz (1992), "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *The Journal of Finance* XLVII:4, September, 1259-81.

Mandl, P. (1968), *Analytical Treatment of One-dimensional Markov Processes*, Academia, Publishing House of the Czechoslovak Academy of Sciences, Prague.

Millet, A., D. Nualart, and M. Sanz (1989), "Integration by Parts and Time Reversal for Diffusion Processes," *The Annals of Probability* 17, 208-238.

Muckenhoupt, B. (1972), "Hardy's Inequality with Weights" *Studia Mathematica* 44, 31-38.

Nelson, E. (1958), "The Adjoint Markoff Process," *Duke Mathematical Journal* 25, 671-690.

Newey, W.K. (1990), "Efficient Estimation of Semiparametric Models Via Moment Restrictions," Bellcore Economics Discussion Paper #65.

Newey, W.K. (1993), "The Asymptotic Variance of Semiparametric Estimators," MIT manuscript.

Pazy, A. (1983), *Semigroups of Linear Operators and Applications to Partial Differential Equations*, New York: Springer-Verlag.

Phillips, A.W. (1959), "The Estimation of Parameters in Systems of Stochastic Differential Equations," *Biometrika* 59, 67-76.

Phillips, P.C.B. (1973), "The Problem of Identification in Finite Parameter Continuous Time Models," *Journal of Econometrics* 1, 351-362.

Powel, J.L., J.H. Stock, and T.M. Stoker (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica* 57, 1403-1430.

Robinson, P.M. (1976), "The Estimation of Linear Differential Equations with Constant Coefficients," *Econometrica* 44, 751-764.

Robinson, P.M. (1989), "Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Review of Economic Studies* 56, 511-534.

Rosenblatt, M. (1969), "Density Estimates and Markov Sequences," *Nonparametric Techniques in Statistical Inference*, ed. M.L. Puri, Cambridge: Cambridge University Press, 199-210.

Rosenblatt, M. (1971), *Markov Processes: Structure and Asymptotic Behavior*, Berlin: Springer-Verlag.

Roussas, G.G. (1969), "Nonparametric Estimation in Markov Processes," *Annals of the Institute of Statistical Mathematics* 21, 73-87.

Rudin, W. (1973), *Functional Analysis*, New York: McGraw-Hill Book Co.

Scheinkman, J.A. (1990), "Nonlinearities in Economic Dynamics," *The Economic Journal* 100, 33-48.

Singer, B. and S. Spillerman (1976), "The Representation of Social Processes by Markov Models," *American Journal of Sociology* 82, 1-54.

Talenti, G. (1969), "Osservazioni Sopra Una Classe di Disuguaglianze," *Rendiconti del Seminario di Matematica e Fisica di Milano* 39, 171-185.

Wong, E. (1964), "The Construction of a Class of Stationary Markoff Processes," in *Sixteenth Symposium in Applied Mathematics - Stochastic Processes in Mathematical Physics and Engineering*, ed. R. Bellman, Providence, RI, American Mathematical Society, 264-276.