

# Operator Methods for Continuous-Time Markov Processes\*

Yacine Aït-Sahalia	Lars Peter Hansen	José A. Scheinkman
Department of Economics	Department of Economics	Department of Economics
Princeton University	The University of Chicago	Princeton University

First Draft: November 2001. This Version: June 2, 2004

## 1 Introduction

Our chapter surveys a set of mathematical and statistical tools that are valuable in understanding and characterizing nonlinear Markov processes. Such processes are used extensively as building blocks in economics and finance. In these literatures, typically the local evolution or short-run transition is specified. We concentrate on the continuous limit in which case it is the instantaneous transition that is specified. In understanding the implications of such a modelling approach we show how to infer the intermediate and long-run properties from the short-run dynamics. To accomplish this we describe operator methods and their use in conjunction with continuous-time stochastic process models.

Operator methods begin with a local characterization of the Markov process dynamics. This local specification takes the form of an *infinitesimal generator*. The infinitesimal generator is itself an operator mapping test functions into other functions. From the infinitesimal generator, we construct a family (semigroup) of conditional expectation operators. The operators exploit the time-invariant Markov structure. Each operator in this family is indexed by the forecast horizon, the interval of time between the information set used for prediction and the object that is being predicted. Operator methods allow us to ascertain global, and in particular, long-run implications from the local or infinitesimal evolution. These global implications are reflected in a) the implied stationary distribution b) the analysis of the eigen-

---

\*All three authors gratefully acknowledge financial support from the National Science Foundation.

functions of the generator that dominate in the long run, c) the construction of likelihood expansions and other estimating equations.

The methods we describe in this chapter are designed to show how global and long-run implications follow from local characterizations of the time series evolution. This connection between local and global properties is particularly challenging for nonlinear time series models. In spite of this complexity, the Markov structure makes such characterizations tractable.

Markov models are designed to be convenient models of nonlinear stochastic processes. We show how operator methods can contribute to useful characterizations of dynamic evolution and approximations of a likelihood function. For many purposes, the Markov models used in practice are formally misspecified as complete descriptions of the time series evolution. Thus formal statistical methods based on likelihood inference or its Bayesian counterpart, while useful, may often not be the most valuable for exploring the nature of the misspecification. Misspecified models are and will continue to be used in practice because they are often built to be pedagogically simple and analytically tractable. Operator methods are useful in exploring how specific families of Markov models might be misspecified.

Section 2 describes the underlying mathematical methods and notation. Section 3 studies Markov models through their implied stationary distributions. Section 4 gives some operator characterizations and related expansions used to characterize transition dynamics. Section 5 investigates alternative ways to characterize the observable implications of various Markov models. Section 6 shows how these methods can be used in the context of some specific applications.

## 2 Alternative Ways to Model a Continuous-Time Markov Process

There are several different but essentially equivalent ways to parameterize continuous time Markov processes, each leading naturally to a distinct estimation strategy. In this section we briefly describe five possible parametrizations.

### 2.1 Transition Functions

In what follows,  $(\Omega, \mathcal{F}, Pr)$  will denote a probability space,  $S$  a locally compact metric space with a countable basis,  $\mathcal{S}$  a  $\sigma$ -field of Borelians in  $S$ ,  $I$  an interval of the real line, and for each  $t \in I$ ,  $X_t : (\Omega, \mathcal{F}, Pr) \rightarrow (S, \mathcal{S})$  a measurable function. We will refer to  $(S, \mathcal{S})$  as the state space and to  $X$  as a stochastic process.

**Definition 1.**  $P : (S \times \mathcal{S}) \rightarrow [0, 1)$  is a transition probability if  $P(x, \cdot)$  is a probability measure

in  $\mathcal{S}$ , and  $P(\cdot, B)$  is measurable, for each  $(x, B) \in (S \times \mathcal{S})$ .

**Definition 2.** A transition function is a family  $P_{s,t}, (s, t) \in I^2, s < t$  that satisfies for each  $s < t < u$  the Chapman-Kolmogorov equation:

$$P_{s,u}(x, B) = \int P_{t,u}(y, B)P_{s,t}(x, dy).$$

A transition function is time homogeneous if  $P_{s,t} = P_{s',t'}$  whenever  $t - s = t' - s'$ . In this case we write  $P_{t-s}$  instead of  $P_{s,t}$ .

**Definition 3.** Let  $\mathcal{F}_t \subset \mathcal{F}$  be an increasing family of  $\sigma$ -algebras, and  $X$  a stochastic process that is adapted to  $\mathcal{F}_t$ .  $X$  is Markov with transition function  $P_{s,t}$  if for each non-negative Borel measurable  $\phi : S \rightarrow \mathbb{R}$  and each  $(s, t) \in I^2, s < t$ ,

$$E[\phi(X_t)|\mathcal{F}_s] = \int \phi(y)P_{s,t}(X_s, dy).$$

The following standard result (for example, Revuz and Yor (1991), Chapter 3, Theorem 1.5) allows one to parameterize Markov processes using transition functions.

**Theorem 1.** Given a transition function  $P_{s,t}$  on  $(S, \mathcal{S})$  and a probability measure  $Q_0$  on  $(S, \mathcal{S})$ , there exists a unique probability measure  $Pr$  on  $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$ , such that the coordinate process  $X$  is Markov with respect to  $\sigma(X_u, u \leq t)$ , with transition function  $P_{s,t}$  and the distribution of  $X_0$  given by  $Q_0$ .

We will interchangeably call transition function the measure  $P_{s,t}$  or its conditional density  $p$  (subject to regularity conditions which guarantee its existence):

$$P_{s,t}(x, dy) = p(y, t|x, s)dy.$$

In the time homogenous case, we write  $\Delta = t - s$  and  $p(y|x, \Delta)$ .

## 2.2 Semigroup of conditional expectations

Let  $P_t$  be a homogeneous transition function and  $L$  be a vector space of real valued functions such that for each  $\phi \in L, \int \phi(y)P_t(x, dy) \in L$ . For each  $t$  define the conditional expectation operator

$$\mathcal{T}_t\phi(x) = \int \phi(y)P_t(x, dy). \tag{2.1}$$

The Chapman-Kolmogorov equation guarantees that the linear operators  $\mathcal{T}_t$  satisfy:

$$\mathcal{T}_{t+s} = \mathcal{T}_t\mathcal{T}_s. \tag{2.2}$$

This suggests another parameterization for Markov processes. Let  $(L, \|\cdot\|)$  be a Banach space.

**Definition 4.** A one-parameter family of linear operators in  $L$ ,  $\{\mathcal{T}_t : t \geq 0\}$  is called a strongly continuous contraction semigroup if (a)  $\mathcal{T}_0 = I$ , (b)  $\mathcal{T}_{t+s} = \mathcal{T}_t \mathcal{T}_s$  for all  $s, t \geq 0$ , (c)  $\lim_{t \rightarrow 0} \mathcal{T}_t \phi = \phi$ , and (d)  $\|\mathcal{T}_t\| \leq 1$

If a semigroup represents conditional expectations, then it must be positive, that is, if  $\phi \geq 0$  then  $\mathcal{T}_t \phi \geq 0$ .

Two useful examples of Banach spaces  $L$  to use in this context are:

**Example 1.** Let  $S$  be a locally compact and separable state space. Let  $L = C_0$  be the space of continuous functions  $\phi : S \rightarrow \mathbb{R}$ , that vanish at infinity. For  $\phi \in C_0$  define:

$$\|\phi\|_\infty = \sup_{x \in S} |\phi(x)|.$$

A strongly continuous contraction positive semigroup on  $C_0$  is called a Feller semigroup.

**Example 2.** Let  $Q$  be a measure on a locally compact subset  $S$  of  $\mathbb{R}^m$ . Let  $L^2(Q)$  be the space of all Borel measurable functions  $\phi : S \rightarrow \mathbb{R}$  that are square integrable with respect to the measure  $Q$  endowed with the norm:

$$\|\phi\|_2 = \left( \int \phi^2 dQ \right)^{\frac{1}{2}}.$$

Typically the semigroup of conditional expectations together with the initial distribution determine the finite-dimensional distributions of the Markov process (see *e.g.* Ethier and Kurtz (1986) Proposition 1.6 of chapter 4). There are also many results (*e.g.* Revuz and Yor (1991) Proposition 2.2 of Chapter 3) concerning whether given a contraction semigroup one can construct a homogeneous transition function such that equation (2.1) is satisfied.

## 2.3 Infinitesimal generators

**Definition 5.** The infinitesimal generator of a semigroup  $\mathcal{T}_t$  on a Banach space  $L$  is the (possibly unbounded) linear operator  $\mathcal{A}$  defined by:

$$\mathcal{A}\phi = \lim_{t \downarrow 0} \frac{\mathcal{T}_t \phi - \phi}{t}.$$

The domain  $D(\mathcal{A})$  is the subspace of  $L$  for which this limit exists.

If  $\mathcal{T}_t$  is a strongly continuous contraction semigroup then  $D(\mathcal{A})$  is dense. In addition  $\mathcal{A}$  is closed, that is if  $\phi_n \in D(\mathcal{A})$  converges to  $\phi$  and  $\mathcal{A}\phi_n$  converges to  $\psi$  then  $\phi \in D(\mathcal{A})$  and  $\mathcal{A}\phi = \psi$ . If  $\mathcal{T}_t$  is a strongly continuous contraction semigroup we can reconstruct  $\mathcal{T}_t$  using its infinitesimal generator  $\mathcal{A}$  (*e.g.* Ethier and Kurtz (1986) Proposition 2.7 of Chapter 2). This

suggests using  $\mathcal{A}$  to parameterize the Markov process. The Hille-Yosida theorem (*e.g.* Ethier and Kurtz (1986) Theorem 2.6 of chapter 1) gives necessary and sufficient conditions for a linear operator to be the generator of a strongly continuous, positive contraction semigroup. Necessary and sufficient conditions to insure that the semigroup can be interpreted as a semigroup of conditional expectations are also known (*e.g.* Ethier and Kurtz (1986) Theorem 2.2 of chapter 4).

As described in Example 1, a possible domain for a semigroup is the space  $C_0$  of continuous functions vanishing at infinity on a locally compact state space endowed with the sup-norm. A process is called a multivariate diffusion if its generator  $\mathcal{A}_d$  is an extension of the second-order differential operator:

$$\mu \cdot \frac{\partial \phi}{\partial x} + \frac{1}{2} \text{trace} \left( \nu \frac{\partial^2 \phi}{\partial x \partial x'} \right) \quad (2.3)$$

where the domain of this second order differential operator is restricted to the space of twice continuously differentiable functions with a compact support. The  $\mathbb{R}^m$ -valued function  $\mu$  is called the drift of the process and the positive semidefinite matrix-valued function  $\nu$  is the diffusion matrix. The generator for the Markov jump process is:

$$\mathcal{A}_p \phi = \lambda (\mathcal{J} \phi - \phi)$$

on the entire space  $C_0$ , where  $\lambda$  is a nonnegative function of the Markov state used to model the jump intensity and  $\mathcal{J}$  is the expectation operator for a conditional distribution that assigns probability zero to staying put.

More generally we can construct a semigroup of conditional expectations operators by considering a generator  $\mathcal{A} = \mathcal{A}_d + \mathcal{A}_p$ .<sup>1</sup> There is a corresponding construction for the case where the domain of the semigroup is  $L^2(Q)$ , as in Example 2.

It is typically difficult to completely characterize  $D(\mathcal{A})$  and instead one parameterizes the generator on a subset of its domain that is ‘big enough.’ Since the generator is not necessarily continuous, one cannot simply parameterize the generator in a dense subset of its domain. Instead one uses a *core*, that is a subspace  $N \subset D(\mathcal{A})$  such that  $(N, \mathcal{A}N)$  is dense in the graph of  $\mathcal{A}$ .

## 2.4 Quadratic forms

Suppose  $L = L^2(Q)$  where we have the natural inner product

$$\langle \phi, \psi \rangle = \int \phi(x) \psi(x) dQ.$$

---

<sup>1</sup>This representation can be generalized to allow for a jump measure that is not finite. For example, see Revuz and Yor (1991) Theorem 1.13 of Chapter 7.

If  $\phi \in \mathcal{D}(\mathcal{A})$  and  $\psi \in L^2(Q)$  then we may define the (closed) form

$$f_2(\phi, \psi) = - \langle \mathcal{A}\phi, \psi \rangle .$$

This leads to another way of parameterizing Markov processes. Instead of writing down a generator one starts with a form. As in the case of a generator it is typically not easy to fully characterize the domain of the form. For this reason one starts by defining a form on a smaller space and showing that it can be extended to a closed form in a subset of  $L^2(Q)$ . When the Markov process can be initialized to be stationary, the measure  $Q$  is typically this stationary distribution. More generally,  $Q$  does not have to be a finite measure.

This approach to Markov processes was pioneered by Beurling and Deny (1958) and Fukushima (1971) for symmetric Markov processes. In this case both the operator  $\mathcal{A}$  and the form  $f$  are symmetric. A stationary, symmetric Markov process is *time-reversible*. If time were reversed, the transition operators would remain the same. On the other hand, multivariate standard Brownian motion is a symmetric (nonstationary) Markov process that is not time reversible. The literature on modelling Markov processes with forms has been extended to the non-symmetric case by Ma and Rockner (1991). In the case of a symmetric diffusion, the form is given by:

$$f_2(\phi, \psi) = \frac{1}{2} \int (\nabla\phi)^* \nu (\nabla\psi) dQ,$$

where  $*$  is used to denote transposition,  $\nabla$  is used to denote the (weak) gradient, and the measure  $Q$  is assumed to be absolutely continuous with respect to Lebesgue measure. The matrix  $\nu$  can be interpreted as the diffusion coefficient. When  $Q$  is a probability measure, it is a stationary distribution. For standard Brownian motion,  $Q$  is Lebesgue measure and  $\nu$  is the identity matrix.

## 2.5 Stochastic differential equations

Another way to generate (homogeneous) Markov processes is to consider solutions to time autonomous stochastic differential equations. Here we start with an  $n$ -dimensional Brownian motion on a probability space  $(\Omega, \mathcal{F}, Pr)$ , and consider  $\{\mathcal{F}_t : t \geq 0\}$ , the (augmented) filtration generated by the Brownian motion. The process  $X_t$  is assumed to satisfy the stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{2.4}$$

$X_0$  given.

Several theorems exist that guarantee that the solution to equation (2.4) exists, is unique and is a Markov diffusion. In this case the coefficients of (2.4) are related to those of the second-order differential operator (2.3) via the formula  $\nu = \sigma\sigma'$ .

### 3 Parametrizations of the Stationary Distribution: Calibrating the Long Run

Over a century ago Karl Pearson (1894) sought to fit flexible models of densities using tractable estimation methods. This led to a method-of-moments approach, an approach that was subsequently criticized by Fisher (1921) on the grounds of statistical efficiency. Fisher (1921) showed that Pearson's estimation method was inefficient relative to maximum likelihood estimation. Nevertheless there has remained a considerable interest in Pearson's family of densities. Wong (1964) provided a diffusion interpretation for members of the Pearson family by producing low-order polynomial models of the drift and diffusion coefficient with stationary densities in the Pearson family. He used operator methods to produce expansions of the transition densities for the processes and hence to characterize the implied dynamics. Wong (1964) is an important precursor to the work that we describe in this and subsequent sections. We begin by generalizing his use of stationary densities to motivate continuous-time models, and we revisit the Fisher (1921) criticism of method-of-moments estimation.

We investigate this approach because modelling in economics and finance often begins with an idea of a target density obtained from empirical observations. Examples are the literature on city sizes, income distribution and the behavior of exchange rates in the presence of bands. In much of this literature, one guesses transition dynamics that might work and then checks this guess. Mathematically speaking this is an inverse problem and is often amenable to formal analysis. As we will see, the *inverse* mapping from stationary densities to the implied transitions or local dynamics can be solved after we specify certain features of the infinitesimal evolution. Wong (1964)'s analysis is a good illustration in which this inverse mapping is transparent. We describe extensions of Wong's approach that exploit the mapping between the infinitesimal coefficients  $(\mu, \sigma^2)$  and the stationary distributions for diffusions.

#### 3.1 Wong's Polynomial Models

To match the Pearson family of densities, Wong (1964) studied the solutions to the stochastic differential equation:

$$dX_t = \varrho_1(X_t)dt + \varrho_2(X_t)^{\frac{1}{2}}dW_t$$

where  $\{X_t\}$  is a scalar diffusion process and  $\{W_t\}$  is a scalar Brownian motion. The polynomial  $\varrho_1$  used to model the drift coefficient is first-order and the polynomial  $\varrho_2$  used to model the diffusion coefficient is no more than second-order. Using arguments we sketch below, the stationary density for this process satisfies the differential equation:

$$(\ln q)' = \frac{2\varrho_1 - \varrho_2'}{\varrho_2} \tag{3.1}$$

where ' denotes differentiation with respect to the state. The logarithmic derivative of the density is the ratio of a first-order to a second-order polynomial as required by Pearson (1894). When the density is restricted to the nonnegative real numbers, we may add a boundary condition that requires the process to reflect at zero.

Wong (1964) identified the diffusion coefficient  $\varrho_2$  up to scale as the denominator of  $(\ln q)'$  expressed as the ratio of polynomials in reduced form. Given  $\varrho_2$  the polynomial  $\varrho_1$  can be constructed from the pair  $((\ln q)', \varrho_2)$  using formula (3.1). In subsection 3.2 we will discuss generalizations of this identification scheme.

Wong (1964) went on to characterize and interpret the stochastic processes whose densities reside in the Pearson class. Many of the resulting processes have been used in economics and finance.

**Example 3.** *When  $\varrho_1$  has a negative slope and  $\varrho_2$  is a positive constant, the implied density is normal and the resulting process is the familiar Ornstein-Uhlenbeck process. This process has been used to model interest rates and volatility. Vasicek (1977) features this process in his construction of an equilibrium model of the real term structure of interest rates.*

**Example 4.** *When  $\varrho_1$  has a negative slope and  $\varrho_2$  is linear with a positive slope, the implied density is gamma and the resulting process is the Feller square-root process. Sometimes zero is an attracting barrier, and to obtain the gamma distribution requires the process to reflect at zero. Cox, Ingersoll, and Ross (1985) feature the Feller square root process in their model of the term structure of interest rates.*

**Example 5.** *When  $\varrho_1$  has a negative slope and  $\varrho_2$  is proportional to  $x^2$ , the stationary density has algebraic tails. This specification is used as a model of volatility and as a model of size distribution. In particular, Nelson (1990) derives this model as the continuous-time limit of the volatility evolution for a GARCH(1,1) model. Nelson (1990) uses the fat (algebraic) tail of the stationary distribution to capture volatility clustering over time.*

*A limiting case of this example also gives a version of Zipf's Law. (See Rapoport (1978) for a nice historical discussion of rank-size relations.) Consider a density of the form:  $q \propto x^{-2}$  defined on  $(y, \infty)$  for  $y > 0$ . This distribution implies that the ranks should be proportional to the size. Notice that the probability of being greater than some value  $x$  is proportional to  $x^{-1}$ . This density satisfies the differential equation:*

$$\frac{d \ln q(x)}{dx} = -\frac{2}{x}.$$

*Zipf's law fits remarkably well the distribution of city sizes. For example, see Auerbach (1913) and Eaton and Eckstein (1997).*

Restrict  $\varrho_2(x) \propto x^2$ . In the context of cities this means that the variance of growth rates is independent of city sizes, which is a reasonable approximation for the data in Japan 1965-1985 and France 1911-1990 discussed in Eaton and Eckstein (1997). (See also Gabaix (1999).) Formula (3.1) implies that

$$(\ln q)' + (\ln \varrho_2)' = \frac{2\varrho_1}{\varrho_2} = 0.$$

Thus the drift is zero and the process is a stationary local martingale. The boundary  $y$  is an attracting barrier which we assume to be reflexive. We will have more to say about this process after we develop spectral tools used in a more refined study of the dynamics.

The density  $q \propto x^{-2}$  has a mode at the left boundary  $y$ . For the corresponding diffusion model,  $y$  is a reflecting barrier. Zipf's Law is typically a statement about the density for large  $x$ , however. Thus we could let the left boundary be at zero (instead of  $y > 0$ ) and set  $\varrho_1$  to a positive constant. The implied density behaves like a constant multiple of  $x^{-2}$  in the right tail, but the zero boundary will not be attainable. The resulting density has an interior mode at one-half times the constant value of  $\varrho_1$ . This density remains within the Pearson family.

**Example 6.** When  $\varrho_1$  is a negative constant and  $\varrho_2$  is a positive constant, the stationary density is exponential and the process is a Brownian motion with a negative drift and a reflecting barrier at zero. This process is related to the one used to produce Zipf's law. Consider the density of the logarithm of  $x$ . The Zipf's Law implied stationary distribution of  $\ln x$  is exponential translated by  $\ln y$ . When the diffusion coefficient is constant, say  $\alpha^2$ , the drift of  $\ln x$  is  $-\frac{\alpha^2}{2}$ .

The Wong (1964) analysis is very nice because it provides a rather complete characterization of the transition dynamics of the alternative processes investigated. Subsequently, we will describe some of the *spectral* or *eigenfunction* characterizations of dynamic evolution used by Wong (1964) and others. It is the ability to characterize the transition dynamics fully that has made the processes studied by Wong (1964) valuable building blocks for models in economics and finance. Nevertheless, it is often convenient to move outside this family of models.

Within the Pearson class,  $(\ln q)'$  can only have one interior zero. Thus stationary densities must have at most one interior mode. To build diffusion processes with multi-modal densities, Cobb, Koppstein, and Chan (1983) consider models in which  $\varrho_1$  or  $\varrho_2$  can be higher-order polynomials. Since Zipf's Law is arguably about tail properties of a density, nonlinear drift specifications (specifications of  $\varrho_1$ ) are compatible with this law. Chan, Karolyi, Longstaff, and Sanders (1992) consider models of short-term interest rates in which the drift remains linear, but the diffusion coefficient is some power of  $x$  other than linear or quadratic. They treat the volatility elasticity as a free parameter to be estimated and a focal point of their investigation. Ait-Sahalia (1996b) compares the constant volatility elasticity model to other

volatility specifications, also allowing for a nonlinear drift. Conley, Hansen, Luttmer, and Scheinkman (1997) study the constant volatility elasticity model but allowing for drift nonlinearity. Jones (2003) uses constant volatility elasticity models to extend Nelson (1990)'s model of the dynamic evolution of volatility.

### 3.2 Stationary Distributions

To generalize the approach of Wong (1964), we study how to go from the infinitesimal generator to the stationary distribution. Given a generator  $\mathcal{A}$  of a Feller process, we can deduce an integral equation for the stationary distribution. This formula is given by:

$$\lim_{\tau \downarrow 0} \int \frac{\mathcal{T}_\tau \phi - \phi}{\tau} dQ = \int \mathcal{A}\phi dQ = 0 \quad (3.2)$$

for *test functions*  $\phi$  in the domain of the generator. (In fact the collection of functions used to check this condition can be reduced to a smaller collection of functions called the *core* of the generator. See Ethier and Kurtz (1986) for a discussion.)

Integral equation (3.2) gives rise to the differential equation used by Wong (1964) [see (3.1)] and others. Consider test functions  $\phi$  that are twice continuously differentiable and have zero derivatives at the boundaries of the scalar state space. Write the integral equation

$$\int (\mu\phi' + \frac{1}{2}\sigma^2\phi'')q = 0.$$

Using integration by parts once, we see that

$$\int \left[ \mu q - \frac{1}{2}(\sigma^2 q)' \right] \phi' = 0.$$

Given the flexibility of our choice of  $\phi'$ , it follows that

$$\mu q - \frac{1}{2}(\sigma^2 q)' = 0. \quad (3.3)$$

From this equation, we may solve for  $\mu$  as a function of  $(q, \sigma^2)$  or for  $q'/q$  as a function of  $(\mu, \sigma^2)$ . Alternatively, integrating as in Ait-Sahalia (1996a), we may solve for  $\sigma^2$  as a function of  $(\mu, q)$ .

Equation (3.3) has a multivariate counterpart used in our treatment of Markov diffusion processes using quadratic forms. Suppose that there is an  $m$ -dimensional Markov state. An  $m$ -dimensional drift vector  $\mu$  that is consistent with a given smooth stationary density  $q$  and a diffusion matrix  $\nu = [\nu_{ij}]$  has component  $j$  given by:

$$\mu_j q = \frac{1}{2} \sum_{i=1}^m \frac{\partial(\nu_{ij} q)}{\partial y_i}.$$

This choice of  $\mu$  is not unique, however. As discussed in Chen, Hansen, and Scheinkman (2000), it is the unique *symmetric* solution where symmetry is defined in terms of quadratic forms. We will have more to say about this parameterization subsequently.

### 3.3 Fitting the Stationary Distribution

In applied research in macroeconomics and international economics, motivation for parameter choice and model selection is sometimes based on whether they produce reasonable steady-state implications. An analysis like that envisioned by Wong (1964) is germane to this estimation problem. A Wong (1964)-type approach goes beyond the fascination of macroeconomists with deterministic steady states and considers the entire steady state distribution under uncertainty. Whereas Wong (1964) produced diffusion models that imply prespecified densities, it is also straightforward to infer or estimate densities from parameterized diffusion models.

We now consider the problem of fitting an identified model of a generator to the stationary distribution. By calibrating to the implied stationary density and ignoring information about transitions, we may gain some robustness to model misspecification. Of course, we will also lose statistical efficiency and may also fail to identify features of the dynamic evolution. From a statistical standpoint, the entire joint distribution of the data should be informative for making inferences about parameters. A misspecified model may, however, continue to imply correct marginal distributions. Knowledge of this implication is valuable information to a model-builder even if the joint distributions are misspecified.

Initially we allow jump processes, diffusion processes and mixtures, although we will subsequently specialize our discussion to diffusion models. Hansen and Scheinkman (1995) use equation (3.2) to produce estimating equations. Their idea is to parameterize the generator and use the empirical distribution of the data to estimate unknown parameters. That is, consider a family of generators  $\mathcal{A}_b$  parameterized by  $b$ . Given time series data  $\{x_t\}$  and a family of test functions,

$$E[\mathcal{A}_\beta \phi(x_t)] = 0 \tag{3.4}$$

for a finite set of test functions where  $\beta$  is the parameter vector for the Markov model used to generate the data. This can be posed as a generalized-method-of-moments (GMM) estimation problem of the form studied by Hansen (1982).

Two questions arise in applying this approach. Can the parameter  $\beta$  in fact be identified? Can such an estimator be efficient? To answer the first question in the affirmative often requires that we limit the parameterization. We may address Fisher (1921)'s concerns about statistical efficiency by looking over a rich (infinite-dimensional) family of test functions using characterizations provided in Hansen (1985). Statistical efficiency is still not attained because this method ignores information in transition densities. Nevertheless, we may consider a more

limited notion of efficiency because our aim is to fit only the stationary distribution.

In some analyses of Markov process models of stationary densities it is sometimes natural to think of the data as being draws from independent stochastic processes with the same stationary density. This is the case for many applications of Zipf's law. This view is also taken by Cobb, Koppstein, and Chan (1983). For this form of data generation, Fisher (1921)'s argument about statistical efficiency continues to apply. We now consider the case in which data are obtained from a single stochastic process. The analysis is greatly simplified by assuming a continuous-time record of the Markov process between date zero and  $T$ . We use a central limit approximation as the horizon  $T$  becomes large. From Bhattacharya (1982) or Hansen and Scheinkman (1995) we know that

$$\frac{1}{\sqrt{T}} \int_0^T \mathcal{A}_\beta \phi \Rightarrow \text{Normal}(0, -2 \langle \mathcal{A}_\beta \phi | \phi \rangle) \quad (3.5)$$

where  $\Rightarrow$  denotes convergence in distribution, and

$$\langle \mathcal{A}_\beta \phi | \phi \rangle \doteq \int \phi (\mathcal{A}_\beta \phi) dQ,$$

for  $\phi$  in the  $L^2(Q)$  domain of  $\mathcal{A}_\beta$ . This central limit approximation is a refinement of (3.4) and uses an explicit martingale approximation. It avoids having to first demonstrate mixing properties.

Using this continuous-time martingale approximation, we may revisit Fisher (1921)'s critique of Pearson (1894). Consider the special case of a scalar stationary diffusion. Fisher (1921) noted that Pearson (1894)'s estimation method was inefficient, because his moment conditions differed from those implicit in maximum likelihood estimation. Pearson (1894) shunned such methods because they were harder to implement in practice. Of course computational costs have been dramatically reduced since the time of this discussion. What is interesting is that when the data come from (a finite interval of) a single realization of a scalar diffusion, then the analysis of efficiency is altered. As shown by Conley, Hansen, Luttmmer, and Scheinkman (1997), instead of using the score vector for building moment conditions the score vector could be used as test functions in relation (3.4).

To use this approach in practice, we need a simple way to compute the requisite derivatives. The score vector for a scalar parameterization is:

$$\phi = \frac{d \ln q_b}{db}(\beta).$$

Recall that what enters the moment conditions are test function first and second derivatives (with respect to the state). That is, we must know  $\phi'$  and  $\phi''$ , but not  $\phi$ . Thus we need not ever compute  $\ln q$  as a function of  $b$ . Instead we may use the formula:

$$\ln q_b' = \frac{2\mu_b}{\sigma_b^2} - \ln \sigma_b^{2'}$$

to compute derivatives with respect to the unknown parameters. Even though the score depends on the true parameter it suffices to use test functions that are depicted in terms of  $b$  instead of  $\beta$ . Asymptotic efficiency will be preserved.

While statistical efficiency presumes a correct specification, any misspecification that leaves intact the parameterized model of the stationary density will remain consistent under ergodicity assumptions. Checking to see whether a model fits the stationary density for some set of parameters is an interesting question in its own right. One possible approach is to add in test functions aimed at specific features of the stationary distribution to obtain an additional set of over-identifying restrictions. Following Bierens (1990), such a method could be refined by using an ever enlarging collection of test functions as the sample size is increased, but the practical impact of this observation seems limited.

An alternative comprehensive comparison of a parametric density estimator can be made to a nonparametric estimator to obtain a specification test. Consider the following comparison criterion:

$$\int (q_b - q)^2 \omega \tag{3.6}$$

where  $q$  is the true density of the data and  $\omega$  a weighting function.<sup>2</sup> Instead of constructing a small number of test functions that feature specific aspects of the distribution, a researcher specifies the weighting function  $\omega$  that dictates which ranges of data receive more emphasis in the statistical test. By design, objective (3.6) is zero only when  $q_b$  and  $q$  coincide for some admissible value of  $b$ . As before, a parameterization of  $q_b$  can be inferred from a parameterization of the generator  $\mathcal{A}$ . The implied model of the stationary density is parameterized correctly when the objective is zero for some choice of  $b$ . Aït-Sahalia (1996b) uses this to devise a statistical test for misspecification of the stationary density.

Following Aït-Sahalia (1996b), the density  $q$  can be estimated consistently from discrete-time data using nonparametric methods. The parameter  $b$  can be estimated using the method previously described or by minimizing the sample-counterpart to (3.6). Aït-Sahalia (1996b) derives the limiting distribution of the resulting test statistic and applies this method to test models of the short-term interest rate process.<sup>3</sup> One challenge facing such nonparametric tests is producing accurate small sample distributions. The convergence to the asymptotic distribution obtained by assuming stationarity of the process can be slow when the data are highly persistent, as is the case with US interest rates. (See Pritsker (1998) and Conley, Hansen, and Liu (1999).)

---

<sup>2</sup>Distance measures other than this  $L^2$  weighted norm can be used, such as an entropy measure.

<sup>3</sup>See Section 5.4 and Aït-Sahalia (1996b) for an analogous test based on transition densities.

### 3.4 Nonparametric Methods for Inferring Drift or Diffusion Coefficients

Recall that for a scalar diffusion, the drift coefficient can be inferred from a stationary density, the diffusion coefficient and their derivatives. Alternatively the diffusion coefficient can be deduced from the density and the drift coefficient. These functional relationships give rise to nonparametric estimation methods for the drift coefficient or the diffusion coefficient. In this subsection we describe how to use local parametrizations of the drift or the diffusion coefficient to obtain nonparametric estimates. The parameterizations become localized by their use of test functions or kernels familiar from the literature on nonparametric estimation.

In the framework of test functions, this estimation method can be viewed as follows. In the case of a scalar diffusion,

$$\int (\mu\phi' + \frac{1}{2}\sigma^2\phi'')q = 0. \quad (3.7)$$

Construct a test function  $\phi$  such that  $\phi'$  is zero everywhere except in the vicinity of some pre-specified point  $y$ . The function  $\phi'$  can be thought of as a kernel and its localization can be governed by the choice of a bandwidth. As in Banon (1978), suppose that the diffusion coefficient is known. We can construct a locally constant estimator of  $\mu$  that is very close to Banon (1978)'s estimator by solving the sample counterpart to (3.7) under the possibly false assumption that  $\mu$  is constant. The local specification of  $\phi'$  limits the range over which constancy of  $\mu$  is a good approximation, and the method produces a local estimator of  $\mu$  at the point  $y$ . This method is easily extended to other local parametrizations of the drift. Conley, Hansen, Luttmer, and Scheinkman (1997) introduce a local linear estimator by using two local test functions to identify the level and the slope of the linear approximation. Using logic closely related to that of Florens-Zmirou (1984), these local estimators sometimes can presumably be justified when the integrability of  $q$  is replaced by a weaker recurrence assumption.

Suppose that a linear function is in the domain of the generator. Then

$$\int \mu q = 0. \quad (3.8)$$

We may now localize the parameterization of the diffusion coefficient by localizing the choice of  $\phi''$ . The specific construction of  $\phi'$  from  $\phi''$  is not essential because moment condition (3.8) is satisfied. For instance, when  $\phi''$  is scaled appropriately to be a density function, we may choose  $\phi'$  to be its corresponding distribution function. The locally constant diffusion estimator obtained by solving the sample counterpart to (3.7) for a local specification of  $\phi''$  is very similar to the one analyzed by Aït-Sahalia (1996a).

Applying integration by parts,

$$\int_l^r \mu(x)\phi'(x)q(x)dx = \int_l^r \left[ \int_x^r \mu q \right] \phi''(x)dx$$

provided that the localization function  $\phi''$  has support in the interior of the state space  $(l, r)$ . By localizing the parameterization of the diffusion coefficient at  $x$ , the limiting version of (3.7) is:

$$\int_x^r \mu q + \frac{\sigma^2(x)q(x)}{2} = 0.$$

Using (3.8), this gives us

$$\sigma^2(x) = \frac{2}{q(x)} \int_l^x \mu(u) q(u) du, \quad (3.9)$$

which is equivalent to solving differential equation (3.3) for the diffusion coefficient. Ait-Sahalia (1996a) constructed a diffusion estimate by substituting a nonparametric density estimator for  $q$ .

The local approaches for constructing estimators of  $\mu$  or  $\sigma^2$  require knowledge of estimates of the other piece. Suppose we parameterize  $\mu$  as in Ait-Sahalia (1996a) to be affine in the state variable,  $\mu(x) = -\kappa(x - \alpha)$ , and a linear function is a valid test function, then

$$\mathcal{A}(x - \alpha) = -\kappa(x - \alpha).$$

This says that  $x - \alpha$  is an eigenfunction of  $\mathcal{A}$ , with eigenvalue  $-\kappa$ . We shall have more to say about eigenfunctions and eigenvalues in section 4. The conditional expectation operator for any interval  $t$  must have the same eigenfunction and an eigenvalue given via the exponential formula:

$$\mathcal{T}_t(x - \alpha) = E[X_t | X_0 = x] = e^{-\kappa t} (x - \alpha).$$

This conditional moment condition applies for any  $t > 0$ . As a consequence,  $(\alpha, \kappa)$  can be recovered by estimating a first order scalar autoregression via least squares for data sampled at any interval  $t = \Delta$ . Following Ait-Sahalia (1996a), the implied drift estimator may be plugged into formula (3.9) to produce a semiparametric estimator of the drift and diffusion function  $\sigma^2$ .

As an alternative, Conley, Hansen, Luttmer, and Scheinkman (1997) produce a semiparametric estimator by adopting a constant volatility elasticity specification of the diffusion coefficient, while letting the drift be nonparametric. The volatility elasticity is identified by using an additional set of moment conditions derived in section 5.4 applicable for some subordinated diffusion models. Subordinated Markov processes will be developed in 6.4.

We will have more to say about observable implications including nonparametric identification in section 5.

## 4 Transition Dynamics and Expansions

In this section we describe some related expansions, which are valuable in characterizing the transitional dynamics. Some of these expansions are model specific and others are designed to

be used across a parameterized family of models. Some are global and some are local. One set of methods is based on a functional analog to principal components. The resulting approximations are global, but they are not always applicable, and they use basis functions (principal components) that differ across models. Nevertheless, they are useful in characterizing the transitional dynamics of specific models.

We then consider expansions that can be applied to all diffusions (subject only to regularity conditions) and lead to *closed form* expressions. As a consequence, they are more readily used in estimation. First, we consider the univariate Hermite expansions of Aït-Sahalia (2002b); by using a judicious change of variable, these expansions use polynomial basis functions that are common across all models. Second, in the multivariate case, we consider the local expansions of Aït-Sahalia (2001) which rely on expansions in both the time and state dimensions and again deliver closed form expressions.

## 4.1 Principal Component Methods

While the stationary density gives one notion of the *long run*, transition distributions are essential to understand the full dynamic implications of nonlinear Markov models. Moreover, stationary distributions are typically not sufficient to identify all of the parameters of interest. Wong (1964) characterized transition dynamics using what is called a spectral decomposition. This decomposition is analogous to the spectral or principal component decomposition of a symmetric matrix. Since we are interested in nonlinear dynamics, we use a functional counterpart to principal component analysis.

We use quadratic forms and eigenfunctions to produce decompositions of both the stationary distribution and the dynamic evolution of the process. These decompositions show what features of the time series dominate in the long run and, more generally, give decompositions of the transient dynamics.

### 4.1.1 Forms

Previously, we demonstrated that a scalar diffusion can be constructed using a density  $q$  and a diffusion coefficient  $\sigma^2$ . By using quadratic forms described in Section 2, we may extend this construction to a broader class of Markov process models. The form construction allows us to define a nonlinear version of principal components.

Let  $Q$  be a Radon measure on the state space  $X$ .<sup>4</sup> For the time being this measure need not be finite, although we will subsequently add this restriction. When  $Q$  is finite, it is proportional to the stationary distribution of the corresponding Markov process. We consider

---

<sup>4</sup>A Radon measure is a Borel measure that assigns finite measure to every compact subset of the state space and strictly positive measure to nonempty open sets.

two positive semi-definite quadratic forms on the space of functions  $L^2(Q)$ . One is given by the usual inner product:

$$f_1(\phi, \psi) \doteq \langle \phi, \psi \rangle = \int \phi \psi dQ.$$

This form is *symmetric* [ $f_1(\phi, \psi) = f_1(\psi, \phi)$ ] and *positive semidefinite* ( $f_1(\phi, \phi) \geq 0$ ).

The second form is constructed from two objects: (a) a state dependent positive semidefinite matrix  $\nu$  and (b) a symmetric, positive Radon measure  $R$  on the product space  $X \times X$  excluding the diagonal  $D \doteq \{(x, x) : x \in X\}$  with

$$\int_{X \times X - D} \frac{|x - y|^2}{1 + |x - y|^2} R(dx, dy) < \infty.$$

It is given by:

$$f_2(\phi, \psi) \doteq \frac{1}{2} \int (\nabla \phi)^* \nu (\nabla \psi) dQ + \frac{1}{2} \int [\phi(y) - \phi(x)][\psi(y) - \psi(x)] R(dx, dy)$$

where  $*$  is used to denote transposition. The form  $f_2$  is well-defined at least on the space  $C_K^2$  of twice continuously differentiable functions with compact support. Under additional regularity conditions, the form  $f_2$  is closable, that is, it has a closed extension in  $L^2(Q)$ .<sup>5</sup> However, even this extension has a limited domain. Like  $f_1$ , the form  $f_2$  is also symmetric and positive semidefinite. Notice that  $f_2$  is the sum of two forms. As we will see the first is associated with a diffusion process and the second with a jump process.<sup>6</sup>

#### 4.1.2 Implied Generator

We may now follow the approach of Beurling and Deny (1958) and Fukushima (1971) by constructing a Markov process associated with the form  $f_1$  and the closed extension of  $f_2$ . In what follows we will sketch only part of this construction. We describe how to go from the forms  $f_1$  and  $f_2$  to an implied generator. The generator  $\mathcal{A}$  is the symmetric solution to:

$$f_2(\phi, \psi) = -f_1[(\mathcal{A}\phi), \psi] = - \int (\mathcal{A}\phi) \psi dQ. \quad (4.1)$$

Since  $f_2$  is a positive semidefinite form,  $\mathcal{A}$  is a negative semidefinite operator.

We explore this construction for each of the two components of  $f_2$  separately. Suppose initially that  $R$  is identically zero. Then

$$f_2(\phi, \psi) \doteq \frac{1}{2} \int (\nabla \phi)^* \nu (\nabla \psi) q \quad (4.2)$$

---

<sup>5</sup>For instance if  $Q$  has density  $q$ , and  $q$  and  $\nu$  are continuously differentiable, then the form  $f_2$  is closable.

<sup>6</sup>In fact there exist generalizations of this representation in which  $\nu$  is replaced by a matrix-valued measure and an additional term  $\int \phi(x) \psi(x) dk(x)$  is introduced where  $k$  is a *killing* measure. See Beurling and Deny (1958) and Fukushima, Oshima, and Takeda (1994).

where  $q$  is the density of  $Q$ . Applying an integration-by-parts argument to (4.2) shows that  $\mathcal{A}_d$  can be depicted as a second order differential operator on the space  $C_K^2$  of twice continuously differentiable functions with compact support:

$$\mathcal{A}_d\phi = \frac{1}{2} \sum_{i,j} \nu_{ij} \frac{\partial^2 \phi}{\partial y_i \partial y_j} + \frac{1}{2q} \sum_{i,j} \frac{\partial(q\nu_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j}$$

provided that both  $q$  and  $\nu$  are continuously differentiable.<sup>7</sup> In this formula we set  $\nu_{ij}$  to be the  $(i, j)$  element of the matrix  $\nu$ . Moreover, the implicit drift is

$$\mu_j = \frac{1}{2q} \sum_{i=1}^m \frac{\partial(\nu_{ij}q)}{\partial y_i}. \quad (4.3)$$

This gives us a multivariate extension to the idea of parameterizing a Markov diffusion process in terms of a density  $q$  and the diffusion matrix  $\nu$ , with the drift being implicit.

Next suppose that  $\nu$  is identically zero, and again assume that  $Q$  has a density  $q$ . Write:

$$\begin{aligned} f_2(\phi, \psi) &= \frac{1}{2} \int [\phi(y) - \phi(x)] [\psi(y) - \psi(x)] R(dx, dy) \\ &= -\frac{1}{2} \int [\phi(y) - \phi(x)] \psi(x) \frac{R(dx, dy)}{q(x)} q(x) dx + \frac{1}{2} \int [\phi(y) - \phi(x)] \psi(y) R(dx, dy) \\ &= -\int [\phi(y) - \phi(x)] \psi(x) \frac{R(dx, dy)}{q(x)} q(x) dx \end{aligned}$$

where we used the symmetry of  $R$ . The joint measure  $R(dx, dy)/q(x)$  implies a conditional measure  $R(dy|x)$  from which we define:

$$\mathcal{A}_p\phi \doteq \int [\phi(y) - \phi(x)] R(dy|x).$$

When the measure  $R(dy|x)$  is finite for all  $x$ , the Poisson intensity parameter is:

$$\lambda(x) = \int R(dy|x),$$

which governs the frequency of the jumps. The probability distribution conditioned on the state  $x$  and a jump occurring is:  $R(dy|x)/\int R(dy|x)$ . Thus  $\mathcal{A}_p$  can be depicted as

$$\mathcal{A}_p = \lambda(\mathcal{J}\phi - \phi)$$

by defining

$$\mathcal{J}\phi = \frac{\int \phi(y) R(dy|x)}{\int R(dy|x)}.$$

---

<sup>7</sup>The continuous differentiability restriction can be weakened by introducing weak derivatives.

We have just shown how to go from the forms to the generator of Markov processes. There is one technical complication that we sidestepped. In general there may be several closed extensions of  $f_2$  depending on boundary restrictions. The smallest of these closed extensions always generates a semigroup of contractions. This semigroup will correspond to a semigroup of conditional expectations provided that the associated operator  $\mathcal{A}$  *conserves probabilities*. When this happens all closed extensions that lead to a Markov process produce exactly the same process constructed with the aid of the minimal extension (*e.g.* Chen, Hansen, and Scheinkman (2000) Proposition 4.6 and references therein.)<sup>8</sup>

Fukushima, Oshima, and Takeda (1994) provide sufficient conditions for conservation of probabilities. An implication of the sufficient conditions of Fukushima, Oshima, and Takeda (1994) is that if  $|\nu_{ij}(x)| \leq c|x|^{2+2\delta}$  and  $q$  has a  $2\delta$  moment, probabilities are conserved. (See also Chen, Hansen, and Scheinkman (2000).) Another set of sufficient conditions can be obtained by observing that a recurrent semigroup conserves probabilities (Fukushima, Oshima, and Takeda (1994) Lemma 1.6.5). Hasminskii (1960) and Stroock and Varadhan (1979) suggest using Liapounov functions to demonstrate recurrence.

### 4.1.3 Symmetrization

There are typically nonsymmetric solutions to (4.1). Given a generator  $\mathcal{A}$ , let  $\mathcal{A}^*$  denote its adjoint. Define a symmetrized generator as:

$$\mathcal{A}^s = \frac{\mathcal{A} + \mathcal{A}^*}{2}.$$

Then  $\mathcal{A}^s$  can be recovered from the forms  $f_1$  and  $f_2$  using the algorithm suggested previously. The symmetrized version of the generator is identified by the forms, while the generator itself is not.

We consider a third form using one-half the difference between  $\mathcal{A}$  and  $\mathcal{A}^*$ . Define:

$$f_3(\phi, \psi) = \int \left( \frac{\mathcal{A} - \mathcal{A}^*}{2} \phi \right) \psi dQ.$$

This form is clearly anti-symmetric. That is

$$f_3(\phi, \psi) = -f_3(\psi, \phi)$$

for all  $\phi$  and  $\psi$  in the common domain of  $\mathcal{A}$  and its adjoint. We may recover a version of  $\frac{\mathcal{A} + \mathcal{A}^*}{2}$  from  $(f_1, f_2)$  and  $\frac{\mathcal{A} - \mathcal{A}^*}{2}$  from  $(f_1, f_3)$ . Taken together we may construct  $\mathcal{A}$ . Thus to

---

<sup>8</sup>When the smallest closed extension fails to conserve probabilities, we may still build an associated Markov process, provided that we allow the process to be *killed* in finite time when it hits a boundary. Other boundary protocols are also possible and lead to the study of alternative closed extensions.

study nonsymmetric Markov processes via forms, we are led to introduce a third form, which is antisymmetric. See Ma and Rockner (1991) for an exposition of nonsymmetric forms and their resulting semigroups.

In what follows we specialize our discussion to the case of multivariate diffusions. When the dimension of the state space is greater than one, there are typically also nonsymmetric solutions to (4.1). Forms do not determine uniquely operators without additional restrictions such as symmetry. These nonsymmetric solutions are also generators of diffusion processes. While the diffusion matrix is the same for the operator and its adjoint, the drift vectors differ. Let  $\mu$  denote the drift for a possibly nonsymmetric solution,  $\mu^s$  denote the drift for the symmetric solution given by (4.3), and let  $\mu^*$  denote the drift for the adjoint of the nonsymmetric solution. Then

$$\mu^s = \frac{\mu^* + \mu}{2}.$$

The form pair  $(f_1, f_2)$  identifies  $\mu^s$  but not necessarily  $\mu$ .

The form  $f_3$  can be depicted as:

$$f_3(\phi, \psi) = \frac{1}{2} \int [(\mu - \mu^*) \cdot (\nabla \phi)] \psi q$$

at least for functions that are twice continuously differentiable and have compact support. For such functions we may use integration by parts to show that in fact:

$$f_3(\phi, \psi) = -f_3(\psi, \phi).$$

Moreover, when  $q$  is a density, we may extend  $f_3$  to include constant functions via

$$f_3(\phi, 1) = \frac{1}{2} \int (\mu - \mu^*) \cdot (\nabla \phi) q = 0.$$

#### 4.1.4 Principal Components

Given two quadratic forms, we define the functional versions of principal components.

**Definition 6.** *Nonlinear principal components are functions  $\psi_j$ ,  $j = 1, 2, \dots$  that solve:*

$$\max_{\phi} f_1(\phi, \phi)$$

subject to

$$\begin{aligned} f_2(\phi, \phi) &= 1 \\ f_1(\phi, \psi_s) &= 0, s = 0, \dots, j-1 \end{aligned}$$

where  $\psi_0$  is initialized to be the constant function one.<sup>9</sup>

---

<sup>9</sup>We may use weak gradients in this construction.

This definition follows Chen, Hansen, and Scheinkman (2000) and is a direct extension of that used by Salinelli (1998) developed for *iid* data. In the case of a diffusion specification, form  $f_2$  is given by (4.2) and induces a quadratic smoothness penalty. Principal components maximize variation subject to a smoothness constraint and orthogonality. These components are a nonlinear counterpart to the more familiar principal component analysis of covariance matrices advocated by Pearson (1901). In the functional version, the state dependent, positive definite matrix  $\nu$  is used to measure smoothness. Salinelli (1998) advocated this version of principal component analysis for  $\nu = I$  to summarize the properties of *i.i.d.* data. As argued by Chen, Hansen, and Scheinkman (2000) they are equally valuable in the analysis of time series data. The principal components, when they exist, will be orthogonal under either form. That is:

$$f_1(\psi_j, \psi_k) = f_2(\psi_j, \psi_k) = 0$$

provided that  $j \neq k$ .

These principal components coincide with the principal components from the canonical analysis used by Darolles, Florens, and Gourieroux (2000) under symmetry, but otherwise they differ. In addition to maximizing variation under smoothness restrictions (subject to orthogonality), they maximize autocorrelation and they maximize the long run variance as measured by the spectral density at frequency zero. See Chen, Hansen, and Scheinkman (2000) for an elaboration.

This form approach and the resulting principal component construction is equally applicable to *i.i.d.* data and to time series data. In the *i.i.d.* case, the matrix  $\nu$  is used to measure function smoothness. Of course in the *i.i.d.* case there is no connection between the properties of  $\nu$  and the data generator. The Markov diffusion model provides this link.

The smoothness penalty is special to diffusion processes. For jump processes, the form  $f_2$  is built using the measure  $R$ , which still can be used to define principal components. These principal components will continue to maximize autocorrelation and long run variance subject to orthogonality constraints.

#### 4.1.5 Existence

It turns out that principal components do not always exist. Existence is straightforward when the state space is compact, the density  $q$  is bounded above and bounded away from zero and the diffusion matrix is uniformly nonsingular on the state space. These restrictions are too severe for many applications. Chen, Hansen, and Scheinkman (2000) treat cases where these conditions fail.

Suppose the state space is not compact. When the density  $q$  has thin tails, the notion of approximation is weaker. Approximation errors are permitted to be larger in the tails. This

turns out to be one mechanism for the existence of principal components. Alternatively,  $\nu$  might increase in the tails of the distribution of  $q$  limiting the admissible functions. This can also be exploited to establish the existence of principal components.

Chen, Hansen, and Scheinkman (2000) exhibit sufficient conditions for existence that require a trade-off between growth in  $\nu$  and tail thinness of the density  $q$ . Consider the (lower) radial bounds,

$$\begin{aligned}\nu(x) &\geq \rho(|x|)^2 I \\ q(x) &\geq \exp[-2\tau(|x|)].\end{aligned}$$

where  $\rho(r) = c(1+r^2)^{\frac{\beta}{2}}$ . Principal components exist when  $0 \leq \beta \leq 1$  and  $\rho\tau' \rightarrow \infty$ . Similarly, they also exist when  $\tau(r) = \frac{\gamma}{2} \ln(1+r^2) + c^*$ , and  $1 < \beta < \gamma - \frac{m}{2} + 1$ . The first set of sufficient conditions deals with a density with exponentially thin tails, the second is useful when  $q$  has an algebraic tail.

We now consider some special results for the case  $m = 1$ . We let the state space be  $(l, r)$ , where either boundary can be infinite. Again  $q$  denotes the stationary density and  $\sigma > 0$  the volatility coefficient (that is,  $\sigma^2 = \nu$ .) Suppose that

$$\int_l^r \left| \int_{x_o}^x \frac{1}{q(y)\sigma^2(y)} dy \right| q(x) dx < \infty \quad (4.4)$$

where  $x_o$  is an interior point in the state space. Then principal components are known to exist. For a proof see, *e.g.* Hansen, Scheinkman, and Touzi (1998), page 13, where this proposition is stated using the *scale function*

$$s(x) \doteq \int_{x_o}^x \frac{1}{q(y)\sigma^2(y)} dy,$$

and it is observed that (4.4) admits entrance boundaries, in addition to attracting boundaries.

When assumption (4.4) is not satisfied, at least one of the boundaries is natural. Recall that the boundary  $l$  ( $r$ ) is natural if  $s(l) = -\infty$  ( $s(r) = +\infty$  resp.) and,

$$\int_l^{x_o} s(x)q(x)dx = -\infty \quad \left( \int_{x_o}^r s(x)q(x)dx = +\infty \text{ resp.} \right)$$

In Hansen, Scheinkman, and Touzi (1998) it is shown that principal components exist whenever

$$\begin{aligned}\limsup_{x \rightarrow r} \frac{\mu}{\sigma} - \frac{\sigma'}{2} &= \limsup_{x \rightarrow r} \frac{\sigma q'}{2q} + \frac{\sigma'}{2} = -\infty \\ \liminf_{x \rightarrow l} \frac{\mu}{\sigma} - \frac{\sigma'}{2} &= \liminf_{x \rightarrow l} \frac{\sigma q'}{2q} + \frac{\sigma'}{2} = +\infty.\end{aligned} \quad (4.5)$$

We can think of the left-hand side of (4.5) as a local measure of *pull* towards the center of the distribution. If one boundary, say  $l$ , is reflexive and  $r$  is natural, then a principal component decomposition exists provided that the  $\liminf$  in (4.5) is  $+\infty$ .

### 4.1.6 Spectral Decomposition

Principal components, when they exist, can be used to construct the semigroup of conditional expectation operators as in Wong (1964). A principal component decomposition is analogous to the spectral decomposition of a symmetric matrix. Each principal component is an eigenfunction of all of the conditional expectation operators and hence behaves like a first-order scalar autoregression (with conditionally heteroskedastic innovations). See Darolles, Florens, and Gouriéroux (2001) for an elaboration. Thus principal components constructed from the stationary distribution must satisfy an extensive family of conditional moment restrictions.

Both the generator and the semigroup of conditional expectations operators have *spectral* (principal component) decompositions. The generator has spectral decomposition:

$$\mathcal{A}\phi = \sum_{j=0}^{\infty} -\delta_j f_1(\psi_j, \phi) \psi_j,$$

where each  $\delta_j > 0$  and,  $\psi_j$  is a principal component (normalized to have a unit second moment) and an eigenvector associated with the eigenvalue  $-\delta_j$ , that is,

$$\mathcal{A}\psi_j = -\delta_j \psi_j.$$

The corresponding decomposition for the semigroup uses an exponential formula:

$$\mathcal{T}_{\Delta}\phi = \sum_{j=0}^{\infty} \exp(-\Delta\delta_j) f_1(\psi_j, \phi) \psi_j. \quad (4.6)$$

This spectral decomposition shows that the principal components of the semigroup are ordered in importance by which dominate in the long run.

Associated with (4.6) for a diffusion is an expansion of the transition density  $p$  over an interval of time  $\Delta$ . Write:

$$p(y|x, \Delta) = \sum_{j=0}^{\infty} \exp(-\Delta\delta_j) \psi_j(y) \psi_j(x) q(y)$$

where  $q$  is the stationary density. Notice that we have constructed  $q(y|x, \Delta)$  so that

$$\mathcal{T}_{\Delta}\phi(x) = \int \phi(y) p(y|x, \Delta) dy.$$

The basis functions used in this density expansion depend on the underlying model. Recall that an *Ornstein-Uhlenbeck* process has a stationary distribution that is normal (see Example 3). Decomposition (4.6) is a Hermite expansion when the stationary distribution has mean zero and variance one. The eigenfunctions are the orthonormal polynomials with respect to a standard normal distribution.

Spectral decomposition does not require the existence of principal components. We have seen how to construct Markov processes with self adjoint generators using forms. A more general version of the spectral decomposition of generators is applicable to the resulting semi-group and generator that generalizes formula (4.6), see Rudin (1973), Hansen and Scheinkman (1995) and Schaumburg (2003). This decomposition is applicable generally for scalar diffusions even when a stationary density fails to exist, for a wide class of Markov processes defined via symmetric forms. The measure  $q$  used in constructing the forms and defining a sense of approximation need not be integrable.

## 4.2 Exponential Expansion

While the spectral depiction (4.6) of the exponential formula is applicable to all functions that are square integrable with respect to  $q$ , it is often difficult to use in practice. Sometimes a power series expansion can be used for a subset of the domain of the generator. By a power series we mean:

$$\mathcal{T}_\Delta \phi \approx \sum_{j=0}^N \frac{\Delta^j \mathcal{A}^j \phi}{j!}. \quad (4.7)$$

Schaumburg (2003) provides an justification for this formula for the following collection of functions. Consider a function  $\phi$  such that  $\phi = \mathcal{T}_t \psi$  for some  $\psi \in L^2(Q)$ . Then under an additional arguably weak regularity condition (see Assumption 2A in Schaumburg (2003)), the power series converges for  $\Delta < t$ .

To illustrate this result, suppose there exists a spectral decomposition of the form given in (4.6) for  $\psi$ . Then

$$\sum_{j=0}^N \frac{\Delta^j \mathcal{A}^j \phi}{j!} - \exp(\mathcal{A}) \phi = \sum_{k=0}^{\infty} \left[ \sum_{j=0}^N \frac{(-\Delta \delta_j)^j}{j!} - \exp(-\Delta \delta_k) \right] \exp(-\delta_k t) f_1(\psi_k, \psi) \psi_k.$$

Notice that

$$\exp(-\Delta \delta_k) \left| \sum_{j=0}^N \frac{(-\Delta \delta_k)^j}{j!} - \exp(-\Delta \delta_k) \right| \leq \exp(-\Delta \delta_k) \left[ \sum_{j=0}^N \frac{(\Delta \delta_k)^j}{j!} + \exp(-\Delta \delta_k) \right] \leq 2.$$

Thus we can bound the approximation error:

$$\left| \sum_{j=0}^N \frac{\Delta^j \mathcal{A}^j \phi}{j!} - \exp(\mathcal{A}) \phi \right| \leq 2 \sum_{k=0}^{\infty} -\exp[-(t - \Delta) \delta_k] f_1(\psi_k, \psi) \psi_k.$$

The exponential discounting of the *coefficients*  $f_1(\psi_j, \psi)$  in this expansion

This approximation can be used to compute approximate conditional moment restrictions, or as we will see to support approximation of a transition density and a likelihood function.

### 4.3 Hermite Expansion

We have already noted that a spectral decomposition of the semigroup for an *Ornstein-Uhlenbeck* process with a standard normal stationary distribution is a Hermite expansion. In problems of estimation it is often convenient to use a common expansion for alternative models, and Hermite expansion is a leading example. Following Aït-Sahalia (1999) and Aït-Sahalia (2002b), we describe Hermite approximation for scalar diffusions, including a method for computing the coefficients in practice.

It is clearly special and limiting to have a stationary distribution that is standard normal. To make the standard normal distribution useful for approximation, we transform the state and rescale the change in the state over an interval of time  $\Delta$ . To understand the construction, the following analogy may be helpful. Consider a standardized sum of random variables to which the Central Limit Theorem (CLT) apply. Often, one is willing to approximate the actual sample size  $N$  by infinity and use the standard normal limiting distribution for the properly standardized transformation of the data. If not, higher-order terms of the limiting distribution (for example the classical Edgeworth expansion based on Hermite polynomials) can be calculated to improve the performance of the approximation.

Consider now approximating the transition density of a diffusion, and think of the sampling interval  $\Delta$  as playing the role of the sample size in the CLT. A property of a diffusion is that over small increments of time, first differences divided by  $\sqrt{\Delta}$  are approximately normal. The normal approximation becomes better as the interval  $\Delta$  becomes small, but the variance may be state dependent. Thus prior to shrinking  $\Delta$  to zero, we transform the state to make the limiting approximation a standard normal. The transformation is:

$$Y_t \doteq \gamma(X_t) = \int^{\cdot X_t} \frac{du}{\sigma(u)}$$

where the lower endpoint of integration is some interior point in the state space. The constructed process  $\{Y_t\}$  has a unit diffusion coefficient and a drift:<sup>10</sup>

$$\mu_y(y) = \frac{\mu[\gamma^{-1}(y)]}{\sigma[\gamma^{-1}(y)]} - \frac{1}{2} \frac{d\sigma}{dx} [\gamma^{-1}(y)]. \quad (4.8)$$

The stationary density  $q_y$  for the transformed process is typically not normal, but it satisfies:

$$q_y(y) \propto \exp \left[ 2 \int^y \mu_y(u) du \right].$$

While it is possible for the transformed state to have finite upper or lower bounds, we focus on the case in which the implied state space is  $\mathbb{R}$ . The stationary density will have exponentially

---

<sup>10</sup>The construction of  $\mu_y$  on the right-hand side of (4.8) is familiar from our discussion of the pull measure (4.5) used to justify the existence of principal components for a scalar diffusion process.

thin tails provided that the drift  $\mu_y$  is negative (positive) for large positive (negative)  $y$  and bounded away from zero in the tails. Thus polynomials have finite second moments after this transformation has been applied, provided that there is some pull towards the origin in the implied drift.<sup>11</sup>

If the drift of the process  $\{Y_t\}$  were zero, then it would be a standard Brownian motion. The first-difference in  $\{Y_t\}$  would have a standard normal density only after dividing by the square root of the sampling interval  $\Delta$ . More generally, let  $p_y$  denote the transition function of the process  $\{Y_t\}$ . Without this scaling, the first-difference of  $\{Y_t\}$  will converge to a degenerate measure with a unit probability mass (a Dirac mass) at zero. To obtain the Hermite refinement of a standard normal approximation, we form

$$Z_\Delta \doteq \Delta^{-1/2} (Y_\Delta - Y_0)$$

and condition on  $Y_0 = y_0 = \gamma(x_0)$ . Let  $p_z$  denote the conditional distribution of  $Z_\Delta$  where  $\Delta$  denotes the time interval used in the approximation.

Since  $Z_\Delta$  is a known transformation of  $X$ , we can recover the transition density of  $X$  from the density of  $Z_\Delta$  using the familiar jacobian formula:

$$p(x|x_0, \Delta) = \frac{p_y[\gamma(x) | \gamma(x_0), \Delta]}{\sigma(x)} = \frac{p_z(\Delta^{-1/2}[\gamma(x) - \gamma(x_0)] | \gamma(x_0), \Delta)}{\sigma(x)\Delta^{1/2}}. \quad (4.9)$$

So this leaves us with the need to approximate the density function  $p_z$ .

Let  $h_j$  denote the Hermite polynomials, the orthogonal polynomials with respect to the standard normal density. They can be computed as:

$$h_j(z) \doteq \exp\left(\frac{z^2}{2}\right) \frac{d^j}{dz^j} \left[ \exp\left(-\frac{z^2}{2}\right) \right], \quad j \geq 0.$$

The Hermite expansion is

$$p_z(z|y_0, \Delta) = \frac{\exp(-z^2/2)}{\sqrt{2\pi}} \sum_{j=0}^{\infty} \eta_j(\Delta, y_0) h_j(z) \quad (4.10)$$

with coefficients given by:

$$\begin{aligned} \eta_j(\Delta, y_0) &= \left(\frac{1}{j!}\right) \int_{-\infty}^{+\infty} h_j(z) p_z(z|y_0, \Delta) dz \\ &= \left(\frac{1}{j!}\right) E(h_j[\Delta^{-1/2}(Y_\Delta - Y_0)] | Y_0 = y_0) \end{aligned}$$

---

<sup>11</sup>As we will see in section 5.7, these conditions on the pull measure imply weak dependence of the diffusion process.

A Hermite approximation to  $p_z$  uses a finite number of terms in expansion (4.10). A corresponding approximation for  $p_x$  follows from (4.9).

Since the coefficients  $\eta_j$  are specific conditional moments of the process  $\{Y_t\}$ , they can be computed using numerical methods such as Monte Carlo integration. A particularly attractive alternative is to exploit the exponential decay in the tails of the stationary density for the transformed process  $\{Y_t\}$  and compute the coefficients using the power series approximation (4.7). This can be justified using the argument in Appendix ?? provided that the requisite regularity conditions are satisfied. Aït-Sahalia (1999) gives the formulae corresponding to popular models in finance. Jensen and Poulsen (2002) show that this Hermite approximation works very well in practice and that it dominates other methods for the benchmark examples they consider. Aït-Sahalia (2002b) uses this approach to approximate numerically a parametric likelihood function for scalar diffusion estimation.

## 4.4 Local Expansion

In the univariate Hermite expansion derived in section 4.3, we first deduced the Hermite expansion in terms of polynomials in  $y - y_0$  for a given  $\Delta$ . Once the Hermite coefficients  $\eta_j(\Delta, y_0)$  are replaced by their Taylor series approximation in  $\Delta$ , the corresponding approximation expansion becomes local in  $\Delta$ . In addition to using a finite number of Hermite polynomials, we limited our use to a finite number of  $\Delta$  terms in the Taylor expansion used to approximate the coefficients.<sup>12</sup> Following Aït-Sahalia (2001) we will use a similar strategy except that we will deduce directly a small  $\Delta$  expansion first. In contrast to the Hermite expansion, this expansion applies directly to the *logarithm* of the transition density and permits the diffusion to be multivariate. After deducing the  $\Delta$  expansion, we will explore an approximation based on the discrepancy between the state to which you move to and the current state. Formally we will deduce this as a small discrepancy approximation. Taken together, this joint expansion provides an operational way to approximate (logarithms) of transition densities for multivariate diffusions.

### 4.4.1 Expansion in $\Delta$

Aït-Sahalia (2001) shows that an expansion at order  $K$  in  $\Delta$  for  $\ell(x|x_0, \Delta) \doteq \ln p(x|x_0, \Delta)$  can be obtained in the form:

$$\ell_K(x|x_0, \Delta) \approx C_{-1}(x|x_0)\Delta^{-1} + \tilde{C}(x|x_0) \ln \Delta + \sum_{k=0}^K C_k(x|x_0) \frac{\Delta^k}{k!}$$

---

<sup>12</sup>Different ways of gathering the terms are available as in the Central Limit Theorem, where both the Edgeworth and Gram-Charlier expansions are based on a Hermite expansion.

The derivative with respect to  $\Delta$  of the approximating function is therefore:

$$\frac{\partial \ell_K}{\partial \Delta}(x|x_0, \Delta) \approx -C_{-1}(x|x_0)\Delta^{-2} + \tilde{C}(x|x_0)\Delta^{-1} + \sum_{k=1}^K C_k(x|x_0) \frac{\Delta^{k-1}}{(k-1)!}.$$

Before computing the coefficients of the expansion, reconsider Example 6.

**Example 7.** Consider a Brownian motion process with a constant drift (see Example 6). The transition density is known to be normal with mean  $x_0 + \Delta\mu$  and variance  $\Delta\sigma^2$ . The log density is:

$$\ell(x|x_0, \Delta) = \frac{1}{2} \left[ -\ln 2\pi - \ln \sigma^2 - \ln \Delta - \frac{(x - x_0 - \mu\Delta)^2}{\Delta\sigma^2} \right].$$

We may compute directly the coefficients of the small  $\Delta$  expansion:

$$\begin{aligned} C_{-1}(x|x_0) &= -\frac{(x - x_0)^2}{2\sigma^2} \\ \tilde{C}(x|x_0) &= -\frac{1}{2} \\ C_0(x|x_0) &= -\ln \sigma + \frac{(x - x_0)\mu}{\sigma^2} - \frac{1}{2} \ln 2\pi \\ C_1(x|x_0) &= -\frac{\mu^2}{2\sigma^2} \end{aligned}$$

More generally, these coefficients can be computed using the Kolmogorov forward and backward equations. In particular, the forward equation is typically stated in terms of the densities, but it has a log-density counterpart:

$$\begin{aligned} \frac{\partial \ell}{\partial \Delta}(x|x_0, \Delta) &= C^*(x) + \sum_{i=1}^m \mu_i(x) \frac{\partial \ell}{\partial x_i}(x|x_0, \Delta) + \sum_{i=1}^m \sum_{j=1}^m \frac{\partial \nu_{ij}(x)}{\partial x_i} \frac{\partial \ell}{\partial x_j}(x|x_0, \Delta) \\ &+ \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \nu_{ij}(x) \frac{\partial^2 \ell}{\partial x_i \partial x_j}(x|x_0, \Delta) \\ &+ \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial \ell}{\partial x_i}(x|x_0, \Delta) \nu_{ij}(x) \frac{\partial \ell}{\partial x_j}(x|x_0, \Delta) \end{aligned} \quad (4.11)$$

where

$$C^*(x) \doteq -\sum_{i=1}^m \frac{\partial \mu_i(x)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 \nu_{ij}(x)}{\partial x_i \partial x_j}$$

This differential equation is linear in the second derivative of  $\ell$  with respect to  $x$  but quadratic in the first derivative.

#### 4.4.2 Leading Term

The leading term in this expansion must solve:

$$-C_{-1}(x|x_0) = \frac{1}{2} \left[ \frac{\partial C_{-1}(x|x_0)}{\partial x} \right]' \nu(x) \left[ \frac{\partial C_{-1}(x|x_0)}{\partial x} \right]. \quad (4.12)$$

This follows because the lowest power in  $\Delta$  on the left-hand side of (4.11) is  $-2$ . Only the last term on the right-hand side contributes to this. We consider the solution that has a strict maximum at  $x = x_0$ .

**Example 8.** Suppose that  $\nu(x) = I$ . Ait-Sahalia (2001) discusses when the state can be transformed so that this restriction is satisfied. The differential equation (4.12) then has as a solution:

$$C_{-1}(x|x_0) = -\frac{|x - x_0|^2}{2}.$$

This suggests a transition density approximation of the form:

$$\exp\left(-\frac{|x - x_0|^2}{2\Delta}\right)$$

over an interval  $\Delta$ . In turn this suggests a normal approximation as the leading term. Since the leading term will not even approximately integrate to one, we will need to explore other terms of the expansion. In this example, by adding the expression

$$-\frac{m}{2} \ln \Delta - \frac{m}{2} \ln 2\pi$$

to the leading term ensures that the resulting approximation is a log density. In fact it is the log density of a multivariate normal with mean  $x_0$  and covariance matrix  $\Delta I$ .

Consider next a quadratic (in  $x - x_0$ ) approximation to the solution to equation (4.12) determining  $C_{-1}(x|x_0)$ . The linear term is necessarily zero when the matrix  $\nu$  is nonsingular. Write the second-order expansion as:

$$C_{-1}(x|x_0) \approx -\frac{1}{2}(x - x_0)'V(x - x_0).$$

Equation (4.12) implies the Riccati equation,

$$V = V\nu(x_0)V$$

with the solution of interest being:

$$V = \nu^{-1}(x_0).$$

As a consequence the leading term in the expansion is:

$$-\frac{1}{2\Delta}(x - x_0)'\nu(x_0)^{-1}(x - x_0)$$

implying an approximate density:

$$\exp \left[ -\frac{1}{2\Delta} (x - x_0)' \nu(x_0)^{-1} (x - x_0) \right]$$

when we localize in both the interval  $\Delta$  and  $x - x_0$ . Adding

$$-\frac{m}{2} (\ln \Delta + \ln 2\pi) - \frac{1}{2} \ln \det \nu(x_0)$$

scales the implied density approximation to integrate to one. The resulting density is normal with mean  $x_0$  and covariance matrix  $\Delta \nu(x_0)$ .

We will have more to say about the  $x - x_0$  component of the expansion subsequently.

#### 4.4.3 Next two terms

We now consider the implications of (4.11) for the next two terms in the small  $\Delta$  expansion. Adding a constant term in  $x$  does not alter the differential equation. Thus we do not expect that the coefficients will be fully determined from this equation alone.

To avoid higher-order terms in  $\ln \Delta$ , we look for solutions in which  $\tilde{C}(x|x_0)$  is independent of  $x$ . Using the previous discussion as motivation, we set

$$\tilde{C}(x|x_0) = -\frac{m}{2}.$$

In addition, we initialize  $C_0(x_0|x_0) = -\frac{1}{2} \ln \det \nu(x_0) - \frac{m}{2} \ln(2\pi)$ .

From the forward equation (4.11), we also have the restriction:

$$\begin{aligned} \tilde{C}(x|x_0) &= \sum_{i=1}^m \mu_i(x) \frac{\partial C_{-1}}{\partial x_i}(x|x_0) + \sum_{i=1}^m \sum_{j=1}^m \frac{\partial \nu_{ij}(x)}{\partial x_i} \frac{\partial C_{-1}}{\partial x_j}(x|x_0) \\ &+ \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \nu_{ij}(x) \frac{\partial^2 C_{-1}}{\partial x_i \partial x_j}(x|x_0) \\ &+ \sum_{i=1}^m \sum_{j=1}^m \frac{\partial C_{-1}}{\partial x_i}(x|x_0) \nu_{ij}(x) \frac{\partial C_0}{\partial x_j}(x|x_0) \end{aligned} \quad (4.13)$$

After substituting the solutions for  $C_1$  and  $\tilde{C}$ , this becomes a first-order partial differential equation in  $C_0(x|x_0)$ .

Recall that in Example 8, we set  $\nu = I$ . In this example differential equation (4.13) simplifies and is satisfied provided that:

$$\sum_{i=1}^m \frac{\partial C_0}{\partial x_i}(x|x_0) (x_i - x_{0i}) = -\sum_{i=1}^m \mu_i(x) (x_i - x_{0i}).$$

Integrating along a line segment between  $x_0$  and  $x$  we obtain:

$$C_0(x|x_0) = - \sum_{i=1}^m (x_i - x_{0i}) \int_0^1 \mu_i [x + u(x - x_0)] du$$

since  $\ln \det I = 0$ .

#### 4.4.4 Remaining Terms

There is a recursive structure to the remaining coefficients. Since the left-hand side of (4.11) entails the derivative with respect to  $\Delta$ , whereas the right-hand side does not,

$$\begin{aligned} C_{k+1}(x|x_0) &= C_k^*(x) + \sum_{i=1}^m \mu_i(x) \frac{\partial C_k}{\partial x_i}(x|x_0) + \sum_{i=1}^m \sum_{j=1}^m \frac{\partial \nu_{ij}(x)}{\partial x_i} \frac{\partial C_k}{\partial x_j}(x|x_0) \\ &\quad + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \nu_{ij}(x) \frac{\partial^2 C_k}{\partial x_i \partial x_j}(x|x_0) \\ &\quad + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \sum_{r=-1}^k \frac{\partial C_r}{\partial x_i}(x|x_0) \nu_{ij}(x) \frac{\partial C_{k-r}}{\partial x_j}(x|x_0). \end{aligned} \tag{4.14}$$

where  $C_0^* = C^*$  and  $C_j^* = 0$  for  $j \geq 1$ . Notice that the right-hand side has a term in

$$\frac{\partial C_{k+1}}{\partial x_j}(x|x_0)$$

obtained when  $r = -1$ . The remaining terms are computed as simple functions of derivatives of lower order coefficients. Thus we are again left with a differential equation to solve, but it is an equation that is linear in this derivative and not quadratic as in partial differential equation (4.12) for  $C_{-1}(x|x_0)$ . We are interested in solutions for which  $C_{k+1}(x_0|x_0) = 0$ .

#### 4.4.5 Expansions in Powers of $x - x_0$

Typically one cannot solve the differential equation (4.14). Instead, we can compute the coefficients of an expansion in powers of  $x - x_0$  that is guaranteed to be accurate for  $x$  close to  $x_0$ . After constructing an expansion to a given order of each coefficient  $C_j(x|x_0)$ , the result is a joint expansion in  $\Delta$  and  $x - x_0$ .

Like the expansion in  $\Delta$ , a polynomial expansion of  $C_j(x|x_0)$  can be computed explicitly in powers of  $x - x_0$ : see Ait-Sahalia (2001) for details, and the order at which to expand the coefficient  $C_j$ . These Taylor expansions of  $C_j(x|x_0)$  may be computed by solving systems of linear equations with one exception, which fortunately also has an explicit expansion in  $x - x_0$ .

Consider the equation (4.12) determining  $C_{-1}(x|x_0)$ . As we have previously argued the first nonzero term in the expansion is quadratic:

$$C_{-1}(x|x_0) = -\frac{1}{2}(x - x_0)' \nu(x_0)^{-1} (x - x_0).$$

obtained by solving a Riccati equation. The higher-order terms  $(x - x_0)$  for  $C_{-1}$  can be calculated by solving linear equations, however.

In conclusion, combining expansions in  $\Delta$  and  $x - x_0$ , as described in Aït-Sahalia (2001), provides a sequence of local approximations to the logarithm  $\ell(x|x_0, \Delta)$  of the transition density. These expansions can be computed conveniently by evaluating derivatives of the drift and diffusion coefficients and, for all but one term, solving linear equations.

## 5 Observable Implications

We have just seen how to characterize transition densities of Markov processes. In this section we explore the inverse problem. Suppose from data we can infer information about transitions. Could these data have come from special classes of continuous-time Markov processes? What are the observable implications of the special types of Markov processes?

### 5.1 Local Characterization

By its very nature the generator gives a local counterpart to conditional moment restrictions. It gives us a formal sense in which:

$$\frac{E\phi(x_{t+\Delta}|x_t) - \phi(x_t)}{\Delta} \approx \mathcal{A}\phi(x_t).$$

Thus estimation of the left-hand side allows for the approximation of  $\mathcal{A}$ . By looking at appropriately chosen families of test functions we can learn about  $\mathcal{A}$  provided discretization errors are small.

First, we consider the identification scheme advocated by Johannes (2004). Consider first linear test functions parameterized as  $\phi(x) = a \cdot (x - x^*)$  for some  $a \in \mathbb{R}^m$  and some  $x^*$ . Then

$$\mathcal{A}\phi(x) = a \cdot \mu(x) + a \cdot \left[ \int (y - x^*) R(dy|x) - (x - x^*) \right].$$

Evaluating this at  $x = x^*$  gives:

$$\mathcal{A}\phi(x^*) = a \cdot \mu(x^*) + a \cdot \int (y - x^*) R(dy|x^*).$$

By letting  $a$  be each of the coordinate vectors we identify:

$$\mu(x^*) + \int (y - x^*) R(dy|x^*).$$

Using an entirely similar argument for quadratic functions of the form  $(x - x^*)'V(x - x^*)$  for symmetric matrices  $V$ , we may infer

$$\nu(x^*) + \int (y - x^*)(y - x^*)'R(dy|x^*).$$

More generally, higher-order polynomials centered around  $x^*$  will reveal higher-order moments of the conditional jump distribution scaled by the jump intensity. The drift and diffusion will only contribute to the first two conditional moments. Johannes (2004) used this observation to infer the importance of jump components in interest rates.

Polynomials will sometimes not be in the domain of the generator. Other collections of localized test functions can be employed in making these approximations. For instance,  $a \cdot (x - x^*)$  might be replaced by  $\phi(x) = a \cdot (x - x^*)\psi(|x - x^*|^2)$  where  $\psi$  is a symmetric twice continuously differentiable function that is one at zero and has compact support. Notice that the derivative of this test function at  $x = x^*$  is  $a$ . In the absence of jumps,

$$\mathcal{A}\phi(x^*) = a \cdot \mu(x^*).$$

Similarly, when  $\phi(x) = (x - x^*)'V(x - x^*)\psi(|x - x^*|^2)$ ,

$$\mathcal{A}\phi(x^*) = \text{trace}[\nu(x^*)V]$$

which can be used to identify  $\nu$ .

Given that the diffusion component is a local operator, localization of first and second-order polynomials continues to permit the identification of the drift and the diffusion coefficients. When the jump component is present, we must add corrections that depend more specifically on the function  $\psi$  used in localization. The corrections will cease to be conditional moments of the jump distribution scaled by the jump intensity parameter  $\lambda$ .

Finally, in the absence of jump components we may also use a localization that is not smooth. For instance, the infinitesimal parameters can be recovered using the familiar formulas:

$$\begin{aligned} \mu(x^*) &= \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int_{|y-x^*|<\varepsilon} (y - x^*) P_{\Delta}(x^*, dy) \\ \nu(x^*) &= \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int_{|y-x^*|<\varepsilon} (y - x^*)(y - x^*)' P_{\Delta}(x^*, dy) \end{aligned}$$

where  $Q_{\Delta}$  is the transition distribution for the diffusion process. Florens-Zmirou (1984), Stanton (1997), Fang and Zhang (2003), Bandi (2002), Bandi and Phillips (2003), and others consider estimation of diffusion based on these local conditional moment restrictions.

## 5.2 Total Positivity

The local characterizations are justified by taking a limit as  $\Delta \rightarrow 0$ . We now examine what can be said if the process is only observed at a finite observation interval  $\Delta$  but arbitrarily large sample sizes. Let  $\mathbb{R}$  be the state space for a Markov process, and consider a family of probability distributions indexed by the time interval  $\Delta$ :  $P_\Delta(\cdot|x)$ . Could this family of densities have come from a scalar diffusion process, i.e., a scalar Markov process with continuous sample paths, or must a more general process be considered? Karlin and McGregor (1959a) address this question using a set of *total positivity* restrictions on transition densities, and Ait-Sahalia (2002c) develops and justifies statistical tests based on these restrictions.

While total positivity has a more general representation and probabilistic interpretation, it implies

$$P_\Delta(x, B) P_\Delta(\tilde{x}, \tilde{B}) - P_\Delta(\tilde{x}, B) P_\Delta(x, \tilde{B}) > 0 \quad (5.1)$$

whenever,  $x < \tilde{x}$  and  $B < \tilde{B}$  (where  $B < \tilde{B}$  is interpreted to mean that every element of  $B$  is less than every element of  $\tilde{B}$ ). Since this must hold for any choice of  $\tilde{x}$  and  $\tilde{B}$ , there is a local (in the state) counterpart that we express using the logarithm of the density:

$$\frac{\partial^2}{\partial x \partial y} \ell(y|x, \Delta) > 0 \quad (5.2)$$

for all  $x$  and  $y$  and interval  $\Delta$ . This cross derivative restriction for each choice of  $x$ ,  $y$  and  $\Delta$  is a necessary condition for transition distributions to be those implied by a scalar diffusion.

A partial converse is also available. Suppose that the family of distribution functions of a Markov process on  $\mathbb{R}$  satisfies (5.1) for any positive  $\Delta$ . Then under a side condition, there exists a realization of the process such that almost all sample paths are continuous.

The following example shows how criterion (5.2) can be used to eliminate some transition densities as coming from a model of a scalar diffusion.

**Example 9.** *Suppose that  $\ell(y|x, \Delta)$  depends on the composite state  $(y, x)$  only through  $y - x$ . Then criterion (5.2) is equivalent to requiring that  $\ell$  be concave in  $y - x$ . It can be shown that the only admissible solution is*

$$\ell(y|x, \Delta) = -\frac{1}{2} \ln(2\pi\beta^2\Delta) - \frac{(y - x - \alpha\Delta)^2}{2\beta^2\Delta}$$

where  $\alpha$  and  $\beta$  are free parameters. That is the transition density is an arithmetic Brownian motion.

As an alternative, consider the generalized Cauchy density

$$\ell(y|x, \Delta) = -\ln \pi + \ln \alpha(\Delta) - \ln [\alpha(\Delta)^2 + (y - x)^2]$$

where  $\alpha(\Delta)$  is positive. Criterion (5.2) fails for large  $y - x$ .

Aït-Sahalia (2002c) contains other examples.

More generally, total positivity implies restrictions on processes defined on state spaces other than  $\mathbb{R}$ . Consider a continuous-time, stationary, Markov chain that can only take countable discrete values, say,  $\{\dots, -1, 0, 1, \dots\}$ . Karlin and McGregor (1959b) study when such a process has the counterpart to a continuous sample path? In a discrete state space, the appropriate notion of continuity of the chain's sample paths is the following intuitive one: the chain never jumps by more than one state at a time, either up or down. It turns out that the restriction on the chain's transition probabilities analogous to (5.1) characterizes precisely this form of continuity: total positivity across all intervals restricts the process to be a so called *birth-and-death* process (see Karlin and McGregor (1959b)). In this sense, a birth-and-death process is the discrete-state analog to a scalar diffusion. See Aït-Sahalia (2002c) for further discussion and implications for derivative pricing methods, such as binomial trees.

For a fixed  $\Delta$ , total positivity is a necessary restriction on the transition distribution but not a sufficient one. Given a candidate transition distribution over an interval  $\Delta$ , we did not construct a diffusion with that transition density. Frydman and Singer (1979) study the analogous question for a finite state birth and death process. In their study they show that to embed a single transition matrix (over an interval  $\Delta$ ) satisfying total positivity in a continuous-time Markov process it is sometimes necessary that the continuous-time process be time-inhomogeneous. They show that the total positivity function is a weaker restriction than embeddability for a continuous-time process that is restricted to be time-homogeneous.

### 5.3 Principal Component Approach

We now explore an alternative approach to the embeddability question in the context of scalar diffusions: when does there exist a (time-homogeneous) scalar diffusion process that is consistent with a given discrete-time transition distribution? We follow Hansen, Scheinkman, and Touzi (1998) by answering this question using a principal component decomposition. As we have seen, the existence of this decomposition is restrictive.

First, consider a scalar diffusion with stationary density  $q$  and diffusion coefficient  $\sigma^2$ . As we have seen there is a corresponding form constructed with these objects. Each principal component satisfies the eigenvalue relation:

$$\frac{1}{2} \int \phi' \psi_j' \sigma^2 q = \delta_j \int \phi \psi_j q.$$

for any  $\phi$  that is twice continuously differentiable for which  $\phi'$  has compact support. An integration-by-parts argument implies that

$$\psi_j'(x) \sigma^2(x) q(x) = -2\delta_j \int^x \psi_j q \tag{5.3}$$

since  $\phi'$  can be localized at the point  $x$ .

To achieve identification, we must construct  $\sigma^2$  from a discrete-time transition operator. The density  $q$  and the principal components  $\psi_j$  and associated eigenvalues  $\delta_j$  are identifiable from discrete-time data. The principle components are identifiable because they maximize autocorrelation. Moreover, they satisfy the discrete-time conditional moment restriction:<sup>13</sup>

$$E [\phi_j(X_{t+\Delta})|X_t] = \exp(-\Delta\delta_j)\phi_j(X_t).$$

We can think of (5.3) as a set of restrictions that can be used to infer  $\sigma^2$ . While  $\sigma^2$  can be identified from one of these equations (except for the constant eigenfunction equation), over-identification comes from the fact that the same  $\sigma^2$  must work for all eigenfunctions.<sup>14</sup> Since  $\sigma^2$  is restricted to be positive, there is a testable implication for even a single choice of  $j$  in (5.3) provided the constant eigenfunction is not used. Unfortunately, statistical testing is likely to be more challenging for testing eigenfunction restrictions than for testing for total positivity.

## 5.4 Testing the Specification of Transitions

The generator of a semigroup commutes with the family of conditional expectation operator that it generates:

$$\mathcal{A}\mathcal{T}_\Delta\phi = \mathcal{T}_\Delta\mathcal{A}\phi \tag{5.4}$$

for any  $\phi$  in the domain of the generator and any  $\Delta$ . This follows from the semigroup property (2.2) and the construction of the generator as the time derivative of the semigroup (at  $t = 0$ ). As emphasized by Hansen and Scheinkman (1995), this gives rise to a set of testable restrictions beyond stationarity which we now explore.

From an abstract perspective, given a candidate generator  $\hat{\mathcal{A}}$  (not necessarily  $\mathcal{A}$ ) and a conditional expectation operator  $\mathcal{T}_\Delta$  suppose

$$\hat{\mathcal{A}}\mathcal{T}_\Delta\phi = \mathcal{T}_\Delta\hat{\mathcal{A}}\phi \tag{5.5}$$

---

<sup>13</sup>An alternative parametric identification and inference approach is suggested by Kessler and Sorensen (1999). They use the fact that principal components satisfy this conditional moment restriction to build estimating equations for parameterized diffusions.

<sup>14</sup>There is a close relation between recovery formula (5.3) and formula (3.9) that we described previously. Suppose that a linear function is in the domain of the generator, the drift is linear. Then the drift coefficient is an eigenfunction and the corresponding value of  $\delta$  is the negative of the derivative of this function. With these substitutions, the two recovery formulas coincide. Demoura (1998) suggests a similar identification by looking across two distinct eigenfunctions and there first two derivatives to identify the pair  $(\mu, \sigma^2)$ . In contrast, recovery formula (5.3) avoids using second derivatives and instead uses a single eigenfunction in conjunction with the stationary density.

for any  $\phi$  among a rich collection of test functions (formally a core of the generator). In what way does this restrict the candidate  $\hat{\mathcal{A}}$ ? How might we actually test this implication?

If the candidate generator  $\hat{\mathcal{A}}$  commutes with  $\mathcal{A}$ , then  $\hat{\mathcal{A}}$  cannot be distinguished from  $\mathcal{A}$  on the basis of (5.4). In particular, when  $\hat{\mathcal{A}}$  is a scalar multiple of  $\mathcal{A}$ , they commute and hence cannot be distinguished. Thus the most one can hope for is from (5.5) is the identification of the generator up to scale. As illustrated by Hansen and Scheinkman (1995), without further restrictions, the identification problem can be more severe than this. On the other hand, Hansen and Scheinkman (1995) show that stationary scalar diffusions can be identified up to scale by (2.2) and the information encoded in the stationary distribution.

Stationary scalar diffusions are examples of processes that are reversible. More generally, stationary Markov processes modelled via symmetric forms are reversible. Such models are identifiable from discrete time data sampled at any fixed interval  $\Delta$ .<sup>15</sup> Thus the commuting restriction does not encode all of the identifying information contained in the transition distribution.

For reversible Markov process models, there is an equivalent statement of restriction (5.5):

$$E \left( \left[ \hat{\mathcal{A}}\phi(x_{t+1}) \right] \psi(x_t) \right) = E \left( \phi(x_{t+1}) \left[ \hat{\mathcal{A}}\psi(x_t) \right] \right) \quad (5.6)$$

for  $\phi$  and  $\psi$  in the domain of  $\hat{\mathcal{A}}$ . The restriction can be tested via statistical methods by focusing on a limited number of test functions, or it can be made comprehensive by adapting the approach of Bierens (1990). In what follows we describe a different approach.

For stationary diffusions, there is an equivalent statement of restriction (5.5) deduced by Aït-Sahalia (1996b). In contrast to (5.6) reversibility is not required. We may deduce this directly from the Kolmogorov forward and backward equations as in Aït-Sahalia (1996b). Alternatively (and essentially equivalently) we may localize the test function  $\phi$  in (5.5). Let  $\hat{\mathcal{A}}$  be a candidate generator of a diffusion with drift  $\hat{\mu}$  and diffusion matrix  $\hat{\nu}$ . After localization, the left-hand side of (5.5) becomes:

$$\sum_i \hat{\mu}_i(x) \frac{\partial}{\partial x_i} p(y|x, \Delta) + \frac{1}{2} \sum_{i,j} \hat{\nu}_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} p(y|x, \Delta).$$

Prior to localizing the right-hand side of (5.5), we apply integration by parts to a test function with compact support in the interior of the state space and write:

$$\begin{aligned} \mathcal{T}_\Delta \hat{\mathcal{A}}\phi(x) = & - \int \left[ \sum_i \frac{\partial}{\partial y_i} \hat{\mu}_i(y) p(y|x, \Delta) \right] \phi(y) dy \\ & + \frac{1}{2} \int \left[ \sum_{i,j} \frac{\partial^2}{\partial y_i \partial y_j} \hat{\nu}_{ij}(y) p(y|x, \Delta) \right] \phi(y) dy \end{aligned}$$

---

<sup>15</sup>See Proposition 5 in Hansen and Scheinkman (1995).

By localizing the test function around a given value of  $y$ , it follows from (5.5) that

$$\begin{aligned} & \sum_i \hat{\mu}_i(x) \frac{\partial}{\partial x_i} p(y|x, \Delta) + \frac{1}{2} \sum_{i,j} \hat{\nu}_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} p(y|x, \Delta) \\ &= - \sum_i \frac{\partial}{\partial y_i} [\hat{\mu}_i(y) p(y|x, \Delta)] + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial y_i \partial y_j} [\hat{\nu}_{ij}(y) p(y|x, \Delta)]. \end{aligned} \quad (5.7)$$

The left-hand side of the inequality is the contribution of the Kolmogorov forward equation and the right-hand side is the contribution from the backward equation. Aït-Sahalia (1996b) calls the difference between the left-hand and right-hand side the *transition discrepancy*.

Restrictions (5.6) or (5.7) could in principle be used to identify a scalar diffusion non-parametrically up to a free scale parameter on the drift and diffusion coefficients. They are also of value in estimating and testing parameterized diffusions processes (again up to free scale parameter). Restriction (5.6) avoids having to estimate second derivatives of transition densities, but it is applicable only to reversible processes and requires a specific selection of test functions.<sup>16</sup> Restriction (5.7) gives rise to a comprehensive test formalized by choosing a weighting function to use in conjunction with the discrepancy measure. The statistically efficient choices of test functions or weighting functions have not been formally analyzed to date.

## 5.5 Testing Markovianity

The specification analysis described above assumes that the process is Markovian. Can this be tested? A continuous time Markov process sampled with an interval  $\Delta$  is a discrete-time Markov process. One common approach to test a discrete-time Markov process is to include additional lags of the state vector into the state evolution equation and test for their statistical significance. Following Aït-Sahalia (2002a), we consider an alternative approach based on the Chapman-Kolmogorov equation given in Definition 2.

Under time-homogeneity, an implication of the Chapman-Kolmogorov equation is that  $\mathcal{T}_{2\Delta} = (\mathcal{T}_\Delta)^2$  as required by the semigroup property. Stated in terms of transition densities, the Markov hypothesis can be tested the form  $H_0$  against  $H_1$ , where

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

with

$$r(y|x, 2\Delta) \doteq \int_{z \in S} p(y|z, \Delta) p(z|x, \Delta) dz. \quad (5.8)$$

---

<sup>16</sup>Hansen and Scheinkman (1995) derive a more general counterpart based also on the generator of the reverse-time process.

Both  $p(y|x, \Delta)$  and  $p(y|x, 2\Delta)$  can be estimated from data sampled at interval  $\Delta$ . The successive pairs of observed data  $(X_0, X_\Delta)$ ,  $(X_\Delta, X_{2\Delta})$ ,  $(X_{2\Delta}, X_{3\Delta})$ , *etc.*, can be used to estimate the density  $p(y|x, \Delta)$  and hence the function  $r$  given by (5.8). Meanwhile, the successive pairs  $(x_0, x_{2\Delta})$ ,  $(x_\Delta, x_{3\Delta})$ ,  $\dots$ , can be used to estimate directly the density  $p(y|x, 2\Delta)$ . In other words, the test compares a *direct* estimator of the  $2\Delta$ -interval conditional density, with the *indirect* estimator of the  $2\Delta$ -interval conditional density based on formula (5.8). If the process is actually Markovian, then the two estimates should be close (for some distance measure) in a sense made precise by the use of the statistical distribution of these estimators.

More generally we could study the  $j\Delta$  transitions where  $j$  is an integer greater than or equal to 2. For larger  $j$ , there are more options for comparison. A test could be based on constructing a  $j\Delta$  period transition could be constructed from shorter ones including shorter ones including  $(\Delta, (j-1)\Delta)$ ,  $(2\Delta, (j-2)\Delta)$ ,  $\dots$  or even from more than two shorter transitions. It is not necessary to check all of these configurations as many will be redundant. In general, a vector of *transition equalities* can be tested in a single pass in a GMM framework with as many moment conditions as transition intervals.

## 5.6 Testing Symmetry

The symmetry of the transition distribution implied by our use of forms to build Markov processes is restrictive. This restriction has motivated the construction of tests of symmetry and as we have seen more general formulations that allow for asymmetry. In one important special case symmetry is not limiting: scalar diffusions on the real line. In higher dimensions, however, symmetry is restrictive even for diffusions. When a Markov process is stationary, the symmetry implied by the forms implies that the process is time reversible when initialized at the stationary distribution. Reversible Markov processes are identifiable from discrete-time data, even without parametric restrictions. There is no aliasing problem for these processes. See Hansen and Scheinkman (1995) for a discussion.

Florens, Renault, and Touzi (1998) propose a test for reversibility as a necessary condition to embed a stationary, reversible continuous-time process in a discrete time process sampled at regular intervals. Their idea is the following. A reversible process should display positive autocorrelation in the following sense. For any test function  $\phi$ ,

$$E\phi(X_t)\phi(X_{t+s}) \geq 0$$

for any interval  $s$ . (See the Theorem in Florens, Renault, and Touzi (1998) on page 75.) To build a statistical test, use a vector of such functions, which we denote by  $\Phi$ . Form the symmetrized autocovariance matrix:

$$\frac{1}{2} [E\Phi(X_t)\Phi(X_{t+1})' + E\Phi(X_{t+1})\Phi(X_t)'] . \quad (5.9)$$

While this matrix has real eigenvalues by construction, the eigenvalues should all be positive if the discretely sampled process can be embedded in a continuous-time, reversible Markov process. Since all linear combinations of test functions in  $\Phi$  should show positive persistence, eigenfunctions should also display positive persistence. Thus eigenvalues must be positive. Florens, Renault, and Touzi (1998) suggest building a test based on the smallest eigenvalue of the sample analog to (5.9).

An alternative approach to testing reversibility is given by Darolles, Florens, and Gouriéroux (2000). It is based on nonlinear canonical analysis of the joint density of adjacent observations, say  $(X_t, X_{t+1})$ . With limitations on the temporal dependence, canonical analysis produces principal component pairs of functions say  $\phi(X_t)$  and  $\psi(X_{t+1})$  that maximize correlation under orthogonality constraints. This becomes a nonlinear analysis because the functions  $\phi$  and  $\psi$  can be nonlinear in the Markov state. These principal components can be used to construct an orthogonal decomposition of the joint density. Dauxois and Nkiet (1998) use canonical analysis as a test of independence between two random vectors and Darolles, Florens, and Gouriéroux (2000) use it to produce a test of reversibility. Their statistical tests are based on the restrictions that reversibility imposes on the canonical analysis. Under reversibility, the two functions  $(\phi, \psi)$  in each orthogonal pair should coincide.

## 5.7 Strong Dependence

The existence of a principal component decomposition typically requires that the underlying Markov process be only weakly dependent. For a weakly dependent process, autocorrelations of test functions decay exponentially. It is possible, however, to build models of Markov processes that are strongly dependent. For such processes, the autocorrelations of some test functions decay at a slower than exponential rate. Operator methods give a convenient way to characterize when a process is strongly dependent.

In our study of strongly dependent, but stationary, Markov processes, we follow Chen, Hansen, and Carrasco (2003) by using two measures of mixing. Both of these measures have been used extensively in the stochastic process literature. The first measure,  $\rho$ -mixing uses the  $L^2(Q)$  formulation. Let

$$U \doteq \left\{ \phi \in L^2(Q) : \int \phi dQ = 0, \int \phi^2 dQ = 1 \right\}.$$

The concept of  $\rho$ -mixing studies the maximal correlation of two functions of the Markov state in different time periods.

**Definition 7.** *The  $\rho$ -mixing coefficients of a Markov process are given by:*

$$\rho_t = \sup_{\psi, \phi \in U} \int \psi(\mathcal{I}_t \phi) dQ.$$

The process  $\{X_t\}$  is  $\rho$ -mixing or weakly dependent if  $\lim_{t \rightarrow \infty} \rho_t = 0$ .

When the  $\rho$ -mixing coefficients of a Markov process decline to zero, they do so exponentially. When a Markov process has a principal component decomposition, it is  $\rho$ -mixing with exponential decay. In fact,  $\rho$ -mixing requires something weaker.

As argued by Banon (1978) and Hansen and Scheinkman (1995),  $\rho$ -mixing is guaranteed by a *gap* in spectrum of the positive semidefinite operator  $-\mathcal{A}$  to the right of zero. Although not always symmetric, the operator  $\mathcal{A}$  is negative semidefinite:

$$\int \phi(\mathcal{A}\phi)dQ \leq 0$$

on the  $L^2(Q)$  domain of  $\mathcal{A}$ . This negative-semidefinite property follows from the restriction that  $\mathcal{T}_t$  is a weak contraction on  $L^2(Q)$  for each  $t$ . A spectral gap is present when we can strengthen this restriction as follows:

$$\inf_{\phi \in U \cap D(\mathcal{A})} - \langle \phi, \mathcal{A}\phi \rangle > 0. \quad (5.10)$$

When this condition is satisfied  $\mathcal{T}_t$  is a strong contraction on the subspace  $U$  for each  $t$ , and the  $\rho$ -mixing coefficients decay exponentially.

In the case of a scalar diffusion, Hansen and Scheinkman (1995) show that this inequality is satisfied provided that

$$\begin{aligned} \limsup_{x \rightarrow r} \frac{\mu}{\sigma} - \frac{\sigma'}{2} &= \limsup_{x \rightarrow r} \frac{\sigma q'}{2q} + \frac{\sigma'}{2} < 0 \\ \liminf_{x \rightarrow \ell} \frac{\mu}{\sigma} - \frac{\sigma'}{2} &= \liminf_{x \rightarrow \ell} \frac{\sigma q'}{2q} + \frac{\sigma'}{2} > 0. \end{aligned} \quad (5.11)$$

where  $r$  is the right boundary and  $\ell$  is the left boundary of the state space. This restriction is a weakening of restriction (4.5), which guaranteed the existence of principal components. Condition (5.11) guarantees that there is sufficient pull from each boundary towards the center of the distribution to imply  $\rho$ -mixing. When one of these two limits is zero, the  $\rho$ -mixing coefficients may be identically equal to one. In this case the Markov process is strongly dependent.<sup>17</sup>

Since the  $\rho$ -mixing coefficients for a Markov process either decay exponentially or are equal to one, we need a different notion of mixing to obtain a more refined analysis of strong dependence. This leads us to consider the  $\beta$ -mixing coefficients:

---

<sup>17</sup>Recall that the term in the left-hand side of (5.11) can be interpreted as the drift of a corresponding diffusion with a unit diffusion coefficient obtained by transforming the scale. As a consequence, condition (5.11) can also be related to Veretennikov (1997)'s drift restriction for a diffusion to be strongly dependent.

**Definition 8.** The  $\beta$ -mixing coefficients for a Markov process are given by:

$$\beta_t = \int \sup_{0 \leq \phi \leq 1} |\mathcal{T}_t \phi - \int \phi dQ| dQ.$$

The process  $\{X_t\}$  is  $\beta$ -mixing if  $\lim_{t \rightarrow \infty} \beta_t = 0$ ; is  $\beta$ -mixing with an exponential decay rate if  $\beta_t \leq \gamma \exp(-\delta t)$  for some  $\delta, \gamma > 0$ .

At least for scalar diffusions, Chen, Hansen, and Carrasco (2003) show that the exponential decay of the  $\rho$ -mixing coefficients is essentially equivalent to the exponential decay of the  $\beta$ -mixing coefficients. When the  $\rho$ -mixing coefficients are identically one, however, the  $\beta$ -mixing coefficients will still decay to zero, but at a rate slower than exponential. Thus the decay properties of the  $\beta$ -mixing coefficients provides a more sensitive characterization of strong dependence.

## 5.8 Maximum Likelihood Estimation

Parametric models of continuous-time Markov processes can be estimated via maximum likelihood (ML) estimation from discrete-time data. In the past, a computational challenge for such estimation was the tractable construction of a likelihood function since existing methods required solving numerically the Fokker-Planck-Kolmogorov partial differential equation (Lo (1988)), approximating the transition function by a Gaussian density (Kessler (1997)) or simulating a large number of sample paths along which the process is sampled very finely (Pedersen (1995)). Ait-Sahalia (2002b) shows how to approximate conveniently ML estimators of diffusions using the Hermite approach described in Section 4.3.

A fixed interval sample of a time-homogenous continuous-time Markov process is a Markov process in discrete time. Given that the Markov state vector is observed and the unknown parameters are identified, properties of the ML estimator follow from what is known about ML estimation of discrete-time Markov processes.<sup>18</sup> There is an extensive literature applicable to discrete-time stationary Markov processes starting with the work of Billingsley (1961). The asymptotic covariance matrix for the ML estimator is the inverse of the score covariance or *information* matrix where the score at date  $t$  is  $\partial \ln p(X_{t+\Delta}|X_t, \Delta, \theta)/\partial \theta$  where  $\ln p(\cdot|x, \Delta, \theta)$  is the logarithm of the conditional density over an interval of time  $\Delta$  and a parameter value  $\theta$ .

When the underlying Markov process is nonstationary, the score process inherits this nonstationarity. The rate of convergence and the limiting distribution of the maximum likelihood

---

<sup>18</sup>Identification of a *multivariate* continuous-time Markov process from discrete-time can be problematic when the process is not reversible. It is well known that an *aliasing problem* can be present. For example, see Phillips (1973) and Hansen and Sargent (1983).

estimator depends upon growth properties of the score process (*e.g.* see Hall and Heyde (1980) Chapter 6.2). A nondegenerate limiting distribution can be obtained when the score process behaves in a sufficiently regular fashion. The limiting distribution can be deduced by showing that general results pertaining to time series asymptotics (see *e.g.*, Jeganathan (1995)) can be applied to the present context. One first establishes that the likelihood ratio has the locally asymptotically quadratic (LAQ) structure, then within that class separates between the locally asymptotically mixed Normal (LAMN), locally asymptotically Normal (LAN) and locally asymptotically Brownian functional (LABF) structures. As we have seen, when the data generating process is stationary and ergodic, the estimation is typically in the LAN class. The LAMN class can be used to justify many of the standard inference methods given the ability to estimate the covariance matrix pertinent for the conditional normal approximating distribution. Rules for inference are special for the LABF case. These structures are familiar from the linear time series literature on unit roots and co-integration. Details for the case of a nonlinear Markov process can be found in Ait-Sahalia (2002b).

**Example 10.** *As an example of the types of results that can be derived, consider the Ornstein-Uhlenbeck specification,  $dX_t = -\kappa X_t dt + \sigma dW_t$ , where  $\theta = (\kappa, \sigma^2)$ . The discrete-time process obtained by sampling at a fixed interval  $\Delta$  is a Gaussian first-order autoregressive process with autoregressive parameter  $\exp(-\kappa\Delta)$  and innovation variance  $\frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa\Delta})$ . White (1958) and Anderson (1959) originally characterized the limiting distribution for the discrete-time autoregressive parameter when the Markov process is not stationary. Alternatively, by specializing the general theory of the limiting behavior of the ML estimation to this model, one obtains the following asymptotic distribution for the the ML estimator of the continuous-time parameterization (see Corollary 2 in Ait-Sahalia (2002b)):*

- If  $\kappa > 0$  (LAN, stationary case):

$$\begin{aligned} & \sqrt{N} \left( \begin{pmatrix} \hat{\kappa}_N \\ \hat{\sigma}_N^2 \end{pmatrix} - \begin{pmatrix} \kappa \\ \sigma^2 \end{pmatrix} \right) \\ \Rightarrow & \text{Normal} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{e^{2\kappa\Delta}-1}{\Delta^2} & \frac{\sigma^2(e^{2\kappa\Delta}-1-2\kappa\Delta)}{\kappa\Delta^2} \\ \frac{\sigma^2(e^{2\kappa\Delta}-1-2\kappa\Delta)}{\kappa\Delta^2} & \frac{\sigma^4((e^{2\kappa\Delta}-1)^2+2\kappa^2\Delta^2(e^{2\kappa\Delta}+1)+4\kappa\Delta(e^{2\kappa\Delta}-1))}{\kappa^2\Delta^2(e^{2\kappa\Delta}-1)} \end{pmatrix} \right) \end{aligned}$$

- If  $\kappa < 0$  (LAMN, explosive case), assume  $X_0 = 0$ , then:

$$\begin{aligned} \frac{e^{-(N+1)\kappa\Delta}}{e^{-2\kappa\Delta}-1} (\hat{\kappa}_N - \kappa) & \Rightarrow G^{-1/2} \times \text{Normal}(0, 1) \\ \sqrt{N} (\hat{\sigma}_N^2 - \sigma^2) & \Rightarrow \text{Normal}(0, 2\sigma^4) \end{aligned}$$

where  $G$  has a  $\chi^2[1]$  distribution independent of the  $\text{Normal}(0, 1)$ .  $G^{-1/2} \times \text{Normal}(0, 1)$  is a Cauchy distribution.

- If  $\kappa = 0$  (LABF, unit root case), assume  $X_0 = 0$ , then:

$$N \hat{\kappa}_N \Rightarrow (1 - W_1^2) \left( 2\Delta \int_0^1 W_t^2 dt \right)^{-1}$$

$$\sqrt{N} (\hat{\sigma}_N^2 - \sigma^2) \Rightarrow \text{Normal}(0, 2\sigma^4)$$

where  $N$  is the sample size,  $\{W_t : t \geq 0\}$  is a standard Brownian motion and  $\Rightarrow$  denotes convergence in distribution.

## 6 Applications

In this section we consider a variety of applications of the operator methods we have described in previous sections. These applications show some of the potency of operator methods as modelling devices.

### 6.1 Zipf's Law

Zipf suggested a generalization of his law in which there was a free parameter that related rank to size. Consider a family of stationary densities that satisfy a power law of the form:  $q_\xi \propto x^{-(2+\xi)}$  defined on  $(y, \infty)$  where  $y > 0$  and  $\xi \geq 0$ . Then the rank-size relation becomes  $\text{size}(\text{rank})^{\frac{1}{1+\xi}} = \text{constant}$ . This family of densities is of interest to economists, because of power-law distributions that seem to describe income distribution and city sizes. We continue to set  $\sigma^2(x) = \alpha^2 x^2$ . The corresponding drift is, using equation (3.3),

$$\mu = -\frac{\xi \alpha^2 x}{2}$$

Notice that  $\mu(y) < 0$ , so that  $y > 0$  is an attainable boundary. We make this barrier reflexive to deliver the requisite stationary density.

To study temporal dependence, we consider the pull measure:

$$\frac{\mu}{\sigma} - \frac{\sigma'}{2} = -\frac{\alpha(1+\xi)}{2},$$

which is negative and independent of the state. The negative pull at the right boundary in conjunction with the reflexive left boundary guarantees that the process has a spectral gap and thus it is weakly dependent even in the case where  $\xi = 0$ . Since the pull measure is constant, it fails to satisfy restriction (4.5). The full principal component decomposition we described in section 4.1 fails to exist because the boundary pull is insufficient.

## 6.2 Stationarity and Volatility

Nonlinearity in a Markov diffusion coefficient changes the appropriate notion of mean reversion. Stationarity can be induced by how volatility changes as a function of the Markov state and may have little to do with the usual notion of mean reversion as measured by the drift of the diffusion process. This phenomenon is most directly seen in scalar diffusion models in which the drift is zero, but the process itself is stationary. Conley, Hansen, Luttmer, and Scheinkman (1997) generalize this notion by arguing that for processes with an infinite right boundary, stationarity is volatility induced when:

$$\int_x^\infty \frac{\mu(y)}{\sigma^2(y)} dy > -\infty \quad (6.1)$$

for some  $x$  in the interior of the state space. This requirement is sufficient for  $+\infty$  not to be attracting. For the process to be stationary the diffusion coefficient must grow sufficiently fast as a function of the state. In effect  $1/\sigma^2$  needs to be integrable. The high volatility in large states is enough to guarantee that the process eventually escapes from those states. Reversion to the center of the distribution is induced by this high volatility and not by the pull from the drift. An example is Zipf's with drift  $\mu = 0$ . Conley, Hansen, Luttmer, and Scheinkman (1997) give examples for models with a constant volatility elasticity.

Jones (2003) uses a stochastic volatility model of equity in which the *volatility of volatility* ensures that the volatility process is stationary. Consider a process for volatility that has a linear drift  $\mu(x) = \alpha - \kappa x$  and constant volatility elasticity:  $\sigma^2(x) \propto x^{2\gamma}$ . Jones estimates that  $\kappa$  is essentially zero. Even with a zero value of  $\kappa$  the pull measure:

$$\frac{\mu}{\sigma} - \frac{\sigma'}{2}$$

diverges to  $-\infty$  at the right boundary provided that  $\gamma$  is greater than one. Jones (2003) in fact estimates a value for  $\gamma$  that exceeds one. The pull measure also diverges at the left boundary to  $+\infty$ . The process is  $\rho$ -mixing and it has a simple spectral decomposition. Stationarity is volatility induced when  $\kappa = 0$  because relation (6.1) is satisfied provided that  $\gamma$  exceeds one. The state-dependence in the volatility (of volatility) is sufficient to pull the process to the center of its distribution even though the pull coming from the drift alone is in the wrong direction at the right boundary.

Using parameter estimates from Jones (2003), we display the first five principal components for the volatility process in Figure 1. For the principal component extraction, we use the two weighting functions described previously. For the quadratic form in function levels we weight by the stationary density implied by these parameter values. The quadratic form in the derivatives is weighted by the stationary density times the diffusion coefficient. As can be

see from Figure 1, this function converges to a constant in the right tail of the stationary distribution.

While they are nonlinear, the principal components evaluated at the underlying stochastic process each behave like a scalar autoregression with heteroskedastic innovations. As expected the higher-order principal components oscillate more as measured by zero crossings.<sup>19</sup> The higher-order principal components are less smooth as measured by the quadratic form in the derivatives. Given the weighting used in the quadratic form for the derivatives, the principal components are flat in the tails.

### 6.3 Approximating Variance Processes

Meddahi (2001) and Andersen, Bollerslev, and Meddahi (2004) use a nonlinear principal component decomposition to study models of volatility. Recall that each principal component behaves as a univariate (heteroskedastic) autoregression and the components are mutually orthogonal. These features of principal components make them attractive for forecasting conditional variances and time-averages of conditional variances. Simple formulas exist for predicting the time-average of a univariate autoregression and Andersen, Bollerslev, and Meddahi (2004) are able to apply those formulas in conjunction with a finite number of the most important principal components to obtain operational prediction formulas.

### 6.4 Random Time Changes

Models with random time changes are common in finance. There are at least two ways to motivate such models. One formulation due to Bochner (1960) and Clark (1973) posits a distinction between calendar time and economic time. The random time changes are used to alter the flow of information in a random way. Alternatively an econometrician might confront a data set with random sample times. Operator methods give a tractable way of modelling randomness of these types.

A model of random time changes requires that we specify two objects. An underlying Markov process  $\{X_t : t \geq 0\}$  that is not subject to distortions in the time scale. For our purposes, this process is modelled using a generator  $\mathcal{A}$ . In addition we introduce a process  $\{\tau_t\}$  for the time scale. This process is increasing and can be specified in continuous as  $\{\tau_t : t \geq 0\}$  and sampled at intervals normalized to have length one or specified directly in discrete time as  $\{\tau_j : j = 0, 1, 2, \dots\}$ . The discrete time process of interest is:

$$Z_j = X_{\tau_j}. \tag{6.2}$$

---

<sup>19</sup>Formally, the as expected comment comes from the Sturm-Liouville theory of second-order differential equations.

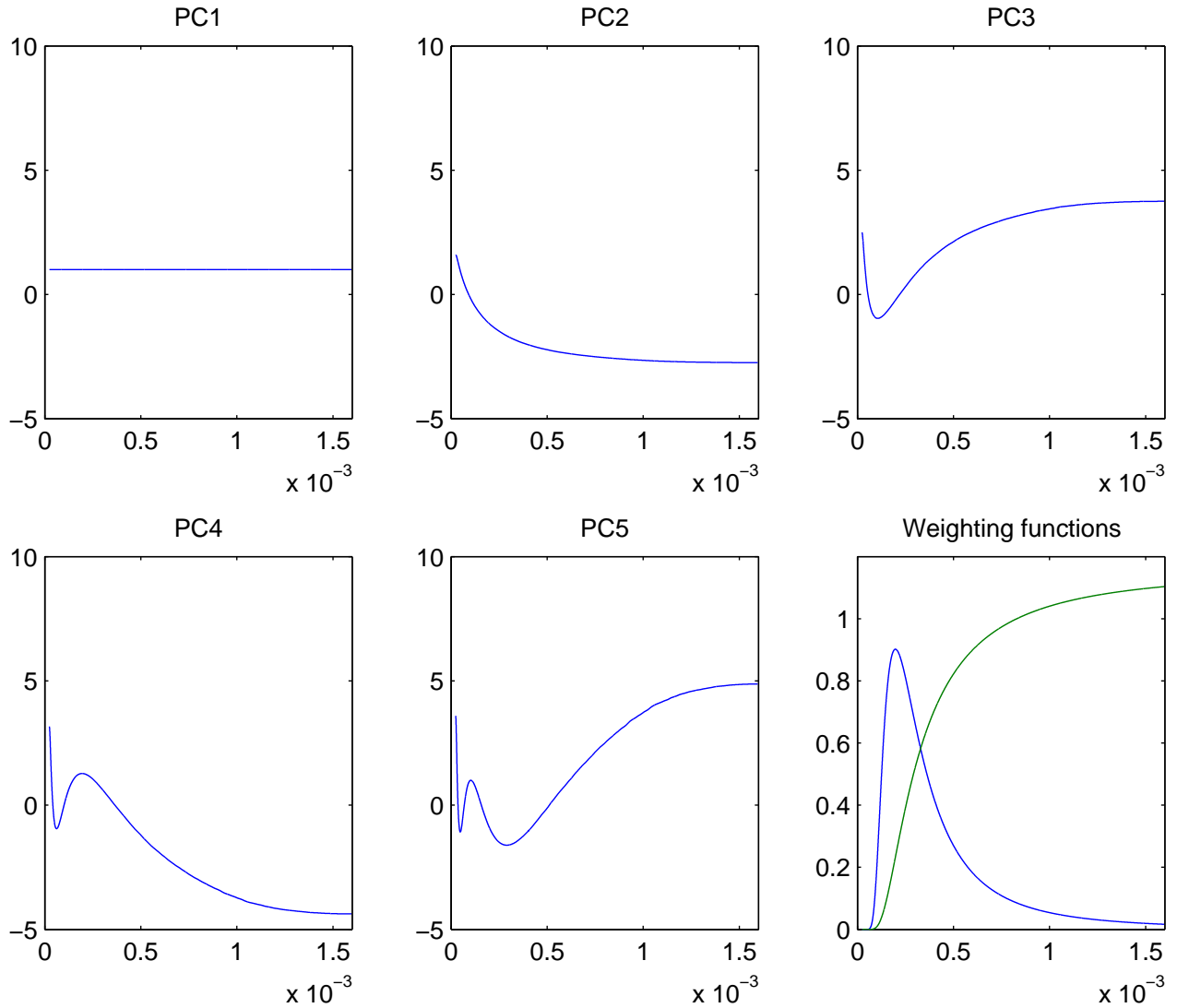


Figure 1: The first five principal components for a volatility model estimated by Jones. The weighting functions are the density and the density scaled by the diffusion coefficient. The parameter values are  $\kappa = 0$ ,  $\alpha = .58 \times 10^{-6}$ , and  $\sigma^2 = 6.1252x^{2.66}$ . Except for  $\kappa$ , the parameter values are taken from the fourth column of Table 1 in Jones. Although the posterior mean for  $\kappa$  is different from zero, it is small relative to its posterior standard deviation.

Clark (1973) refers to  $\{\tau_j\}$  as the *directing process* and the process  $\{X_t\}$  is *subordinated* to the directing process in the construction of  $\{Z_j\}$ . In what follows we consider two related constructions.

**Example 11.** *Following Ethier and Kurtz (1986), consider a process specified recursively in terms of two objects: a generator  $\mathcal{A}$  of a Markov process  $\{X_t\}$  and a nonnegative continuous function  $\zeta$  used to distort calendar time. The process that interests us satisfies the equation:*

$$Z_t = X_{\int_0^t \zeta(Z_s) ds}.$$

In this construction, we think of

$$\tau_t = \int_0^t \zeta(Z_s) ds$$

as the random distortion in the time of the process we observe. Using the time distortion we may write:

$$Z_t = X_{\tau_t},$$

which is the continuous-time counterpart to formula (6.2). By sampling this process at the integers, we obtain the discrete-time process of interest.

This construction allows for dependence between the sampling or directing process and the underlying process  $\{X_t\}$ . By construction the directing process has increments that depend on  $Z_t$ . Ethier and Kurtz (1986) show that under some additional regularity conditions, the continuous-time process  $\{Z_t\}$  is itself Markovian with generator  $\zeta\mathcal{A}$  (see Theorem 1.4 on page 309). Since the time derivative of  $\tau_t$  is  $\zeta(Z_t)$ , this scaling of the generator is to be expected. In the case of a Markov diffusion process the drift  $\mu$  and the diffusion matrix  $\nu$  are both scaled by the function  $\zeta$  of the Markov state. In the case of a Markov jump process,  $\zeta$  alters the jump frequency by scaling the intensity parameter.

**Example 12.** *Consider next a specification suggested by Duffie and Glynn (2001). Following Clark (1973), they use a Poisson specification of the directing process. In contrast to Clark (1973), Duffie and Glynn (2001) allow the Poisson intensity parameter to be state dependent. Thus consider an underlying continuous time process  $\{(X_t, \tau_t)\}$  where  $\tau_t$  is a process that jumps by one unit where the jump times are dictated by an intensity function  $\lambda(X_t)$ . The observed process has a continuous-time depiction as:*

$$Z_t = X_{\tau_t}$$

although it is the version sampled at the integers that interests us. (The process  $\{Z_t\}$  is in fact constant between jumps in  $\{\tau_t\}$ .) The jump probabilities between say  $\tau_{j+1}$  and  $\tau_j$  can depend on the realized  $X_t$  between  $\tau_j \leq t < \tau_{j+1}$  through the specification of  $\lambda$ .

Consider now the conditional distribution of  $Z_{j+1}$  conditioned on  $Z_j$  and a function  $\psi$  that satisfies:

$$\mathcal{A}\phi - \lambda\phi = -\lambda\psi$$

for some  $\phi$  in the generator's domain. For this convenient but flexible choice of  $\psi$ , under some additional regularity conditions Duffie and Glynn (2001) deduce the conditional moment restriction:<sup>20</sup>

$$E[\psi(Z_{j+1})|Z_j] = \phi(Z_j),$$

or

$$E\left[-\frac{\mathcal{A}}{\lambda}\phi(Z_{j+1}) + \phi(Z_{j+1}) - \phi(Z_j)|Z_j\right] = 0. \quad (6.3)$$

They derive this moment restriction by using the implied jump distribution conditioned on the  $X_t$  for all  $t \geq \tau_j$ .

There is in fact a close connection between these two examples. Suppose we set  $\zeta = \frac{1}{\lambda}$  in Example 11. We then know that the resulting process  $\{Z_t^*\}$  has generator  $\mathcal{A}^* \doteq \zeta\mathcal{A} = \frac{1}{\lambda}\mathcal{A}$ . In addition to this smooth time distortion, suppose we sample this process as in Example 12 according to a Poisson process, but specify the intensity to be one. Call this process  $\{Z_j\}$ . For this simple Poisson specification, the Duffie and Glynn (2001) conditional moment restriction can be computed as:

$$E\left[\int_0^\infty \exp(-t)\psi(Z_{t+\tau_j}^*)dt|Z_{\tau_j}\right] = \left[\int_0^\infty \exp(\mathcal{A}^*t)dt\right]\psi(Z_j) = (\mathcal{I} - \mathcal{A}^*)^{-1}\psi(Z_j)$$

where  $\mathcal{I}$  is the identity operator. As suggested by Duffie and Glynn (2001), we can avoid calculating the inverse by setting

$$(\mathcal{I} - \mathcal{A}^*)\phi = \psi$$

for some  $\phi$  in the generator's domain. This two-step construction also results in a process that satisfies conditional moment restriction (6.3).

Models with random time distortions present additional challenges for identification and estimation. Without observations on the directing process or sampling times, nonparametric identification of even reversible processes breaks down. If the directing process  $\{\tau_j\}$  is independent of the underlying process  $\{X_t\}$ , then the most we can hope for is identification of  $\mathcal{A}$  up to scale. It will not be possible to distinguish an original process from one that moves though time say twice as fast. Hansen and Scheinkman (1995) establish that scalar diffusions can be identified up to a free constant scale parameter without data on observation times. Identification is more challenging when the sampling or directing process is dependent on the

---

<sup>20</sup>In fact Duffie and Glynn (2001) deduce a more general class of conditional moment restrictions by allowing for test functions that depend on  $Z_j$ 's at adjacent integers.

underlying process. As we have seen in both the constructions of Ethier and Kurtz (1986) and Duffie and Glynn (2001), the generator of the original process is scaled by a scalar function of the underlying Markov state in the characterization of the generator for the subordinated process or in the construction of the conditional moment restrictions. Thus without data on the process  $\{\tau_j\}$  we are left not being able to distinguish  $\mathcal{A}$  from  $\zeta^* \mathcal{A}$  for some positive function  $\zeta^*$  of the Markov state. The free scale factor is a function not a constant. Finite-dimensional parameterizations will simplify or in some cases even solve this identification problem.

Consider the case in which  $\{\tau_j\}$  is directly interpreted as a set of sample times and not some unobserved distortion in the time scale. These sampling times provide important identifying information about the possibly dependent sampling scheme and about the underlying process  $\{X_t\}$ . We expect direct or indirect (through say trading volume) data on the directing process to be useful in inferring the underlying process. Our discussion in the next subsection studies this question.

## 6.5 Value of Sampling Time Information

Ait-Sahalia and Mykland (2003) quantify the value of sample time information for the following specification. They consider the estimation of the parameters  $\theta = (\kappa, \sigma)$  of

$$dX_t = \mu(X_t; \kappa)dt + \sigma dW_t.$$

Sampling intervals  $\{\Delta_j : j = 1, 2, \dots\}$  are random where  $\Delta_j = \tau_j - \tau_{j-1}$  is drawn conditionally upon  $X_{\tau_{j-1}}$  from a known distribution. Formally this model differs from those described in the previous subsection because movements in  $X_t$ , for  $\tau_{j-1} < t < \tau_j$ , are not allowed to influence the  $\tau_j$ . Nevertheless, by letting  $\Delta_j$  be drawn conditionally on  $X_{\tau_{j-1}}$ , one can still capture effects such as an increase in trading activity following a large price movement say at  $\tau_{j-1}$ .

To study the value of observing  $\{\Delta_j : j = 1, 2, \dots\}$ , Ait-Sahalia and Mykland (2003) study four possible estimators  $\theta$  designed so that each one of them is subject to a specific subset of the different effects they wish to measure. The estimators are:

- FIMLC: Full Information Maximum Likelihood, using a continuous record of  $X_t$ ;
- FIMLD: Full Information Maximum Likelihood, using the bivariate observations  $(X_{\tau_j}, \Delta_j)$ ;
- IOML: Partial information maximum likelihood estimator using only the state observations  $X_{\tau_j}$ , with the sampling intervals *integrated out*;
- PFML: Pseudo maximum likelihood estimator *pretending* that the sampling intervals are *fixed* at  $\Delta_j = \bar{\Delta}$ .

FIMLC, FIMLD and IOML are asymptotically efficient given their respective data sets. Under stationarity and ergodicity, they will be asymptotically normal with asymptotic covariance matrix  $\Omega_\theta$ . FIMLC uses the most data and IOML the least. The extent to which FIMLD is less efficient than FIMLC is the *cost of discreteness*, and the extent to which IOML is less efficient than FIMLD is the *cost of ignoring sampling times*. PFML will be asymptotically biased with population limit  $\bar{\theta}$ . The extent to which PFML performs worse than FIMLD is the *cost of ignoring randomness*.

Ait-Sahalia and Mykland (2003) derive Taylor expansions of the asymptotic variance and bias of these estimators. A random variable from the common distribution of the sampling intervals is  $\Delta = \varepsilon\Delta_0$ , where  $\varepsilon$  is deterministic and  $\Delta_0$  has a given finite distribution conditional on  $X_0$ . They compute Taylor expansions in  $\varepsilon$  of the expectations of interest, around  $\varepsilon = 0$  (the limiting case were the full continuous-time sample path is observable), leading to results of the type:

$$\Omega_\theta = \Omega_\theta^{(0)} + \varepsilon \Omega_\theta^{(1)} + \varepsilon^2 \Omega_\theta^{(2)} + O(\varepsilon^3) \quad (6.4)$$

$$\bar{\theta} - \theta_0 = \varepsilon b_\theta^{(1)} + \varepsilon^2 b_\theta^{(2)} + O(\varepsilon^3) \quad (6.5)$$

where the higher order terms in  $\varepsilon$  correct the leading term for the discreteness of the sampling. Differences between estimation methods and data use show up in the functions  $\Omega_\theta^{(i)}$  and  $b_\theta^{(i)}$ ,  $i = 0, 1, \dots$

The two equations (6.4)-(6.5) are then used to analyze the effects of a given sampling scheme on parameter estimation. The cost of discreteness is measured by the coefficient at the first order  $i$  in  $\varepsilon$  for which the FIMLD variance differs from its continuous-time limit  $\Omega_\theta^{(0)}$ . It is also the error that one would make if one were to use continuous-time asymptotics ( $\Omega_\theta^{(0)}$ ) instead of the full  $\Omega_\theta$  when the data are in fact discretely sampled.

The cost of ignoring sampling times is quantified by examining the first order  $i$  in  $\varepsilon$  at which the coefficient  $\Omega_\theta^{(i)}$  for IOML differs from the corresponding coefficient  $\Omega_\theta^{(i)}$  for FIMLD. The cost is measured by how much bigger the IOML coefficient at that order is than the FIML coefficient. For this example, the cost of randomness is at least as great, and often substantially greater than the cost of discreteness.

Since the PFML estimator is asymptotically biased its asymptotic mean-square error is dominated by the square of the bias. Its performance under an asymptotic mean-square error loss function will always be worse than an estimator that is asymptotically unbiased. Expansion (6.5) can be used to quantify the squared bias.

The main conclusion is that the loss from not observing, or not using, the sampling intervals, will be at least as great, and often substantially greater, than the loss due to the fact that the data are discrete rather than continuous. Introducing unknown parameters in the sampling distribution for  $\Delta_j$  will alter the quantitative comparison, but we know from

the related results in the previous subsection that full identification of the diffusion can fail without some knowledge of the sampling distribution.

While correcting for the latter effect has been the main focus of the literature in recent years, these results suggest that empirical researchers using randomly spaced data should pay as much attention, if not more, to sampling randomness as they do to sampling discreteness.

## 6.6 Measuring the Diffusion Coefficient in the Presence of Jumps

Suppose now that jumps are in fact present, in addition to the usual Brownian noise, as in

$$dX_t = \mu dt + \sigma dW_t + dY_t,$$

where  $Y_t$  is a pure jump Lévy process with jump measure  $\nu$  and independent of the Brownian motion  $W_t$ . When we first introduced jump process in Section 2, we noted that the jump distribution could be generalized to allow for an infinite number of jumps in a finite interval. By restricting  $\{Y_t\}$  to be a pure Lévy process, we eliminate state dependence but include jump measures that are not finite. When  $\nu$  is a finite measure the jump process is referred to as a compound Poisson process. Other Lévy processes are such that  $\nu([-\varepsilon, +\varepsilon]) = \infty$  for any  $\varepsilon > 0$ , so that the process exhibits an infinite number of small jumps in any finite time interval. Typical examples are members of the class of symmetric stable processes of index  $0 < \alpha < 2$  and rate  $\xi > 0$ , for which  $\nu(dy) = \alpha \xi^\alpha dy / |y|^{1+\alpha}$ . The Cauchy process corresponds to  $\alpha = 1$ , while the limit  $\alpha \rightarrow 2$  (from below) produces a Gaussian distribution. Following Aït-Sahalia (2003), we assess the effect of jumps on the estimation of the Brownian variance parameter  $\sigma^2$ .

When the Lévy measure is finite, the tiny jumps ought to be harder to distinguish from Brownian noise. Surprisingly, using maximum likelihood, it is possible to identify  $\sigma^2$  with the same degree of precision as if there were no jumps. Specifically, when the Brownian motion is contaminated by jumps, with a known measure, the asymptotic variance AVAR of the maximum likelihood estimator ML for the diffusion coefficient estimator satisfies

$$\text{AVAR}_{\text{ML}}(\sigma^2) = 2\sigma^4\Delta + o(\Delta) \tag{6.6}$$

so that in the limit when the sample interval shrinks to zero ( $\Delta \rightarrow 0$ ), the ML estimator of  $\sigma^2$  has the same asymptotic distribution as if no jumps were present. This result holds not only for the specific examples considered in Aït-Sahalia (2003) but for all Lévy processes which stay at a finite distance from the limiting case  $\alpha = 2$  (see Aït-Sahalia and Jacod (2003)).

This result also states that the presence of the jumps imposes no cost on our ability to estimate  $\sigma^2$ . From (6.6), the leading term in the asymptotic variance expansion is the asymptotic variance that applies when jumps are absent. In contrast, suppose we contaminated

the Brownian motion with another independent Brownian motion with known variance  $\varsigma^2$ . In that case, we could still estimate  $\sigma^2$ , but the asymptotic variance of the MLE would be  $2(\sigma^2 + \varsigma)^2 \Delta$ .

Aït-Sahalia (2003) also studies the ability of method-of-moments to reproduce the efficiency of ML, considering in particular absolute moments of order  $r$  and shows that the optimal choice of moment functions involves non integer values of  $r$  which are less than one.

## 6.7 Strong Dependence

Linear characterizations of time series typically define long memory in terms of the behavior of the spectral density function (the Fourier transform of the autocovariance function). When the spectral density diverges to infinity at zero, there is strong linear dependence. The degree of fractional integration is defined using the rate at which this spectral density diverges. As we have seen, stationary Markov processes can be strongly dependent as characterized by the behavior of the implied mixing coefficients.

The spectral density function at frequency zero is typically the variance used in a central limit approximation. From Bhattacharya (1982) and Hansen and Scheinkman (1995) and formula (3.5), we know that the asymptotic variance for the central limit approximation for  $\frac{1}{\sqrt{T}} \int_0^T \mathcal{A}\phi(x_t)$  is:

$$\begin{aligned} 2f_2(\phi, \phi) &= -2 \int \phi(\mathcal{A}\phi) dQ \\ &= \int (\nabla\phi)^* \nu(\nabla\phi) q \end{aligned}$$

where the second right-hand side expression is the formula for diffusion processes. The formula for jump processes is different. Thus the long-run variance for the process  $\{\mathcal{A}\phi(x_t)\}$  is given by the form  $2f_2$  applied to the test function  $\phi$ . This long-run variance is also the spectral density at frequency zero.

This long-run variance is not always finite, however. Using this long-run variance, we may define weak dependence as:

$$\sup_{\phi \in D(\mathcal{A}), \int \mathcal{A}(\phi)^2 dQ = 1} f_2(\phi, \phi) < \infty.$$

This is in effect the inverse counterpart to (5.10), and is equivalent to the restriction that the  $\rho$ -mixing coefficients have exponential decay. This criterion also suggests how we might construct strongly dependent diffusion processes with a divergent spectral density. Find a pair  $(\nu, Q)$  and a test function  $\phi$  such that for

$$\psi \doteq \mu \cdot \frac{\partial\phi}{\partial x} + \frac{1}{2} \text{trace} \left( \nu \frac{\partial^2\phi}{\partial x \partial x'} \right)$$

- $\int |\psi|^2 q < \infty$ ;
- $\int \psi q = 0$ ;
- $\int \left(\frac{\partial \phi}{\partial x}\right)^* \nu \left(\frac{\partial \phi}{\partial x}\right) = \infty$

Such a process gives an alternative way to produce long range dependence to the self similar fractional Brownian motion model of Mandelbrot and Ness (1968). While these diffusions are not self-similar, they have the mathematical advantage of being semimartingales.

We illustrate a family of scalar diffusion models that are strongly dependent. It is often argued that strong dependence is a feature of volatility models. One important source of evidence for strong dependence is a spectral density matrix that diverges at frequency zero. We now display one construction of a nonlinear diffusion model that is strongly dependent. This example is taken from Chen, Hansen, and Carrasco (2003).

Consider a scalar process with a zero mean and a diffusion coefficient  $\sigma^2(x) = (1+x^2)^\gamma$  for  $1/2 < \gamma < 1$ . The candidate stationary density is proportional to  $1/\sigma^2$ . In fact this process is stationary, but its  $\rho$ -mixing coefficients are unity. In particular, the pull measure

$$\frac{\mu}{\sigma} - \frac{\sigma'}{2}$$

is zero at both boundaries.

Form a new process by taking a time invariant transformation of the original process. That is, let

$$\psi = \frac{\sigma^2}{2} \phi''$$

where  $\phi$  is such that  $\phi'(x) = (1+x^2)^{-\eta/2}$ . Restrict  $\eta$  to satisfy:  $\gamma - 1/2 \leq \eta \leq 1/2$ . Then  $\psi$  has mean zero and finite variance when integrated against the stationary density.<sup>21</sup> Its long run variance, however, is infinite. Notice that

$$\int \sigma^2 (\phi')^2 q = \infty$$

because  $\eta \leq 1/2$ . The divergence spectral density function near frequency zero is illustrated by Figure 2. The rate of divergence of this spectral density function at frequency zero is shown in Figure 3 using logarithmic scaling.

## 6.8 Pricing

As an alternative application, Darolles and Laurent (2000) use a principal component decomposition for scalar diffusions to approximate asset payoffs and prices under a risk neutral

---

<sup>21</sup>The function  $\phi$  will not typically be in the  $L^2(Q)$  domain of the generator.

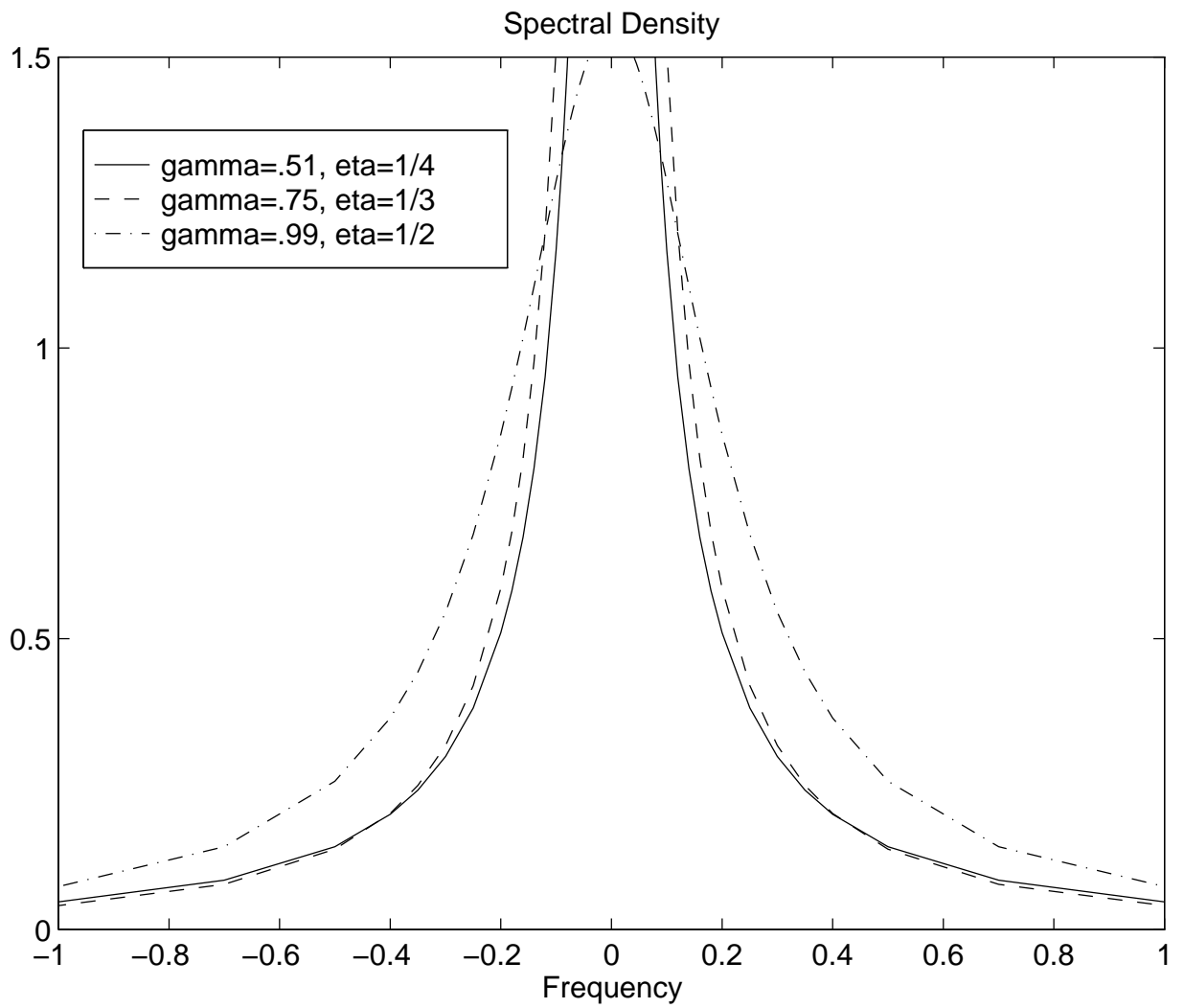


Figure 2: Spectral density functions for different pairs  $(\gamma, \eta)$ . Spectral densities are re-scaled to integrate to one.

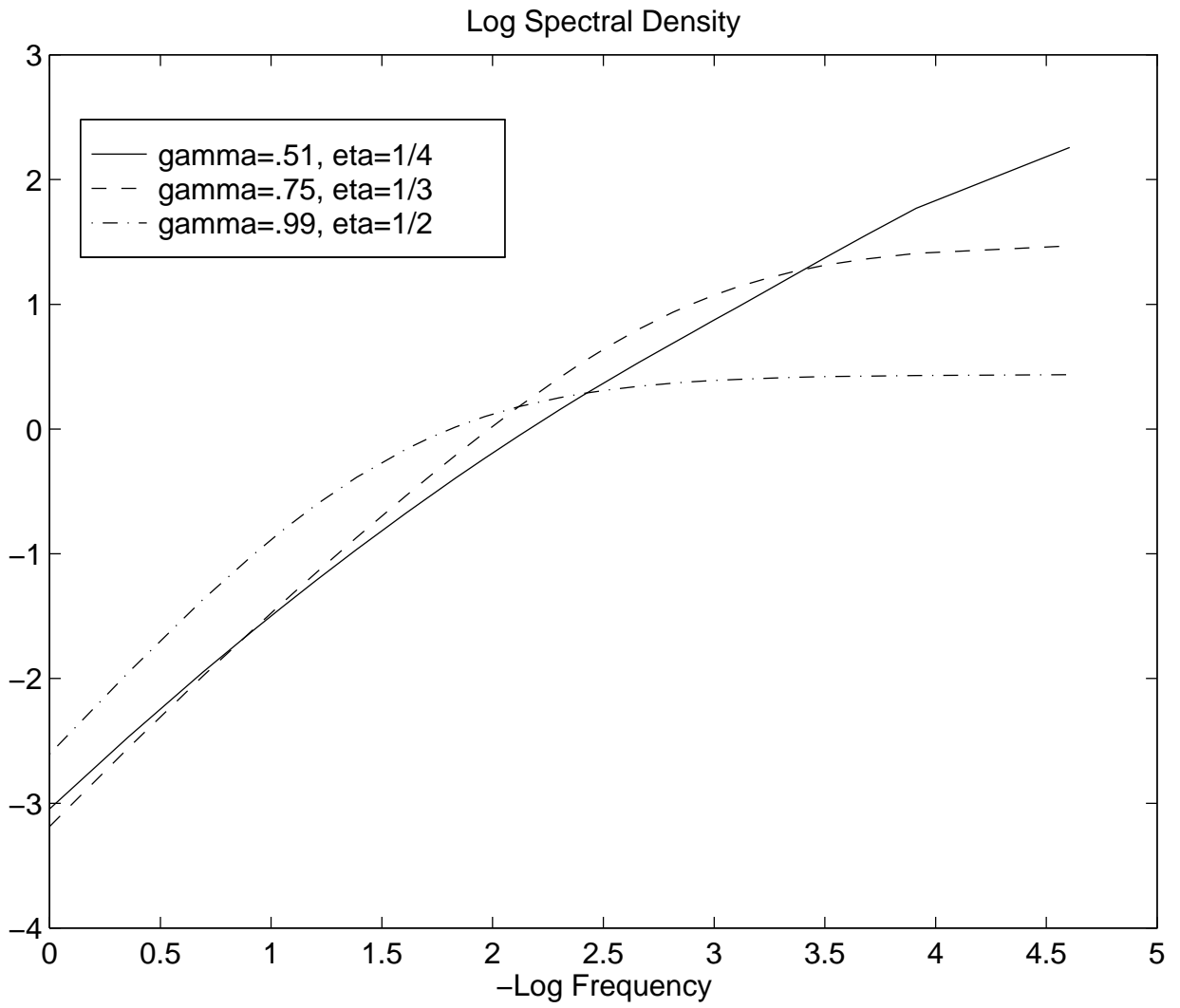


Figure 3: Spectral density functions for different pairs  $(\gamma, \eta)$  plotted on a  $\ln - \ln$  scale. Spectral densities are re-scaled to integrate to one.

probability distribution. Limiting attention to a finite number of terms of the principal component decomposition simplifies the calculation of prices.

More generally, Hansen and Scheinkman (2003) develop a semigroup theory for Markov pricing. In their framework, a semigroup is a family of operators that assigns prices today to payoffs that are functions of the Markov state in the future. Like semigroups for Markov processes, the Markov pricing semigroup has a generator. In contrast to Darolles and Laurent (2000), it explicitly incorporates jump components and interest rates. As an application Hansen and Scheinkman (2003) study the implications of the existence of an eigenfunction for the semigroup that dominates pricing in the long run.

## References

- Aït-Sahalia, Y. (1996a). Nonparametric pricing of interest rate derivative securities. *Econometrica* 64, 527–560.
- Aït-Sahalia, Y. (1996b). Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9, 385–426.
- Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* 54, 1361–1395.
- Aït-Sahalia, Y. (2001). Closed-form likelihood expansions for multivariate diffusions. Technical report, Princeton University.
- Aït-Sahalia, Y. (2002a). Empirical option pricing and the markov property. Technical report, Princeton University.
- Aït-Sahalia, Y. (2002b). Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. *Econometrica* 70, 223–262.
- Aït-Sahalia, Y. (2002c). Telling from discrete data whether the underlying continuous-time model is a diffusion. *Journal of Finance* 57, 2075–2112.
- Aït-Sahalia, Y. (2003). Disentangling diffusion from jumps. *Journal of Financial Economics*. forthcoming.
- Aït-Sahalia, Y. and J. Jacod (2003). Which pure jump lévy processes can be distinguished from brownian motion? Technical report, Princeton University.
- Aït-Sahalia, Y. and P. A. Mykland (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* 71, 483–549.
- Andersen, T. G., T. Bollerslev, and N. Meddahi (2004). Analytic evaluation of volatility forecasts. *International Economic Review* forthcoming.
- Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676–687.
- Auerbach, F. (1913). Das gbesetz der bevölkerungskonzentration. *Petermanns Geographische Mitteilungen* 59, 74–76.
- Bandi, F. and P. C. B. Phillips (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241–283.
- Bandi, F. M. (2002). Short-term interest rate dynamics: A spatial approach. *Journal of Financial Economics* 65, 73–110.

- Banon, G. (1978). Nonparametric identification for diffusion processes. *SIAM Journal of Control and Optimization* 16, 380–395.
- Beurling, A. and J. Deny (1958). Espaces de dirichlet i, le cas elementaire. *Acta Math.* 99, 203–224.
- Bhattacharya, R. N. (1982). On the functional central limit theorem and the law of the iterated logarithm for markov processes. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete* 60, 185–201.
- Bierens, H. J. (1990). A consistent conditional moment test of functional form. 58.
- Billingsley, P. (1961). *Statistical Inferences for Markov Processes*. Chicago: University of Chicago Press.
- Bochner, S. (1960). *Harmonic Analysis and the Theory of Probability*. Berkley: University of California Press.
- Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 48, 1209–1227.
- Chen, X., L. P. Hansen, and M. Carrasco (2003). Nonlinearity and temporal dependence. manuscript.
- Chen, X., L. P. Hansen, and J. A. Scheinkman (2000). Principal components and the long run. manuscript.
- Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41, 135–155.
- Cobb, L., P. Koppstein, and N. Y. Chan (1983). Estimation and moment recursions for multimodal distributions of the exponential family. *Journal of the Americal Statistical Association* 78, 124–130.
- Conley, T., L. P. Hansen, and W. F. Liu (1999). Bootstrapping the long run. *Macroeconomic Dynamics* 1, 279–311.
- Conley, T. G., L. P. Hansen, E. G. J. Luttmer, and J. A. Scheinkman (1997). Short-term interest rates as subordinated diffusions. *Review of Financial Studies* 10, 525–577.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica* 53, 385–408.
- Darolles, S., J. P. Florens, and C. Gourieroux (2000). Kernel based nonlinear canonical analysis and time reversibility. manuscript.

- Darolles, S., J. P. Florens, and C. Gourieroux (2001). Factor arma representation of a markov process. *Economic Letters* 71, 165–171.
- Darolles, S. and J. P. Laurent (2000). Approximating payoffs and pricing formulas. *Journal of Economic Dynamics and Control* 24, 1721–1746.
- Dauxois, J. and G. M. Nkiet (1998). Nonlinear canonical analysis and independence tests. *Annals of Statistics* 26, 1254–1278.
- Demoura, S. G. (1998). The nonparametric estimation of the expected value operator. manuscript.
- Duffie, D. and P. Glynn (2001). Estimation of continuous-time markov processes sampled at random time intervals. Stanford University.
- Eaton, J. and Z. Eckstein (1997). Cities and growth: Theory and evidence from france and japan. *Regional Science and Urban Economics* 27, 443–474.
- Ethier, S. N. and T. G. Kurtz (1986). *Markov Processes: Characterization and Convergence*. New York: John Wiley and Sons.
- Fang, J. and C. Zhang (2003). A re-examination of stanton’s diffusion estimations with applications to financial market validation. *Journal of the American Statistical Association* 98, 118–134.
- Fisher, R. A. (1921). On the mathematical foundation of theoretical statistics. *Phil. Trans. Roy. Soc.* 222, 309.
- Florens, J. P., E. Renault, and N. Touzi (1998). Testing for embeddability by stationary reversible continuous-time markov processes. *Econometric Theory* 14, 744–769.
- Florens-Zmirou, D. (1984). Theorem de limite central pour une diffusion et pour sa discretisee. *C. R. Acad. Sc. Paris, Series I t 299 n 9*, 995–998.
- Frydman, H. and B. Singer (1979). Total positivity and the embedding problem for markov chains. *Math. Proc. Camb. Phil. Soc* 86, 339–344.
- Fukushima, M. (1971). Dirichlet spaces and strong markov processes. *Transactions of the American Mathematics Society* 162, 185–224.
- Fukushima, M., Y. Oshima, and M. Takeda (1994). *Dirichlet Forms and Symmetric Markov Processes*. Walter de Gruyter.
- Gabaix, Z. (1999). Zipf’s law for cities: An explanation. *Quarterly Journal of Economics* 111, 738–767.
- Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*. Boston: Academic Press.

- 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 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.
- Hansen, L. P. and J. Scheinkman (2003). Semigroup pricing. manuscript.
- Hansen, L. P., J. Scheinkman, and N. Touzi (1998). Spectral methods for identifying scalar diffusion processes. *Journal of Econometrics* 86, 1–32.
- Hansen, L. P. and J. A. Scheinkman (1995). Back to the future: Generating moment implications for continuous-time markov processes. *Econometrica* 63, 767–804.
- Hasminskii, R. Z. (1960). Ergodic properties of recurrent diffusion processes and stabilization of the solution of the cauchy problem for parabolic equations. *Theory, Probability and Applications* 5, 179–196.
- Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–887.
- Jensen, B. and R. Poulsen (2002). Transition densities of diffusion processes: Numerical comparison of approximation techniques. *Journal of Derivatives* 9, 18–32.
- Johannes, M. (2004). The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance* forthcoming.
- Jones, C. S. (2003). The dynamics of stochastic volatility: Evidence from the underlying and options market. *Journal of Econometrics* 116, 181–224.
- Karlin, S. and J. McGregor (1959a). Coincidence probabilities. *Pacific Journal of Mathematics* 9, 1141–1164.
- Karlin, S. and J. McGregor (1959b). Coincidence properties of birth and death processes. *Pacific Journal of Mathematics* 9, 1109–1140.
- Kessler, M. (1997). On the estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics* 24, 211–229.
- Kessler, M. and M. Sorensen (1999). Estimation equations based on eigenfunctions for discretely observed diffusion processes. *Bernoulli* 5, 299–314.
- Lo, A. (1988). Maximum likelihood estimation of generalized ito processes with discretely sampled data. *Journal of Econometric Theory* 4, 231–247.

- Ma, Z. M. and M. Rockner (1991). *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms*. Berlin: Springer-Verlag.
- Mandelbrot, B. B. and J. W. V. Ness (1968). Fractional brownian motions, fractional noises and applications. *SIAM Review* 10, 422–437.
- Meddahi, N. (2001). An eigenfunction approach for modeling volatility modeling. CIRANO working paper 2001s-70.
- Nelson, D. B. (1990). Arch models as diffusion approximations. *Journal of Econometrics* 45, 7–38.
- Pearson, K. (1894). On the dissection of assymmetric frequency curves. *Philosophical Transactions of the Royal Society of London: Series A* 185, 71–110.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2, 559–572.
- Pedersen, A. R. (1995). A new approach to maximum-likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics* 22, 55–71.
- Phillips, P. C. B. (1973). The problem of identification in finite parameter continuous time models. *Journal of Econometrics* 1, 351–362.
- Pritsker, M. (1998). Nonparametric density estimation and tests of continuous-time interest rate models. *Review of Financial Studies* 11, 449–489.
- Rapoport, A. (1978). Rank-size relations. In W. H. Kruskal and J. M. Tanur (Eds.), *International Encyclopedia of Statistics*, pp. 847–854. The Free Press.
- Revuz, D. and M. Yor (1991). *Continuous Martingales and Brownian Motion* (second ed.). Berlin: Springer-Verlag.
- Rudin, W. (1973). *Functional Analysis*. New York: McGraw-Hill, Inc.
- Salinelli, E. (1998). Nonlinear principal components i: Absolutely continuous variables. *Annals of Statistics* 26, 596–616.
- Schaumburg, E. (2003). Estimation of time reversible markov processes. manuscript, Northwestern University.
- Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52, 1973–2002.
- Stroock, D. W. and S. R. S. Varadhan (1979). *Multidimensional Diffusion Processes*. Springer.

- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.
- Veretennikov, A. Y. (1997). On polynomial bounds for stochastic differential equations. *Stochastic Processes and their Applications* 70, 115–127.
- White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29, 1188–1197.
- Wong, E. (1964). The construction of a class of stationary markoff processes. In R. Bellman (Ed.), *Sixteenth Symposium in Applied Mathematics - Stochastic Processes in Mathematical Physics and Engineering*, pp. 264 – 276. Providence, RI: American Mathematical Society.