Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals Forthcoming: *Econometrica*

Darrell Duffie and Peter Glynn
Stanford University

Current Version: August 19, 2004

Abstract

We introduce a family of generalized-method-of-moments estimators of the parameters of a continuous-time Markov process observed at random time intervals. The results include strong consistency, asymptotic normality, and a characterization of standard errors. Sampling is at an arrival intensity that is allowed to depend on the underlying Markov process and on the parameter vector to be estimated. We focus on financial applications, including tick-based sampling, allowing for jump diffusions, regime-switching diffusions, and reflected

*Duffie is at the Graduate School of Business, Stanford University, Stanford CA 94305-5015 (email: duffie@stanford.edu). Glynn is at the Management Science and Engineering Department, Stanford University, Stanford CA 94305 (email: glynn@leland.stanford.edu).
We are especially grateful to Qiang Dai for extensive research assistance and discussions, to Jun Liu and Mark Garmaise for additional research assistance, to Tom Kurtz for helpful suggestions, and to Yacine Aït-Sahalia for early formative discussions that contributed to the choice of this approach. We are also grateful for comments from several anonymous referees, Lars Hansen, Joel Hasbrouck, Mathias Kessler, Alain Monfort, and Michael Sørensen; and from participants of conferences or seminars at INRIA, Sophia-Antipolis, France; Department of Management, Odense University, Denmark; The Department of Mathematics, Humboldt University, Berlin; the Department of Mathematics, Stanford University; The Department of Statistics, Columbia University; The Institute of Advanced Studies, Princeton University; The Mathematical Sciences Research Institute, Berkeley; The Graduate School of Business, University of Chicago; and CREST-INSEE, Paris. We acknowledge hospitality offered to the first author by Mishkenot Sha'ananim, Jerusalem.

diffusions. **Keywords:** Method of moments, parameter estimation, Markov process, continuous-time.

1 Introduction

We introduce a family of generalized-method-of-moments (GMM) estimators for continuous-time Markov processes observed at random time intervals. The results, in parallel with GMM estimation in a discrete-time setting, include strong consistency, asymptotic normality, and a characterization of standard errors. We allow for a range of sampling schemes. A special case is sampling at the event epochs of a Poisson process. More generally, we allow for the arrival of observations at an intensity that varies with the underlying Markov process. The unknown statistical parameters may determine the arrival intensity that governs the sampling times. Our approach is motivated by *(i)* the fact that certain financial data, particularly intra-day, are sampled at random times, *(ii)* by the fact that it offers structural econometric identification, and *(iii)* by its computational advantages in calculating moment conditions. The approach does not apply if the sampling times are deterministic unless the model assumptions apply after a random time change induced, for example, by variation in the arrival rate of market information.

We are particularly interested in applications to financial time series, including tick-based sampling, allowing for jump diffusions, regime-switching diffusions, and reflected diffusions. A companion paper, Dai, Duffie, and Glynn [1997], gives a fully worked example, with explicit solutions for the asymptotic variances associated with "typical" moment conditions suggested by our approach, including maximum-likelihood estimation, for Ornstein-Uhlenbeck processes sampled at Poisson times.

The goal is to estimate the parameters governing the probabilistic behavior of a time-homogeneous continuous-time Markov process X. We defer to Section 3 a more careful and complete description of the problem setting. For informal purposes, we define X in terms of its state space S (for example, a subset of $\mathbb{R}^k$ for some $k \geq 1$) and its infinitesimal generator $\mathcal{A}$. At each appropriately well behaved $G : S \to \mathbb{R}$, the function $\mathcal{A}G : S \to \mathbb{R}$ is defined by

$$\mathcal{A}G(x) = \frac{d}{dt} E_x[G(X(t))]\Big|_{t=0+}, \qquad x \in S, \tag{1.1}$$

where E_x denotes expectation associated with a given initial condition x for

X. The transition probabilities of X are determined by its generator $\mathcal{A}$. In our setting, the generator $\mathcal{A}$ is unknown, and assumed to be one of a family $\{\mathcal{A}_\theta : \theta \in \Theta\}$ of generators that is one-to-one with a set $\Theta \subset \mathbb{R}^d$ of parameters. (Some of our results apply to non-parametric settings.)

In many financial and other applications, the Markov process X cannot be observed continuously. We propose an estimator of the unknown "true" parameter $\theta^* \in \Theta$ based on observation of X at random times $T_1, T_2, T_3, \ldots$. We assume, for later discussion and generalization, that $T_i = \inf\{t : N(t) = i\}$, the i-th event time of a doubly-stochastic counting process (Brémaud [1981]) N driven by X, for some state-dependent arrival intensity function $\lambda : S \to (0, \infty)$. That is, conditional on $\{X(t) : t \geq 0\}$, the counting process N is distributed as a Poisson process with time-varying intensity $\{\lambda(X(t)) : t \geq 0\}$. The idea is roughly that, conditional on $\{(X(u), N(u)) : 0 \leq u \leq t\}$, the probability of an observation between t and $t+\Delta t$ for small Δt is approximately $\lambda(X(t))\Delta t$. We allow for the possibility that the intensity function $\lambda(\,\cdot\,)$ is not "known." That is, we suppose that $\lambda(x) = \lambda(\theta^*, x)$ for all x, where $\lambda : \Theta \times S \to (0, \infty)$. For example, N could be a Poisson process of unknown intensity.

In some cases, we interpret the sampling times $T_1, T_2, \ldots$ as the times at which observations are generated by the underlying economic process, for example, the times of quotes or reported trades of a financial security whose prices are determined by X. In other cases, an econometrician might simulate times at which to sample X. An extension to the case of subordination of X to a "market-time" process is discussed in Section 7.

Our objective is to estimate the "true" generator $\mathcal{A} = \mathcal{A}_{\theta^*}$ based on observation of $Z_0, Z_1, Z_2, \ldots, Z_n$, where $Z_i = X(T_i)$, taking $T_0 = 0$. In some cases, we also assume observation of the arrival times $T_1, T_2, \ldots, T_n$. The basic idea of our approach is as follows. We choose some "test function" of the form $g : \Theta \times S \times S \to \mathbb{R}$. We suppose that g is measurable and, for each state x in S and each θ in Θ, that $\mathcal{A}_\theta g^{(\theta,x)}(\,\cdot\,) = \mathcal{A}_\theta g(\theta, x, \,\cdot\,)$ is well defined. We show, under technical regularity, that

$$E\left[g(\theta^*, Z_i, Z_{i+1}) - \frac{\mathcal{A} g^{(\theta^*, Z_i)}(Z_{i+1})}{\lambda(\theta^*, Z_{i+1})} - g(\theta^*, Z_i, Z_i) \;\middle|\; Z_i \right] = 0. \tag{1.2}$$

We therefore construct the sample-moment analogue

$$\Gamma_n(\theta) = \frac{1}{n} \sum_{i=0}^{n-1} \gamma(\theta, Z_i, Z_{i+1}),$$

where

$$\gamma(\theta, x, y) = g(\theta, x, y) - \frac{\mathcal{A}_\theta g^{(\theta,x)}(y)}{\lambda(\theta, y)} - g(\theta, x, x).$$

Assuming the positive recurrence of X and technical conditions provided below, the sample moment $\Gamma_n(\theta)$ converges with n to its population counterpart almost surely, uniformly in the parameter θ. From (1.2), in particular, $\Gamma_n(\theta^*) \to 0$ almost surely. This suggests stacking together some number $m \geq d = \dim(\Theta)$ of such moment conditions, so that Γ_n takes values in $\mathbb{R}^m$. Then, under identification conditions, the parameter θ_n that minimizes $\|\Gamma_n(\theta)\|$ converges to θ^*. (Here, we could take $\|\cdot\|$ to be the usual Euclidean norm, but for reasons of efficiency some other norm that may depend on the data is often chosen.) This is standard for GMM estimation (Hansen [1982]). Efficiency and identification are influenced by the number and choice of different such test functions g, by the use of instrumental variables, and by suitable "weighting" of different moment conditions, as explained in Appendix B. In principle, maximum-likelihood estimation (MLE) is included as a special case, but may be computationally intractable. In such cases, it may be reasonable to use moment conditions based on the first-order conditions for MLE associated with a related model for which MLE is tractable. We do not, however, have any theory showing that this would achieve "near efficiency" if the related model is "sufficiently near."

Section 2 gives some motivation for our approach. Section 3 provides a more careful problem statement and defines our class of estimators. Section 4 shows that there exist test functions that identify the "true" parameter. Section 5 summarizes the use of our approach to estimation. Section 6 gives some illustrative examples. Section 7 discusses some alternative formulations for random sampling times, including subordination to "market time," particularly with an eye toward financial time series. Appendices A through E contain proofs and supporting technical results.

2 Motivation and Alternative Approaches

This section is an informal introduction to our approach, and offers some comparison with other approaches.

We are partly motivated by the computational difficulties associated with traditional moment conditions that are based on deterministic sampling times.[1]

[1]See, for example, Bibby and Sorensen [1995], Bibby and Sorensen [1997], Broze, Scail-

In this previous work, the general idea is to obtain moment conditions for inference based on some function $f : \Theta \times S \times S \to \mathbb{R}^m$, for some $m \geq d$, such that $h : \Theta \times S \to \mathbb{R}^m$ is well defined by

$$h(\theta, x) = E_x^\theta[f(\theta, x, Z_1)], \tag{2.1}$$

where E_x^θ denotes expectation under the assumption that X has infinitesimal generator $\mathcal{A}_\theta$ and initial condition x. One then considers an estimator θ_n that minimizes a sample-moment criterion such as

$$\left\| \sum_{i=0}^{n-1} f(\theta, Z_i, Z_{i+1}) - h(\theta, Z_i) \right\|. \tag{2.2}$$

A difficulty with this traditional approach is that $h(\theta, \cdot\,)$ is typically difficult to compute from (2.1), whether or not T_1 is a random time. We now consider this computational issue.

Under the model with parameter θ, the density $p(\,\cdot\,)$ of T_1 conditional on $\{X(t) : t \geq 0\}$ is that of a Poisson arrival time with time-varying intensity $\lambda(\theta, X(t))$ (Brémaud [1981]), so that

$$p(t) = \exp\left(\int_0^t -\lambda(\theta, X(s))\, ds\right) \lambda(\theta, X(t)), \quad t \geq 0.$$

The definition (2.1) of h and the law of iterated expectations then imply that

$$\begin{aligned} h(\theta, x) &= E_x^\theta\left[E_x^\theta\left(f(\theta, x, X(T_1)) \;\middle|\; \{X(t) : t \geq 0\}\right)\right] \qquad (2.3)\\ &= E_x^\theta\left[\int_0^\infty p(t) f(\theta, x, X(t))\, dt\right] \\ &= E_x^\theta\left[\int_0^\infty \exp\left(\int_0^t -\lambda(\theta, X(s))\, ds\right) \lambda(\theta, X(t)) f(\theta, x, X(t))\, dt\right]. \end{aligned}$$

Provided technical conditions are satisfied, it follows that $h(\theta, x) = g(\theta, x, x)$, where $g : \Theta \times S \times S \to \mathbb{R}^m$ solves

$$\lambda(\theta, y) g(\theta, x, y) - \mathcal{A}_\theta g^{(\theta, x)}(y) = \lambda(\theta, y) f(\theta, x, y). \tag{2.4}$$

---

let, and Zakoïan [1998], Clement [1995], Gallant and Long [1997], Gallant and Tauchen [1996], Gallant and Tauchen [1997], He [1990], Kessler [1997], Kessler [1996], Kessler [2000], Kessler and Sorensen [1999], and Stanton [1997].

(We apply $\mathcal{A}_\theta$ to $\mathbb{R}^m$-valued functions, component-wise.) For the case of constant λ, (2.4) follows from the fact that $g(\theta, x, \cdot\,)$ is seen from (2.4) to be the λ-potential of $\lambda f(\theta, x, \cdot\,)$. (See Ethier and Kurtz [1986], Proposition 2.1, page 10.) More generally, (2.4) is established later in Proposition 3.

For example, if X is a diffusion process, then (2.4) is the probabilistic "Feynman-Kac" solution of the elliptic partial differential equation (PDE) (2.4). In this case, X satisfies a stochastic differential equation of the form

$$dX(t) = \mu(\theta^*, X(t))\, dt + \sigma(\theta^*, X(t))\, dB_t, \tag{2.5}$$

where B is a standard Brownian motion in $\mathbb{R}^k$ for some integer $k \geq 1$, and where μ and σ are measurable functions on $\Theta \times \mathbb{R}^k$ into $\mathbb{R}^k$ and $\mathbb{R}^{k\times k}$, respectively, such that there is a unique (strong) solution[2] to (2.5) for each θ in Θ and each initial condition x in the state space $S \subset \mathbb{R}^k$. With (2.5) and any twice continuously differentiable $G : S \to \mathbb{R}$ such that $\mathcal{A}_\theta G$ is well defined,

$$\mathcal{A}_\theta G(x) = \partial_x G(x)\mu(\theta, x) + \frac{1}{2}\text{trace}\left[\partial^2_{xx} G(x)\sigma(\theta, x)\sigma(\theta, x)^\top\right]. \tag{2.6}$$

The PDE (2.4) for this generator $\mathcal{A}_\theta$ could typically be solved numerically, say by finite-difference methods, except for special cases that admit explicit solutions. The computational burden of numerical solutions to (2.4), however, is an impediment. In any case, this introduces a source of approximation that, while potentially negligible with a sufficient computational budget, is difficult to treat theoretically.

An alternative numerical approach is Monte Carlo simulation of X, using the distribution associated with θ, for each candidate parameter θ. In some cases, one can directly simulate from the distribution of $X(T_1) = Z_1$ given $X(0)$. Because $Z_1, Z_2, \ldots$ is a discrete-time time-homogeneous Markov process, moment conditions developed in this fashion amount to simulated-method-of-moments estimation. (See Duffie and Singleton [1993] or Ingram and Lee [1991].) If one assumes, for example, that λ is constant, then in order to simulate $X(T_{i+1})$ given $X(T_i)$ it would be enough to simulate $T_{i+1} - T_i$ with the exponential (λ) density, and then to simulate the outcome of the increment of X associated with a deterministic time period whose length is the outcome of $T_{i+1} - T_i$. For related econometric methods, see Corradi and

[2]For sufficient technical conditions, see for example Karatzas and Shreve [1988]. It is enough that μ and σ are Lipschitz with respect to x.

Swanson [2001], Gallant and Tauchen [1996], Gallant and Tauchen [1997], and Gourieroux, Monfort, and Renault [1993]. Simulation with a stochastic arrival intensity driven by X is outlined in Section 7. For cases in which simulation directly from the distribution of the increments of X is unavailable, one could simulate a discrete-time approximation of X. For example, the Euler approximation of a stochastic differential equation, or the approximations of Milshtein [1978] and Milshtein [1985], have weak convergence properties that have been characterized and extended.[3] Simulated method-of-moments estimation has been based on the assumption that the approximation error associated with time discretization is negligible, at least asymptotically, as the length of a time increment shrinks to zero. (See, for example, Clement [1995], Gallant and Long [1997], Gallant and Tauchen [1996], Gallant and Tauchen [1997], Lesne and Renault [1995].)

Even simulation based on the exact probability distribution of increments of X involves loss of efficiency over explicitly given moments, and an associated computational burden. With time discretization, one is also faced with the theoretical issue of joint convergence, with both discretization interval and sample size, that is uniform over the parameter space.

Yet another computational alternative, proposed by Stanton [1997] and Aït-Sahalia [2002], for deterministic observation times, is to approximate the transition operator that maps f to h defined by (2.1) through an analytic expansion of the infinitesimal generator $\mathcal{A}$. This is relatively more effective for small time periods, and provides straightforward approximate non-parametric estimators for the drift and diffusion functions associated with stochastic differential equations. In another non-parametric estimation approach, Aït-Sahalia [1996] uses a mixed procedure involving transition moment conditions for a parametrized drift function, as well as a non-parametric estimator for the diffusion in terms of the estimated drift and estimated stationary density of the process.

Our approach is to fix some judiciously chosen test function $g : \Theta \times S \times S \to \mathbb{R}^m$ such that $\mathcal{A}_\theta g^{(\theta,x)}$ is well defined, and only then to define f via (2.4). That is, we let

$$f(\theta, x, y) = g(\theta, x, y) - \frac{\mathcal{A}_\theta g^{(\theta,x)}(y)}{\lambda(\theta, y)}, \quad (\theta, x, y) \in \Theta \times S \times S. \tag{2.7}$$

[3] See Bally and Talay [1996], Duffie and Glynn [1995], Klöden and Platen [1992], Protter and Talay [1997], and Talay and Tubaro [1990].

Proposition 3, to follow, gives technical conditions under which

$$g(\theta, x, x) = E_x^\theta[f(\theta, x, Z_1)]. \tag{2.8}$$

We emphasize that, by first selecting g and only then evaluating f from (2.7), we avoid the difficult numerical step of computing $E_x^\theta[f(\theta, x, Z_1)]$. Applying the law of iterated expectations to (2.8), we can therefore base our estimating equations on the moment condition

$$E_\pi\left[f(\theta^*, Z_i, Z_{i+1}) - g(\theta^*, Z_i, Z_i)\right] = 0, \tag{2.9}$$

where π is the invariant probability measure for $\{Z_1, Z_2, \ldots\}$, to be established in Proposition 3 under recurrence conditions on X.

Fixing a test function g, our estimator θ_n for θ^* given $\{Z_0, \ldots, Z_n\}$ is an element of Θ that minimizes the norm (possibly after applying a weighting matrix) of

$$\Gamma_n(\theta) = \frac{1}{n}\sum_{i=0}^{n-1} \gamma(\theta, Z_i, Z_{i+1}), \tag{2.10}$$

where $\gamma : \Theta \times S \times S \to \mathbb{R}^m$ is defined by

$$\gamma(\theta, x, y) = f(\theta, x, y) - g(\theta, x, x). \tag{2.11}$$

Appendix B offers a typical extension with instrumental variables based on lagged values of Z_i. We are exploiting the fact that, under stationarity conditions given in Proposition 3,

$$\lim_n \Gamma_n(\theta^*) = E_\pi[\gamma(\theta^*, Z_i, Z_{i+1})] = 0 \quad a.s. \tag{2.12}$$

A comparison of the estimator proposed here with that of Hansen and Scheinkman [1995] is provided in Appendix E.

3 Basic Moment Condition

We now give a more complete statement of the econometric setting and justification of our class of moment conditions.

3.1 Setup: Time-Homogeneous Markov Processes

The following definitions are typically found in any basic source treating continuous-parameter Markov processes. [See, for example, Dellacherie and Meyer [1988], Ethier and Kurtz [1986], Meyer [1966], and Sigman [1990].]

We fix a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-algebras of $\mathcal{F}$ satisfying the usual conditions.[4] For our purposes, a *state process* X is defined to be a progressively measurable[5] time-homogeneous Harris-recurrent Markov process valued in a complete separable metric space S, with transition function $\mathcal{P} : S \times [0, \infty) \times \mathcal{B}(S) \to [0, 1]$, where $\mathcal{B}(S)$ denotes the Borel subsets of S. In particular, for any t and $u \geq 0$, and any measurable subset B of the state space, the $\mathcal{F}_t$-conditional probability that $X(t+u)$ is in B is $\mathcal{P}(X(t), u, B)$. We let P_x denote the associated distribution of the sample paths of X determined by[6] initial condition x.

Because of its Harris-recurrence, X has a unique (up to constant multiples) non-trivial σ-finite invariant measure η, with the property that $\eta(B) > 0$ if and only if, for any x in S,

$$P_x\left(\int_0^\infty I(X(t) \in B)\, dt = +\infty\right) = 1,$$

where $I(\,\cdot\,)$ denotes the indicator function. We let L be the space of bounded measurable $f : S \to \mathbb{R}$, endowed with the norm $\|\cdot\|$ defined by

$$\|f\| = \eta\text{−ess} \sup_{x \in S} |f(x)|.$$

This implies that we equate two functions in L if they agree almost everywhere with respect to the invariant measure of X. The transition function $\mathcal{P}$ has a transition semi-group $\mathcal{T} = \{\mathcal{T}(t) : L \to L : t \geq 0\}$ defined by

$$[\mathcal{T}(t)f](x) = E_x[f(X(t))] = \int_S f(y)\, \mathcal{P}(x, t, dy),$$

[4]See, for example, Ethier and Kurtz [1986] for technical definitions not given here.

[5]For progressive measurability, it is enough that X has right- or left-continuous sample paths.

[6]For each probability measure ν on S (endowed with the σ-algebra of its Borel sets), the transition function $\mathcal{P}$ determines a unique probability distribution, denoted P_ν, on the space $S^{[0,\infty)}$ of sample paths of X, endowed with the product σ-algebra, such that the distribution of $X(0)$ under P_ν is ν. See Ethier and Kurtz [1986], page 157. For any given x in S, we let P_x denote the distribution P_ν for the case of $\nu = \delta_x$, the Dirac measure at x.

where E_x denotes expectation with respect to P_x. We always assume that $\{\mathcal{T}(t) : t \geq 0\}$ is strongly continuous.[7]

The generator $\mathcal{A}$ of the semi-group $\{\mathcal{T}(t) : t \geq 0\}$ is defined at some g in L if $\mathcal{A}g$ is well defined in L by

$$\mathcal{A}g(x) = \lim_{t \downarrow 0} \frac{[\mathcal{T}(t)g](x) - g(x)}{t}, \quad x \in S.$$

The set of such g is the domain of $\mathcal{A}$, denoted $D(\mathcal{A})$. In some applications, we might consider $\mathcal{A}g$, defined by (1.1), even if g or $\mathcal{A}g$ are not in L, and treat integrability as a separate issue.

Given a constant $\rho \in (0, \infty)$, the ρ-resolvent operator $U_\rho : L \to L$ of X is defined by

$$[U_\rho f](x) = E_x \left[\int_0^\infty e^{-\rho t} f(X(t))\, dt \right] = \int_0^\infty e^{-\rho t} [\mathcal{T}(t) f](x)\, dt.$$

With no further assumptions, for any f in L, we have $U_\rho f$ in $D(\mathcal{A})$ and

$$f = (\rho I - \mathcal{A})[U_\rho f], \tag{3.1}$$

where I is the identity operator. [See Ethier and Kurtz [1986], Proposition 1.2.1.] For constant observation arrival intensity, (3.1) justifies our basic moment condition (2.8), using the calculation (2.4).

For some integer $d \geq 1$, a measurable set $\Theta \subset \mathbb{R}^d$ parameterizes a family $\{\mathcal{P}_\theta :\, \theta \in \Theta\}$ of transition functions on the same state space S, with strongly continuous semi-groups. For each θ in Θ, we let $P_x^\theta, E_x^\theta, \mathcal{T}^\theta, U_\rho^\theta, \mathcal{A}_\theta$, and $D(\mathcal{A}_\theta)$ be defined in relation to $\mathcal{P}_\theta$ just as $P_x, E_x, \mathcal{T}, U_\rho, \mathcal{A}$, and $D(\mathcal{A})$ were defined[8] in relation to $\mathcal{P}$. The "true" parameter is some θ^* in Θ, so that $P_x^{\theta^*} = P_x$, $E_x^{\theta^*} = E_x$, and so on.

[7] Strong continuity means that for each bounded measurable $f : S \to \mathbb{R}$, we have $\lim_{t \to 0} \|\mathcal{T}(t)f - f\| = 0$. Hansen and Scheinkman [1995] propose, for the domain of $\mathcal{T}$, the space of functions that are square-integrable with respect to η. They establish the analogous notion of strong continuity under weak conditions. A similar approach could be used here.

[8] The fact that $D(\mathcal{A}_\theta)$ is treated as a subset of L, in which two functions are the same if equal η-almost everywhere, is a limitation of this modeling approach that can perhaps be relaxed.

3.2 The Data-Generating Process

For our purposes, a *data-generating process* is a pair (X, N) consisting of a state process X and a counting process N whose i-th jump time $T_i = \inf\{t : N(t) = i\}$ is the i-th time at which X is sampled. Provisionally, with a generalization adopted below, we take it that N is a doubly-stochastic point process driven by X. (See, for example, Daley and Vere-Jones [1988].) That is, conditional on $\{X(t) : t \geq 0\}$, N has the law of a Poisson process with time-varying intensity $\{\lambda(X(t)) : t \geq 0\}$, for some measurable $\lambda : S \to (0, \infty)$.

For each parameter θ, we let $\lambda(\theta, \cdot\,) : S \to (0, \infty)$ denote the intensity associated with θ. In particular, $\lambda(\,\cdot\,) = \lambda(\theta^*, \cdot\,)$. The data available for inference are $\{Z_0, Z_1, Z_2, \ldots\}$, where $Z_i = X(T_i)$, with $T_0 = 0$. We always take it that there is a unique parameter vector θ associated with each generator-intensity pair $(\mathcal{A}_\theta, \lambda(\theta, \cdot\,))$, because the probability distribution of a data-generating process (X, N) is uniquely determined by its generator-intensity pair (once the distribution of the initial stae X_0 is fixed).

If the sampling times $T_0, T_1, T_2, \ldots, T_n$ are observable, and thus available for purposes of forming moment conditions, we can assume without loss of generality that the inter-sampling times are "part of" of the state vector in the sense of the above definition. That is, we can take it that $X(t) = (Y(t), t - T(t))$, where $T(t) = t - \max_i\{T_i : t > T_i\}$, that Y is a state process with some generator $\mathcal{A}_Y$ and state space S_Y, and that the point process N is doubly stochastic driven by Y, with an intensity $\lambda_Y(Y_t)$. Then, for X, we take the state space $S = S_Y \times (0, \infty)$ and the generator $\mathcal{A}$ defined by

$$[\mathcal{A}g(y, t) = \mathcal{A}_Y g(\,\cdot\,, t)](y) + \lambda_Y(y)[g(y, 0) - g(y, t)] + \frac{\partial}{\partial t} g(y, t).$$

We thus use the following definitions to distinguish cases in which sampling times are observable or not, for purposes of inference.

Definition 1 *For a given data-generating process (X, N), sampling times are not observable if N is a doubly-stochastic point process driven by X with intensity $\{\lambda(X_t) : t \geq 0\}$. Sampling times are observable if there is a state process Y such that $X_t = (Y_t, t - T(t))$, and N is doubly stochastic driven by Y.*

If sampling times are observable, then, by definition, the intensity $\lambda(Y_t, t - T(t))$ does not depend on its second argument. These definitions allow us to treat most of the theory with results that apply whether or not sampling times are observable.

3.3 The Law of Large Numbers

In order to obtain the effect of the law of large numbers for our sequence of observations of X over time, we require X to be positive-recurrent, so that its invariant measure η may be taken to be a probability measure. Sufficient conditions for this can be based, for example, on the conditions of Meyn and Tweedie [1994] for geometric ergodicity. For the case of diffusions, Has'minskiǐ [1980] has sufficient conditions.

We are concerned with the stationary behavior of the observed discrete-time process $Z = \{Z_0, Z_1, Z_2, \ldots\}$. Proofs of the following are found in Appendix A.

Proposition 1 *Suppose that X is Harris-recurrent with invariant probability measure η. If $\int_S \lambda(x)\,\eta(dx) > 0$, then Z is a Harris-recurrent Markov chain in discrete time.*

Proposition 2 *Suppose X is Harris-recurrent with invariant probability measure η and $0 < \int_S \lambda(x)\,\eta(dx) < \infty$. Then Z is a Harris-recurrent Markov process in discrete time with invariant probability measure π defined by*

$$\pi(B) = \frac{\int_B \lambda(x)\,\eta(dx)}{\int_S \lambda(x)\,\eta(dx)}.$$

In order to address the case in which sampling times are observable, we also establish Harris-recurrence for the augmented state process $(Z, \tau) = \{(Z_i, \tau_i) : i \geq 0\}$, with $\tau_i = T_i - T_{i-1}$ and $\tau_0 = 0$.

Corollary 1 *Suppose the conditions of Proposition 2 apply. Then (Z, τ) is a Harris-recurrent Markov chain in discrete time, with invariant probability measure $\hat{\pi}$ defined by $\hat{\pi}(B) = E_\pi\left(I[(Z_1, \tau_1) \in B]\right)$, where $I[\,\cdot\,]$ is the event-indicator function. In particular, for any measurable $f : S \times [0, \infty) \to \mathbb{R}$ that is integrable with respect to $\hat{\pi}$, and for any x in S,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(Z_i, \tau_i) = E_\pi[f(X(T_1), T_1)] \quad P_x\text{ -a.s.}$$

3.4 Generic Moment Condition

Our basic moment condition (2.12) is a consequence of the following result. The proof of the first assertion is in Appendix A; the remainder then follows from Corollary 1. We call a measurable $g : \Theta \times S \times S \to \mathbb{R}^m$ a *test function* if, for each x and θ, we have $g^{(\theta,x)} = g(\theta, x, \cdot\,)$ in $D(\mathcal{A}_\theta)$ and if $f(\theta, x, Z_1)$, defined by (2.7), is integrable with respect to P_x^θ.

Proposition 3 *Let g be a test function. Then, for each θ, we have $g(\theta, x, x) = E_x[f(\theta, x, Z_1)]$. Suppose, moreover, that the conditions of Proposition 2 apply. Let $\gamma(\theta, x, y) = g(\theta, x, y) - f(\theta, x, y)$. If $\gamma(\theta, Z_i, Z_{i+1})$ is integrable with respect to π, then, for any x in S,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \gamma(\theta, Z_i, Z_{i+1}) = E_\pi[\gamma(\theta, Z_1, Z_2)] \quad P_x\text{−a.s.}$$

4 Identification

Following Definition 3.10 of Gourieroux and Monfort [1995], we describe our model as "identified" if there is a one-to-one mapping between the parameter θ and the probability distribution of the data $\{Z_0, Z_1, Z_2, \ldots\}$. Identification may be impossible if the observation times are not observable *and* the sampling intensity $\lambda(\theta, X_t)$ depends non-trivially on the unknown parameter vector θ. For example, suppose that Y is a Brownian motion with drift parameter μ and variance parameter σ^2, and suppose that the sampling intensity is an unknown constant $\lambda(\theta, x) = \rho \in (0, \infty)$. The parameter vector may then be taken to be $\theta = (\mu, \sigma^2, \rho)$. From Definition 1, to say that the sampling times are not observable means that we observe $Z_i = Y(T_i)$, but not T_i. In this case, the model is clearly not identified, for (from the scaling property of Brownian motion) the distribution of $\{Y(T_1), Y(T_2), \ldots\}$ is the same for all parameter vectors in a set of the form $\{(k\mu, k\sigma^2, k\rho) : k \in (0, \infty)\}$. For instance, it would be impossible to tell from observing the data if the parameters were doubled. On the other hand, as we shall show, it is possible (quite generally) to identify the model if the sampling times are observable.

It is not immediately obvious whether a model that admits identification can actually be identified by using moment conditions of the test-function type that we propose. We show, however, that this is indeed the case. That is, suppose $\phi \neq \theta^*$. The two propositions in this section show that, provided the sampling times are observable or the intensity function $\lambda(\theta, \cdot)$ is known (that is, does not depend on θ), then there is a test function g such that the associated moment condition $\Gamma_n(\theta^*)$ converges to 0 (under the usual technical conditions) as required by (2.12), but $\Gamma_n(\varphi)$ converges to something non-zero. That is, we can tell the two models apart with a moment condition of the test-function type. We do not, however, propose a recipe for finding such a test function; we merely show that it exists. This is not an unusual shortcoming of GMM estimation theory, which rarely provides a practical method for the selection of moment conditions. In practice, GMM moment conditions are typically *ad hoc*.

The fact that our moment conditions can distinguish, in this sense, between distinct underlying continuous-time models depends in part on the randomness of the sampling times. The well known "aliasing problem" implies that the same probability distribution may apply to observations at fixed deterministic time intervals of two distinct continuous-time Markov processes, despite the fact that the two processes have different probability transition functions in continuous-time.[9] In our setting, the aliasing problem is avoided because random sampling at an intensity effectively captures information regarding transition behavior over arbitrarily short time intervals.[10]

Definition 2 *The model is is identified by test functions if, whenever $\varphi \neq \theta^*$, there is some test function g satisfying*

$$E_\pi \left[f(\theta^*, Z_i, Z_{i+1}) - g(\theta^*, Z_i, Z_{i+1}) \right] = 0 \tag{4.1}$$

and

$$E_\pi \left[f(\varphi, Z_i, Z_{i+1}) - g(\varphi, Z_i, Z_{i+1}) \right] > 0, \tag{4.2}$$

where f is defined by (2.7).

Under the recurrence conditions of Proposition 3, we could equally well have replaced (4.1) and (4.2) with their asymptotic sample counterparts.

[9]See Bergstrom [1990], Harvey and Stock [1985], Phillips [1973], and Robinson [1976] for Gaussian vector-autoregressive processes, and Banon [1978] for more general aliasing issues.

[10]Masry [1983] and Solo [1983] have previously noted that random sampling times may defeat the aliasing problem.

Proposition 4 *Suppose λ is bounded away from zero. If $\lambda(\theta, \cdot)$ does not depend on θ, then the model is identified by test functions.*

The condition that λ is bounded away from zero is used only to ensure that a strongly continuous semi-group is associated with the "time-changed" generator $\mathcal{A}/\lambda$. Weaker technical conditions would suffice.

The idea of the proof is roughly that, under technical conditions, for any fixed $\rho \in (0, 1)$, two state processes have the same generator if and only if they have the same ρ-resolvent operators. That is, $\mathcal{A} = \mathcal{A}_\varphi$ if and only if $U_\rho = U_\rho^\varphi$. This would be enough if the observation arrival intensity function λ is some constant, say ρ, for in this case $g(\theta, \cdot) = \rho[U_\rho^\theta f(\theta, \cdot)]$. We can, however, reduce to the case of constant arrival intensity by a random time change. A complete proof is found in Appendix A.

For the general case of an arrival intensity $\lambda(\theta, X(t))$ that depends on θ, we consider the "speed-corrected" generators

$$\mathcal{B} \triangleq \frac{\mathcal{A}}{\lambda(\theta^*, \cdot)} \quad \text{and} \quad \mathcal{B}^\varphi \triangleq \frac{\mathcal{A}_\varphi}{\lambda(\varphi, \cdot)}. \tag{4.3}$$

If the sampling times are not observable, it may be that any difference between the two underlying generators $\mathcal{A}$ and $\mathcal{A}_\varphi$ is precisely offset by the associated arrival intensities, in that $\mathcal{B} = \mathcal{B}^\varphi$. For example, if $\mathcal{A} = 2\mathcal{A}_\varphi$ and $\lambda(\theta^*, x) = 2\lambda(\varphi, x)$, then the distribution of $Z_1, Z_2, \ldots$ could be identical under P_x and under P_x^φ, although we will collect data at twice the average speed that we would under the "incorrect" generator $\mathcal{A}_\varphi$. Without use of the sampling-time data, there might be no way to distinguish whether $\mathcal{A}$ or $\mathcal{A}_\varphi$ is the correct generator.

Proposition 5 *Suppose, for each θ, that $\lambda(\theta, \cdot)$ is bounded away from zero. If sampling times are observable, then the model is identified by test functions.*

5 Estimation

With the results at hand, we have reduced the problem of estimation to that of a relatively standard GMM setting. The data $\{Z_0, Z_1, Z_2, \ldots\}$ form a Harris-recurrent Markov chain under the technical conditions of Proposition 2, for the case without observation of sampling times, and by its Corollary 1, for the case with observation of sampling times. Thus, under integrability, sample moments converge to their population counterparts (Proposition 3), and it is a question of selecting moments. The test-function moments of our approach can in principle identify the model provided the sampling times are observable (Proposition 5), or if the sampling intensity function is known (Proposition 4). At this point, one can bring in standard GMM theory (Hansen [1982]) for consistency and asymptotic normality of the estimators. Because GMM estimation theory is relatively well known, we relegate typical conditions for consistency and asymptotic normality, as well as the use of instrumental variables, to Appendix B. Some of the integrability conditions for this theory are treated for our setting in Appendix D.

Under technical conditions, maximum-likelihood estimation (MLE) is included as a special case of our estimation approach, as shown in Appendix C, by using a special test function $g(\,\cdot\,)$, which solves

$$\mathcal{A}_\theta g(\theta,x,y)-\lambda(\theta,x)g(\theta,x,y)-f_X(\theta,x,y)=0, \tag{5.4}$$

where f_X is defined by Appendix Equation (C.1) in terms of the log-likelihood gradient. Computation of this likelihood gradient, however, is often difficult in practice.[11] On the other hand, there may be a Markov process W on the same state space S whose dynamics are "similar" to those of X, but parameterized by a different family $\{\mathcal{A}_\theta^W:\theta\in\Theta\}$ of generators, for which one can more easily (or explicitly) compute the log-likelihood gradient f_W. This is related to the idea of "quasi-maximum-likelihood estimation" of Gallant and Tauchen [1996] and Gallant and Tauchen [1997].

In our setting, for example, it would be useful to choose the process W so that we can solve for g_W the analogue of (5.4),

$$\mathcal{A}_\theta^W g_W(\theta,x,y)-\lambda(\theta,x)g_W(\theta,x,y)-f_W(\theta,x,y)=0.$$

With this, one can evaluate

$$f_{XW}(\theta,x,y)=g_W(\theta,x,y)-\frac{\mathcal{A}_\theta g_W^{(\theta,x)}(y)}{\lambda(\theta,y)}.$$

A moment condition for estimation of X is obtained from the fact that

$$E_\pi[g_W(\theta^*,Z_i,Z_{i+1})-f_{XW}(\theta^*,Z_i,Z_{i+1})]=0.$$

[11] See, for example, Chuang [1997], Durham and Gallant [2002], Pedersen [1995b], and Pedersen [1995a].

This moment condition may (or, perhaps, may not) capture some of the benefits of maximum-likelihood estimation, to the extent that the transition behaviors of W and X are indeed similar. We provide an illustration of this concept as Example 6, at the end of the next section. One can of course exploit instrumental variables with this approach, and append (possibly over-identifying) restrictions based on other moment conditions.

Unfortunately, we have no theory providing conditions under which the test functions associated with MLE for a "nearby" model provide "near efficiency" when used as test functions for the actual target model.

6 Examples

A few examples are considered in this section. In all cases, twice-continuously differentiable functions with compact support can be considered as test functions.

Example 1: Continuous-Time Markov Chains. For finite or countably infinite S, under quite general conditions, we have

$$\mathcal{A}_\theta g(i) = \sum_j R(\theta)_{ij}[g(j) - g(i)],$$

where $R(\theta)_{ij}$ is the intensity of transition from state i to state j for parameter θ.

Example 2: Diffusions without Boundaries. For $\theta \in \Theta$, suppose that X satisfies the SDE (2.5), and that $\mathcal{A}_\theta$ is specified by (2.6) on the subset of twice-continuously differentiable functions in its domain. One cannot identify the "speed measure" of X without observation times. It would be sufficient for identification to choose an instrumental variable of the form $\frac{\partial}{\partial x} f(\theta, x) \stackrel{\Delta}{=} H(\theta, x)$ such that $E_{\pi(\theta)}[H(\theta, X(t))^2] < \infty$.

Example 3: Jump Diffusions. Consider a jump-diffusion whose generator $\mathcal{A}_\theta$ takes the form (for smooth g)

$$\mathcal{A}_\theta g(x) = \mathcal{A}_{\mu,\sigma} g(x) + \int_S [g(x+y) - g(x)]\, \nu(\theta, x, dy),$$

where $\mathcal{A}_{\mu,\sigma}$ is the diffusion generator defined by the right-hand side of (2.6), and $\nu(\theta, x, \cdot\,)$ is the jump measure, which controls both the arrival intensities and probability distribution of jumps of various types.[12] (This is not the most general form of jump-diffusion.) If the Laplace transform of the jump distribution $\nu(\theta, x, \cdot)$ is known, it may be useful to consider a test function g of the form $g(x) = e^{\alpha \cdot x}$, for some coefficient vector $\alpha \in \mathbb{R}^k$. Barndorff-Nielsen, Jensen, and Sorensen [1998] characterize the stationary distributions of special cases of Ornstein-Uhlenbeck processes driven by homogeneous Lévy processes. This class generalizes the usual Ornstein-Uhlenbeck process and may be suitable for applications in finance.

Example 4: Diffusion with Reflection. For certain applications, such as the former European Exchange Rate Mechanism (see, for example, Krugman [1991] and Froot and Obstfeld [1991]), the case of a diffusion process with reflection has been considered. To illustrate, suppose that X is a 1-dimensional Brownian motion "reflecting" at zero, with constant drift and variance coefficients μ and σ, respectively. That is,

$$dX(t) = \mu\, dt + \sigma\, dB_t + dU_t,$$

where B is a standard Brownian motion and U increases only "when" $X(t) = 0$, so that X is non-negative. Then, for twice continuously differentiable G in $D(\mathcal{A})$ with $G'(0) = 0$, Ito's Formula implies that

$$\mathcal{A}G(x) = G'(x)\mu + \frac{\sigma^2}{2} G''(x),$$

supplying a family of useful test functions. A similar approach applies to more general stochastic differential equations with reflection (or other boundary conditions, for that matter). For approximate computation of the likelihood function, see Aït-Sahalia [2002] and Chuang [1997].

Example 5: MLE without observation of sampling times. Suppose that we observe a Brownian motion with variance parameter $\sigma^2 > 0$ at the event times $T_1, T_2, \ldots$ of an independent Poisson process with rate λ. In other words, we observe $(Z_n : n \geq 0)$, where $Z_n = X(T_n) = \sigma B(T_n)$, for a standard Brownian motion B, with a goal of estimating σ^2. Then, the likelihood function for $\{Z_0, \ldots, Z_n\}$ is

$$\prod_{i=0}^{n-1} p(\sigma, Z_i, Z_{i+1}),$$

[12]See Gihman and Skorohod [1972] and Protter [1990].

where

$$p(\sigma, x, y) = \int_0^\infty \lambda e^{-\lambda t} \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(- \frac{(y-x)^2}{2\sigma^2 t} \right) dt$$
$$= \sqrt{\frac{\lambda}{2}} \frac{1}{\sigma} \exp\left(-|y-x| \frac{\sqrt{2\lambda}}{\sigma} \right).$$

The MLE first-order condition from from which σ^2 can be estimated is therefore

$$\sum_{i=0}^{n-1} \frac{\partial}{\partial \sigma} \log\, p(\sigma, Z_i, Z_{i+1}) \Big|_{\sigma=\sigma_n} = \sum_{i=0}^{n-1} f\left(\sigma_n, X(T_i), X(T_{i+1})\right) = 0, \qquad (6.1)$$

where

$$f(\sigma, x, y) = \frac{-1}{\sigma} + |x-y| \frac{\sqrt{2\lambda}}{\sigma^2}.$$

In order to establish that MLE can be recovered as a special case of our framework, we therefore want a function $g_x(\,\cdot\,)$ for which

$$(\mathcal{A} g_x)(y) - \lambda g_x(y) = \frac{-1}{\sigma} + |x-y| \frac{\sqrt{2\lambda}}{\sigma^2},$$

where $\mathcal{A}$ is the infinitesimal generator associated with Brownian motion having variance parameter σ^2, namely $\mathcal{A} = (\sigma^2/2) d^2/dx^2$. In order that the solution g_x can be represented as the expectation

$$g_x(y) = \frac{-1}{\lambda} E_y \Big[f(\sigma, x, X(T_1)) \Big],$$

g_x and g_x' must be continuous at $y = x$, and $g_x(\cdot)$ must grow slowly enough at infinity in order that appropriate martingale arguments can be invoked. Subject to these boundary conditions,

$$g_x(y) = \frac{1}{\sigma\lambda} - \frac{1}{\sigma\lambda} \exp\left(- \frac{\sqrt{2\lambda}}{\sigma} |y-x| \right) - \frac{\sqrt{2/\lambda}}{\sigma^2} |y-x|.$$

As expected, $g_x(x) = 0$, so the moment condition (2.9) does indeed therefore coincide with the maximum-likelihood estimating equation.

Example 6: Using MLE test functions from an auxiliary model.

Now, suppose that X is an Ornstein-Uhlenbeck process, with $dX_t = -\mu X_t \, dt + \sigma \, dB_t$, for a standard Brownian motion B, and that X is observed at the event times $(T_n : n \geq 0)$ of an independent Poisson process with rate $\lambda > 0$. If the mean-reversion rate μ is small, X has a transition distribution close to that of Brownian motion, at least perhaps over time scales of the order of the Poisson inter-event times. In Example 5, we computed the function $g_x(\,\cdot\,)$ that corresponds to maximum likelihood estimation in the Brownian setting. If we use this test function g_x in the Ornstein-Uhlenbeck context, then we obtain the estimating equation

$$
\begin{aligned}
0 \;=\; & \sum_{i=0}^{n-1} \Bigg[\lambda - |X(T_i) - X(T_{i+1})\mid \sqrt{2}\frac{\lambda^{\frac{3}{2}}}{\hat{\sigma}_n} - \mu X(T_{i+1}) \frac{\sqrt{2\lambda}}{\hat{\sigma}_n} \\
& \cdot \Big(\exp\left(-\frac{\sqrt{2\lambda}}{\hat{\sigma}_n} (X(T_{i+1}) - X(T_i - 1) - 1 \right) \cdot \operatorname{sign}[X(T_{i+1}) - X(T_i)] \Big) \Bigg],
\end{aligned}
$$

from which the estimator $\hat{\sigma}_n^2$ for the parameter σ^2 can be compared.

7 Data Arrival

This section describes some issues related to observation times and incompletely observed state information.

7.1 Market-Time Subordination

It has been noted that transactions frequency and volume are related to the distribution of asset returns. One may think in terms of a measure of "market time," differing from calendar time, under which returns are stationary. For example, one's intuition may be that trading frequency (or volume) is higher during periods of faster information arrival. Mandelbrot and Taylor [1967] and Clark [1973] have proposed influential subordination models, based on a "random time change" under which stationarity applies. Recent work includes that of Conley, Hansen, Luttmer, and Scheinkman [1997], Geman, Madan, and Yor [2001], Ghysels and Jasiak [1994], Ghysels, Gourerioux, and Jasiak [1997], Ghysels, Gourerioux, and Jasiak [2004], Gourieroux, Jasiak, and Fol [1999], Jasiak [1998], Redekop [1995], and Russel and Engle [1998].

In that spirit, we could suppose that the arrival intensity of data is controlled in part by a strictly increasing continuous adapted random market-time process ℓ, with $\ell(0) = 0$, so that $\ell(t)$ represents the amount of market time that has transpired after t units of real time. The underlying state process $X = \{X(t) : t \geq 0\}$, in real time t, can be viewed in market time as the process Y defined by $Y(s) = X(\ell^{-1}(s))$. Then, the number of data points by real time t is $N(\ell(t))$, where $\{N(s) : s \geq 0\}$ is assumed to be a doubly-stochastic counting process driven by Y, with market-time intensity $\{\lambda(Y(s)) : s \geq 0\}$. (Our base case is equivalent to $\ell(t) = t$.) In this case, we could also suppose that the state process Y, measured in market time, is a time-homogeneous state process, in the sense assumed in our basic results.

7.2 Observation at Fixed Deterministic Time Intervals

In the spirit of Conley, Hansen, Luttmer, and Scheinkman [1997], we might even rely on the notion of random time subordination to allow for real-time sampling at fixed intervals (say, hourly).

Specifically, suppose that we observe the underlying state process X at integer real times $1, 2, \ldots$. We let $U_i = \ell(i) - \ell(i-1)$ denote the amount of market time that passes between real times $i-1$ and i. It is conventional to measure market time via some observable process, such as volume of trade. We have some reservations about this interpretation, although one can certainly take it as a definition of market time. We proceed in any case with the assumption that U_i is observable, a common assumption in the market-time literature, and defer a discussion of the treatment of unobserved state variables until later in this section.

We let $Y(s) = X(\ell^{-1}(s))$ define the state process, in market time, with generator $\mathcal{A}$. Conditional on the state path $Y = \{Y(s) : s \geq 0\}$, we suppose, for each integer i, that the amount $U_i = \ell(i) - \ell(i-1)$ of market time that transpires between real times $i-1$ and i has a density $p_i(\,\cdot\,|\,Y) : (0, \infty) \to (0, \infty)$. The assumption that this conditional density is strictly positive implies that there is no upper bound on the expected rate of passage of market time, per unit of real time. Then we can define the conditional hazard rate h_i for U_i given Y by

$$h_i(u) = \frac{p_i(u\,|\,Y)}{P(U_i > u\,|\,Y)}, \quad u \in (0, \infty). \tag{7.1}$$

It is not unreasonable in a stationary setting to suppose that $h_i(u)$ depends only on the amount u of market time that has passed since the last observation and on the current state $Y(\ell(i-1) + u)$, so that we could define a fixed

$\lambda : S \times (0, \infty) \to (0, \infty)$ by $\lambda(Y(\ell(i-1)+u), u) = h_i(u)$. It follows from (7.1) that, for a given $u > 0$, conditional on Y and $U_i \geq u$, the density of U_i at any $s > u$ is

$$\exp\left(-\int_u^s \lambda(Y_r, r)\, dr\right) \lambda(Y_s, s). \tag{7.2}$$

For a given bounded measurable $f : S \times (0, \infty) \to \mathbb{R}$, if we let

$$g(Y(\ell(i-1)+u), u) = E[f(Y(\ell(i)), U_i) \,|\, Y(\ell(i-1)+u), U_i \geq u],$$

then it follows from iterated expectations and the form (7.2) of the conditional density that

$$g(Y(\ell(i-1)+u), u) = E\left[\int_u^\infty \delta_{u,s} \lambda(Y_s, s) f(Y(s), s)\, ds \;\middle|\; Y(\ell(i-1)+u)\right],$$

where $\delta_{u,s} = \exp\left(-\int_u^s \lambda(Y_r, r)\, dr\right)$. Therefore, by the same arguments used in Appendix A,

$$\mathcal{A}g(y, u) + \frac{\partial g(y, u)}{\partial u} - \lambda(y, u) g(y, u) = f(y, u)\lambda(y, u), \tag{7.3}$$

for (y, u) in $S \times (0, \infty)$.

Now, if we let $Z_i = (X_i, U_i)$, then $Z = \{Z_1, Z_2, \ldots\}$ is a time-homogeneous Markov chain. Following our established line of attack, for a given g such that the lefthand side of (7.3) is well defined, we can use (7.3) to define f. Then, under integrability,

$$g(X_i, 0) = E[f(X(i+1), U_{i+1}) \,|\, X_i].$$

Under positive recurrence, this defines a moment condition that can be used for estimation and inference, as we have shown. We omit the details.

7.3 Latent States and Simulated Method of Moments

Clark [1973] estimates a latent-factor model, in which, effectively, one cannot observe the market time process driving the observation point process N. In some cases, latent state variables can be observed up to parameters, and method-of-moments can be applied directly, with care, as in Dai and Singleton [2000].

Simulated-method-of-moment estimators can in principle deal quite generally with latent state variables (as, for example, in Duffie and Singleton [1993]). One would want, however, a tractable method by which to simulate data. In our random-sampling-time setting, this can be done, in certain cases, as follows.

Let (X, N) be a data-generating process, with stochastic arrival intensity $\{\lambda(X_t) : t \geq 0\}$. We can simulate T_{i+1} given $X(T_i)$ by inverse-CDF simulation from a uniform-$[0,1]$ simulated random variable, using the fact that

$$\zeta(t \,|\, X(T_i)) \triangleq P(T_{i+1}-T_i > t \,|\, X(T_i)) = E\left[\exp\left(\int_{T_i}^{T_i+t} -\lambda(X_s)\, ds\right) \,\middle|\, X(T_i)\right].$$

For example, in setting of affine jump diffusions, as explained by Duffie and Kan [1996], $\zeta(t \,|\, x)$ is explicit in many cases, or easily obtained numerically by solving an ordinary differential equation that does not depend on x.

Then, under technical regularity, the density $q(\,\cdot \mid X(T_i), T_{i+1} - T_i)$ of $X(T_{i+1})$ given $(X(T_i), T_{i+1})$ is given by

$$q(y \,|\, x, t) = \frac{\xi(y \,|\, x, t)\lambda(y, t)\zeta(t \,|\, x)}{\zeta_t(t \,|x)},$$

where $\xi(y \,|\, x, t)$ is the transition density of X and ζ_t is the conditional density of $T_{i+1} - T_i$ given $X(T_i)$. This follows from Bayes' Rule and the likelihood calculations in Appendix C.

If ξ and ζ_t are known explicitly, then simulated method of moments may be computationally feasible. For example, if simulation directly from the density $q(\,\cdot\,|x, t)$ is intractable, one can nevertheless use importance sampling and simulate from alternative distributions (with the same support) for which simulation is tractable. For example, in the affine jump-diffusion setting, both ζ_t and the Fourier transform $\hat{\xi}(\,\cdot \,|\, x, t)$ of $\xi(\,\cdot \,|\, x, t)$ are known analytically. We can thus simulate $X(T_{i+1})$ from an alternative distribution with a conditional density $\eta(\,\cdot \mid X(T_i), T_{i+1} - T_i)$, and correct for use of the "wrong" density by scaling the candidate test function by the Radon-Nikodym derivative

$$\frac{q(X(T_{i+1}) \,|\, X(T_i), T_{i+1} - T_i)}{\eta(X(T_{i+1}) \,|\, X(T_i), T_{i+1} - T_i)}.$$

Evaluation of $\xi(X(T_{i+1} \,|\, X(T_i), T_{i+1}-T_i)$ can be done, for example, by Fourier transform from $\hat{\xi}(\,\cdot \,|\, x, t)$.

7.4 Asynchronously Observed Coordinate Processes

Our approach cannot be applied directly to cases in which the state vector is of the form $X = (X^{(a)}, X^{(b)})$, with component processes $X^{(a)}$ and $X^{(b)}$ (each possibly multi-dimensional) that are asynchronously observed. One can treat this with two-stage procedures if one of the component processes, say $X^{(a)}$, is autonomous (that is, $X^{(a)}$ is a Markov process on its own). One can also treat this problem by approximation. In Duffie and Glynn [1997], we provide somewhat more complicated results for the general case of asynchronous arrival, relying to some degree on observation of sampling times.

Appendices

A Proofs

Throughout, we treat the cases of observable and unobservable sampling times simultaneously by letting $Q(X_t) = X_t$ if sampling times are not observable, and, for observable sampling times, as defined in Section 3.2, by letting $Q((Y_t, t - T(t))) = Y_t$. Then, after conditioning on $\{Q(X_t) : t \geq 0\}$, the observation counting process N is Poisson with intensity $\{\lambda(X_t) : t \geq 0\}$.

Proposition 1.

We must show that there is a non-trivial σ-finite reference measure $\tilde{\eta}$ on S such that if $\tilde{\eta}(B) > 0$, then for any x in S, $P_x\left(\min\{n \geq 0 : Z_n \in B\} < \infty\right) = 1$. The strong law of large numbers for Harris-recurrent Markov processes (see Sigman [1990]) guarantees that, for any measurable $B \subset S$ and any x in S,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t I(X(s) \in B)\lambda(X(s))\, ds = \tilde{\eta}(B) \triangleq \int_B \lambda(y)\, \eta(dy) \quad P_x\text{-a.s.}$$

Thus, whenever $\tilde{\eta}(B) > 0$, for any x in S,

$$P_x\left(\int_0^\infty I(X(t) \in B)\lambda(X(t))\ dt = +\infty\right) = 1.$$

For each x in S, however, using the assumption in Section 3.2 on the distribution of T_i,

$$\begin{aligned}
&P_x\left(\min\{n \geq 0 : Z_n \in B\} = \infty\right) \\
&\quad = E_x\left[P_x\left(\min\{n \geq 0 : Z_n \in B\} = \infty \;\middle|\; \{Q(X(t)) : t \geq 0\}\right)\right] \\
&\quad = E_x\left[\exp\left(-\int_0^\infty \lambda(X(t)) I(X(t) \in B)\, dt\right)\right] \\
&\quad = 0,
\end{aligned}$$

completing the proof.

Proposition 2.

Suppose $f : S \to \mathbb{R}$ is measurable, non-negative, and bounded. Then, for any x in S,

$$\begin{aligned}
E_x\left[\sum_{i=1}^{N(t)} f(Z_i)\right] &= E_x\left(E_x\left[\sum_{i=1}^{N(t)} f(Z_i) \;\middle|\; \{Q(X(t)) : t \geq 0\}\right]\right) \\
&= E_x\left[\int_0^t f(X(s))\lambda(X(s))\, ds\right].
\end{aligned}$$

Because f is non-negative,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t E_x[f(X(s))\lambda(X(s))]\, ds = \int_S \lambda(y) f(y)\, \eta(dy).$$

For this fact, see Sigman [1990] and Glynn and Sigman [1998]). Furthermore, we can represent N in the form $N(t) = \tilde{N}(\Lambda^{-1}(t))$, where $\tilde{N}$ is a unit-rate Poisson process independent of X and

$$\Lambda(t) = \int_0^t \lambda(X(s))\, ds.$$

We fix some x in S. By the strong law for Harris-recurrent Markov processes,

$$\lim_{t\to\infty} \frac{1}{t}\Lambda(t) = \int_S \lambda(y)\, \eta(dy) \quad P_x\text{–a.s.},$$

so that

$$\lim_{t\to\infty} \frac{1}{t}\Lambda^{-1}(t) = \frac{1}{\int_S \lambda(y)\, \eta(dy)} \quad P_x\text{–a.s.},$$

and hence

$$\lim_{t\to\infty} \frac{1}{t} N(t) = \int_S \lambda(y)\, \eta(dy) \quad P_x\text{–a.s.}$$

Because $n = N(T_n)$, this ensures that

$$\lim_{n\to\infty} \frac{T_n}{n} = \kappa \triangleq \frac{1}{\int_S \lambda(y)\, \eta(dy)} \quad P_x\text{–a.s.}$$

Now, because f is non-negative and bounded, for any $\epsilon > 0$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} E_x\left[\sum_{i=1}^{n} f(Z_i)\right] \\
&= \lim_{n\to\infty} \frac{1}{n} E_x\left[\sum_{i=1}^{N(T_n)} f(Z_i)\right] \\
&\le \lim_{n\to\infty} \frac{1}{n} E_x\left[\sum_{i=1}^{N(n(\kappa+\epsilon))} f(Z_i)\right] + \lim_{n\to\infty} \sup_{y\in S} |f(y)|\, P_x\left(\frac{T_n}{n} \ge \kappa+\epsilon\right) \\
&= (\kappa+\epsilon) \int_S \lambda(y) f(y)\, \eta(dy),
\end{aligned}$$

using the fact that $P_x\left(T_n/n \ge \kappa+\epsilon\right) \to 0$ as $n \to \infty$. Because ϵ was arbitrary,

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E_x[f(X_i)] \le \frac{\int_S \lambda(y) f(y)\, \eta(dy)}{\int_S \lambda(y)\, \eta(dy)}.$$

Similarly, we can show that

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E_x[f(X_i)] \ge \frac{\int_S \lambda(y) f(y)\, \eta(dy)}{\int_S \lambda(y)\, \eta(dy)}.$$

Hence,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E_x[f(X_i)] = \frac{\int_S \lambda(y) f(y)\, \eta(dy)}{\int_S \lambda(y)\, \eta(dy)}.$$

As x was chosen arbitrarily, the result follows from Glynn [1994].

Corollary to Proposition 2.

We note that (Z, τ) is a Markov chain, because, for any measurable $B \subset S \times [0, \infty)$,

$$P\left((Z_{n+1}, \tau_{n+1}) \in B \;\middle|\; (Z_0, \tau_0), \ldots, (Z_n \tau_n)\right) = P\left((Z_{n+1}, \tau_{n+1}) \in B \;\middle|\; Z_n\right).$$

Furthermore, for $f : S \times [0, \infty) \to \mathbb{R}$ bounded, measurable, and non-negative,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} E_x \left[\sum_{i=1}^{n} f(Z_i, \tau_i)\right] &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E_x \left(E_x \left[f(Z_i, \tau_i) \;\middle|\; Z_{i-1}\right]\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n-1} E_x \left[\tilde{f}(Z_i)\right] \\
&= \frac{\int_S \tilde{f}(y)\lambda(y)\, \eta(dy)}{\int_S \lambda(y)\, \eta(dy)} \quad P_x\text{-a.s.},
\end{aligned}$$

where $\tilde{f}(x) = E_x[f(Z_1, \tau_1)]$. Appealing to Glynn [1994], this establishes that (Z, τ) is a Harris-recurrent Markov chain with invariant probability measure $\hat{\pi}$ as defined, recalling that π is the invariant measure of Z.

Proposition 3.

We will show that $g(\theta, x, x) = E_x^\theta[f(\theta, x, X(T_1))]$. The remainder of the result then follows by Proposition 2, and its Corollary. It is enough to take the case of $\theta = \theta^*$, and for some g in $D(\mathcal{A})$, to let $f(x) = g(x) - \mathcal{A}g(x)/\lambda(x)$. Assuming that $E_x[|f(X(T_1))|] < \infty$, we need to show that $g(x) = E_x[f(X(T_1))]$.

As with (2.4),

$$E_x[f(X(T_1))] = E_x \left[\int_0^\infty \lambda(X_t) \exp\left(\int_0^t -\lambda(X_s)\, ds\right) f(X_t)\, dt\right]. \tag{A.1}$$

Let V be the process defined by $V_t = g(X(t))$. Then V is a semimartingale with $dV_t = \mathcal{A}g(X(t))\, dt + dM_t$, where M is a martingale. [For this, see Ethier and Kurtz [1986], Proposition 1.7, page 156.] Let $U_t = \exp\left(\int_0^t -\lambda(X(s))\, ds\right) V_t$. Then, by Ito's Formula for semimartingales,

$$\begin{aligned}
U_T &= U_0 + \int_0^T -\lambda(X(t)) U_t\, dt + \int_0^T \delta_{0,t}\, dV_t \\
&= g(X(0)) + \int_0^T -\lambda(X(t))\delta_{0,t} \left[g(X(t)) - \frac{\mathcal{A}g(X(t))}{\lambda(X(t))}\right] dt \\
&\quad + \int_0^T \delta_{0,t}\, dM_t,
\end{aligned}$$

where $\delta_{0,t} = \exp\left(\int_0^t -\lambda(X(u))\,du\right)$. Because λ is positive, the last term is a martingale. It follows that, for all t,

$$\begin{aligned} g(x) &= E_x\left[U_T - \int_0^T dU_t\right] \\ &= E_x\left[\delta_{0,T}\, g(X(T))\right] \\ &\quad + E_x\left[\int_0^T \lambda(X(t))\delta_{0,t}\left[g(X(t)) - \frac{\mathcal{A}g(X(t))}{\lambda(X(t))}\right] dt\right]. \end{aligned}$$

From the fact that g is in L and from the condition in Proposition 2 that $\int_S \lambda(x)\,d\eta(x) > 0$, we have $\lim_{t\to+\infty}\int_0^t \lambda(X(s))\,ds = +\infty$ almost surely. We can therefore let T go to infinity and use dominated convergence to get $g(x) = E_x[f(X(T_1))]$ from (A.1). This completes the proof.

Propositions 4 and 5.

Suppose that $\varphi \neq \theta^*$. Because, by definition, each parameter defines a unique generator-intensity pair, we have $(\mathcal{A}, \lambda(\theta^*, \cdot\,)) \neq (\mathcal{A}_\varphi, \lambda(\varphi, \cdot\,))$. Because $\lambda(\theta^*, \cdot\,)$ and $\lambda(\varphi, \cdot\,)$ are bounded away from zero,[13] $\mathcal{B}$ and $\mathcal{B}^\varphi$ are generators for strongly continuous contraction semi-groups.

There are two cases to consider. The first case is that with $\mathcal{B} \neq \mathcal{B}^\varphi$. By the Hille-Yosida Theorem (we use only the part of it in Proposition 2.1, Ethier and Kurtz [1986], page 10), both $I - \mathcal{B} : D(\mathcal{B}) \to L$ and $I - \mathcal{B}^\varphi : D(\mathcal{B}^\varphi) \to L$ are one-to-one and onto. It follows that $(I - \mathcal{B})^{-1} : L \to D(\mathcal{B})$ and $(I - \mathcal{B}^\varphi)^{-1} : L \to D(\mathcal{B}^\varphi)$ are not the same maps.[14] Thus, there is some F in L such that

$$(I - \mathcal{B})^{-1}F \neq (I - \mathcal{B}^\varphi)^{-1}F.$$

We let $G(\varphi, \cdot\,) = (I - \mathcal{B}^\varphi)^{-1}F$ and $G(\theta^*, \cdot\,) = (I - \mathcal{B})^{-1}F$. For both $\theta = \theta^*$ and $\theta = \varphi$, we have

$$G(\theta, x) - \frac{\mathcal{A}_\theta G(\theta, x)}{\lambda(\theta, x)} = F(x).$$

Because two functions in L are the same, by definition, if they are equal η-almost surely, and therefore π-almost surely, there is some measurable set $C \subset S$ with $\pi(C) > 0$, such that

$$G(\theta^*, x) = E_x[F(X(T_1))] \neq G(\varphi, x) = E_x^\varphi[F(X(T_1))], \quad x \in C.$$

[13] We are grateful to Tom Kurtz for suggesting a time change by λ.

[14] If they were the same maps, they would have the same image, and as they are both one-to-one and onto, we would have $D(\mathcal{B}) = D(\mathcal{B}^\varphi)$, contradicting the fact that $\mathcal{B} \neq \mathcal{B}^\varphi$.

Letting

$$\begin{aligned} g(\theta, x, y) &= G(\theta, y) \quad \text{for} \quad G(\theta^*, x) > G(\varphi, x), \\ &= -G(\theta, y), \quad \text{otherwise,} \end{aligned}$$

Proposition 4 follows.

Suppose the other case is true, and $\mathcal{B}^* = \mathcal{B}$. If sampling times are observable, we can define $h : S \to [0, \infty)$ by $h(X_t) = t - T(t)$, the time since the previous observation. Let $q : S \to \mathbb{R}$ be defined by $q(x) = e^{-h(x)}$. We can calculate that $q \in D(\mathcal{A}) \cap D(\mathcal{A}_\varphi)$ and that

$$\mathcal{A}q(x) = (1 - \lambda(\theta^*, x))q(x); \quad \mathcal{A}_\varphi q(x) = (1 - \lambda(\varphi, x))q(x), \quad x \in S. \tag{A.2}$$

From division of the first expression in $(A.2)$ by $\lambda(\theta^*, x)$ and the second by $\lambda(\varphi, x)$, and using the fact that $\mathcal{B} = \mathcal{B}^\varphi$, we see that $\lambda(\theta^*, \cdot\,) = \lambda(\varphi, \cdot\,)$ π-a.s. Thus $\mathcal{A} = \mathcal{A}_\varphi$. Thus, if sampling times are observable, it must be the case that $\mathcal{B} \neq \mathcal{B}^\varphi$ unless $\lambda(\theta^*, \cdot\,) = \lambda(\varphi, \cdot\,)$ π-a.s., and we therefore have identification, proving Proposition 5.

B Consistency and Asymptotic Normality

In this appendix, we summarize conditions, adapted to our setting, for consistency and asymptotic normality. As this material is relatively standard for GMM estimation, once we have reduced the inference problem to that of a stationary Markov process in discrete time, we are brief, and defer to basic sources, such as Hansen [1982], for details and extensions.

B.1 Consistency

For some integer $\ell \geq 1$, we let $Z_{i,\ell} = (Z_i, Z_{i-1}, \ldots, Z_{i-\ell-1})$, a vector of lagged states on which we allow instrumental variables to depend.

Motivated by Proposition 3, we take as given some measurable $\gamma : \Theta \times S^\ell \times S \to \mathbb{R}^m$, for some positive integers ℓ and m, such that

$$E_\pi\left[\|\gamma(\theta^*, Z_{i,\ell}, Z_{i+1})\|^2\right] < \infty, \tag{B.1}$$

with

$$E_\pi\left[\gamma(\theta^*, Z_{i,\ell}, Z_{i+1}) \,|\, Z_{i,\ell}\right] = 0. \tag{B.2}$$

For our application, we have in mind that, for some test function g and some f defined by (2.7), we would take $H(\theta, Z_{i,\ell})$ to be an instrumental variable and

$$\gamma(\theta, Z_{i,\ell}, Z_{i+1}) = H(\theta, Z_{i,\ell})[f(\theta, Z_i, Z_{i+1}) - g(\theta, Z_i, Z_i)]. \tag{B.3}$$

Condition (B.1) is a technical integrability condition. Condition (B.2) is a moment condition for estimation, which is satisfied in our setting for γ as constructed in Proposition 3, and more generally. Our identification condition is that

$$E_\pi\left[\gamma(\theta, Z_{i,\ell}, Z_{i+1})\right] \neq 0, \quad \theta \neq \theta^*. \tag{B.4}$$

For each integer $n \geq \ell$, we let

$$\Gamma_n(\theta) = \frac{1}{n} \sum_{i=\ell-1}^{n-1} \gamma(\theta, Z_{i,\ell}, Z_{i+1}),$$

let W_n be an $\mathcal{F}_n$-measurable $\mathbb{R}^{m \times m}$-valued positive-definite symmetric "weighting" matrix converging almost surely to a positive-definite symmetric matrix measurable with respect to $\sigma(Z_0, Z_1, \ldots)$, and let

$$\theta_n \in \arg \min_{\theta \in \Theta} \Gamma_n(\theta)^\top W_n \Gamma_n(\theta), \tag{B.5}$$

taking a measurable selection from the arg min$(\,\cdot\,)$, which is possible, for example, if γ is continuous and Θ is compact.[15] Newey and West [1987] have developed asymptotic covariance estimators that allow "optimal" choice of the weighting matrices $\{W_n\}$, from the viewpoint of asymptotic efficiency.

Strong consistency depends on technical conditions, such as those of Hansen [1982], appropriate for the uniform strong law of large numbers. For example, with (B.1)-(B.4), we need only add some variation of first-moment continuity, in the sense of Hansen [1982]. For example, consider the "Lipschitz" assumption:

Definition 3 *(First-Moment Continuity). There is some measurable $K : S^\ell \times S \to \mathbb{R}_+$ satisfying $E_\pi[K(Z_{i,\ell}, Z_{i+1})] < \infty$ such that, for each $(z, z_1) \in S^\ell \times S$ and $(\theta, \varphi) \in \Theta^2$,*

$$\|\gamma(\theta, z, z_1) - \gamma(\varphi, z, z_1)\| \leq K(z, z_1)\|\theta - \varphi\|.$$

[15] See Hildenbrand [1974].

For example, it is enough for First-Moment Continuity, when Θ is compact and γ is continuously differentiable with respect to θ, that

$$E_\pi\left(\max_{\theta\in\Theta}\left\|\frac{\partial\gamma}{\partial\theta}(\theta, Z_{i,\ell}, Z_{i+1})\right\|\right) < \infty. \tag{B.6}$$

Theorem 1 *Suppose (X, N) is a data-generating process satisfying the conditions of Proposition 2. Suppose γ satisfies (B.1)-(B.4) and First-Moment Continuity, for a compact parameter set Θ. Let θ_n be defined by (B.5). Then $\theta_n \to \theta^*$ P_x-a.s. for each $x \in S$.*

Proof. It suffices to establish the result under P_π, and then to use shift coupling, obtaining the result under P_x, as in Glynn and Meyn [1995]. In order to invoke our version of the uniform strong law in Theorem 2 below, we let $\mathcal{S} = S^\ell \times S$, let $Y_i = (Z_{i,\ell}, Z_{i+1})$, and set $\varphi(r) = r$. We have, by Cauchy-Schwarz,

$$\begin{aligned} E_\pi\left(\|\gamma(\theta, Z_{i,\ell}, Z_{i+1})\|\right) \ &\leq \ E_\pi^{1/2}\left(\|\gamma(\theta^*, Z_{i,\ell}, Z_{i+1})\|^2\right) \\ &\quad +\|\theta^* - \theta\|\, E_\pi[K(Z_{i,\ell}, Z_{i+1})] < \infty, \end{aligned}$$

for $\theta \in \Theta$. We can now use the identification hypothesis (B.4) to finish, noting that $\Gamma_n(\theta)/(n-\ell-1) \to E_\pi[\gamma(\theta, Z_{i,\ell}, Z_{i+1})]$ uniformly P_π-a.s., by the following slightly more general result.

In order to state a general version of the uniform strong law suitable for this setting, let $Y = \{Y_1, Y_2, \ldots\}$ be a positive Harris-recurrent Markov chain with stationary distribution π, living on a state space $\mathcal{S}$. Let $F : \Theta \times \mathcal{S} \to \mathbb{R}$ satisfy, for some relatively compact Θ, some measurable $K : \mathcal{S} \to \mathbb{R}_+$ with $E_\pi[K(Y_1)] < \infty$, and some $\varphi : [0,\infty) \to [0,\infty)$, with $\varphi(r) \downarrow 0$ as $r \downarrow 0$:

(i) $|F(\theta_1, y) - F(\theta_2, y)| \leq K(y)\varphi(\| \theta_1 - \theta_2 \|)$, for all $y \in \mathcal{S}$ and $\theta_1, \theta_2 \in \Theta$.

(ii) $E_\pi[F(\theta, Y_1)] < \infty$ for all $\theta \in \Theta$. (It suffices that this condition holds for one $\theta \in \Theta$.)

Theorem 2 *Under the above conditions,*

$$\lim_{n\to\infty}\ \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{j=0}^{n-1} F(\theta, Y_j) - E_\pi[F(\theta, Y_0)]\right| = 0 \quad a.s.$$

Our proof is conventional. Fix $\epsilon < 0$ and choose $\delta = \delta(\epsilon) > 0$ so that $\varphi(\delta) < \epsilon$. Then, cover Θ by a finite number $m = m(\epsilon)$ of δ-balls having centers $\theta_1, \theta_2, \ldots, \theta_m$ lying in Θ. Then, any $\theta \in \Theta$ lies in one of the m δ-balls, say the i^{th}, so that

$$\begin{aligned}
\left| \frac{1}{n} \sum_{j=0}^{n-1} F(\theta, Y_j) - E_\pi F(\theta, Y_0) \right| \quad \le \quad & \left| \frac{1}{n} \sum_{j=0}^{n-1} F(\theta, Y_j) - F(\theta_i, Y_j) \right| \\
& + \left| \frac{1}{n} \sum_{j=0}^{n-1} F(\theta_i, Y_j) - E_\pi F(\theta_i, Y_0) \right| \\
& + | \; E_\pi \, F(\theta_i, Y_0) - E_\pi \, F(\theta, Y_0) \, | \\
\le \quad & \frac{1}{n} \sum_{j=0}^{n-1} | \; F(\theta, Y_j) - F(\theta_i, Y_j) \; | \\
& + \max_{1 \le i \le m} \left| \frac{1}{n} \sum_{j=0}^{n-1} F(\theta_i, Y_j) - E_\pi F(\theta_i, Y_0) \right| \\
& + E_\pi \; | \; F(\theta_i, Y_0) \; - \; F(\theta, Y_0) \; | \\
\le \quad & \epsilon \cdot \frac{1}{n} \sum_{j=0}^{n-1} K(Y_j) \\
& + \max_{1 \le i \le m} \left| \frac{1}{n} \sum_{j=0}^{n-1} F(\theta_i, Y_j) - E_\pi F(\theta_i, Y_0) \right| \\
& + \epsilon \, E_\pi \, K(Y_0).
\end{aligned}$$

That is,

$$\begin{aligned}
\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{j=0}^{n-1} F(\theta, Y_j) - E_\pi F(\theta, Y_0) \right| \quad \le \quad & \epsilon \left(\frac{1}{n} \sum_{j=0}^{n-1} K(Y_j) + E_\pi \, K(Y_0) \right) \\
& + \max_{1 \le i \le m} \left| \frac{1}{n} \sum_{j=0}^{n-1} F(\theta_i, Y_j) - E_\pi F(\theta_i, Y_0) \right| .
\end{aligned}$$

Sending $n \to \infty$, we find that

$$\overline{\lim}_{n \to \infty} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{j=0}^{n-1} F(\theta, Y_j) - E_\pi F(\theta, Y_0) \right| \le 2\epsilon \, E_\pi K(Y_0).$$

Since ϵ was arbitrary, this finishes the proof.

Remark: The condition that $E_\pi[K(Y_0)]$ is finite can easily be checked via a Lyapunov function of the form: Find a non-negative $h : \mathcal{S} \to \mathbb{R}$ and a compact A, with h bounded on A, such that

$$E_y[h(Y_1)] \leq h(y) - K(y), \quad y \in A^c,$$

where E_y denotes expectation over the distribution on Y_1 induced by the initial condition $Y_0 = y$. (See page 337 of Meyn and Tweedie [1993a] for details.)

B.2 Asymptotic Normality

As with consistency, one can treat asymptotic normality and calculation of the asymptotic covariance as in a standard GMM framework. The following result is merely for illustration in the case of a one-dimensional parameterization. We refer to Hansen [1982] for typical conditions for more abstract cases.

Theorem 3 *Suppose (X, N) is a data-generating process satisfying the conditions of Proposition 2. Suppose γ satisfies (B.1)-(B.4) and First-Moment Continuity, for a compact interval of parameters $\Theta \subset \mathbb{R}$. Let θ_n be defined by (B.5). Suppose also that γ, of the form (B.3), is continuously differentiable with respect to θ, (B.6) holds, and that*

$$E_\pi\left[\frac{\partial \gamma}{\partial \theta}(\theta^*, Z_{i,\ell}, Z_{i+1})\right] \triangleq a \neq 0.$$

Then, for each $x \in S$,

$$n^{1/2}(\theta_n - \theta^*) \Rightarrow N(0, \sigma^2) \quad P_x\text{−weakly},$$

as $n \to \infty$, where $N(0, \sigma^2)$ denotes the normal distribution with mean zero and variance

$$\sigma^2 = \frac{1}{a^2} E_\pi[\gamma(\theta^*, Z_{i,\ell}, Z_{i+1})^2].$$

Proof. As with the proof of Theorem 1, it suffices to establish the result under P_π. Let γ_θ denote the partial derivative of γ with respect to its first (θ) argument. Note that, for any $\epsilon > 0$ sufficiently small and any $(z, z_1) \in S^\ell \times S$,

$$\left| \frac{\gamma(\theta^* + \epsilon, z, z_1) - \gamma(\theta^*, z, z_1)}{\epsilon} \right| \le \sup_{\theta \in \Theta} |\gamma_\theta(\theta, z, z_1)|.$$

The Dominated Convergence Theorem and (B.6) then imply that

$$E_\pi\left[\gamma_\theta(\theta^*, Z_{i,\ell}, Z_{i+1})\right] = \frac{d}{d\theta} E_\pi\left[\gamma(\theta, Z_{i,\ell}, Z_{i+1})\right] \Bigg|_{\theta=\theta^*}.$$

Since $a \neq 0$, it follows that there exists $\epsilon > 0$ such that

$$E_\pi\left[\gamma(\theta^* - \epsilon, Z_{i,\ell}, Z_{i+1})\right] < 0 = E_\pi\left[\gamma(\theta^*, Z_{i,\ell}, Z_{i+1})\right] < E_\pi\left[\gamma(\theta^* + \epsilon, Z_{i,\ell}, Z_{i+1})\right].$$

Hence, for n sufficiently large,

$$\Gamma_n(\theta^* - \epsilon) < 0 < \Gamma_n(\theta^* + \epsilon),$$

so it follows that $\{\theta \in \Theta : \Gamma_n(\theta) = 0\} \neq \emptyset$. Thus, for sufficiently large n, we have $\Gamma_n(\theta_n) = 0$, and

$$\Gamma_n(\theta_n) - \Gamma_n(\theta^*) = -\Gamma_n(\theta^*).$$

So, there exists $\xi_n \to \theta^*$ P_π-a.s. such that $\Gamma_n'(\xi_n)(\theta_n - \theta^*) = -\Gamma_n(\theta^*)$. Now, for n sufficiently large and $\epsilon > 0$,

$$\begin{aligned} \left| \Gamma_n'(\xi_n) - \Gamma_n'(\theta^*) \right| &= \left| \frac{1}{n} \sum_{i=0}^{n-1} \big(\gamma_\theta(\xi_n, Z_{i,\ell}, Z_{i+1}) - \gamma_\theta(\theta^*, Z_{i,\ell}, Z_{i+1}) \big) \right| \\ &\le \frac{1}{n} \sum_{i=0}^{n-1} R_i(\epsilon), \end{aligned}$$

where

$$R_i(\epsilon) = \sup_{|\theta - \theta^*| < \epsilon} \left| \gamma_\theta(\theta, Z_{i,\ell}, Z_{i+1}) - \gamma_\theta(\theta^*, Z_{i,\ell}, Z_{i+1}) \right|.$$

It follows that

$$\overline{\lim}_{n \to \infty} \left| \Gamma_n'(\xi_n) - \Gamma_n'(\theta^*) \right| \le E_\pi[R_i(\epsilon)].$$

Since γ_θ is continuous by assumption, $\lim_{\epsilon \to 0} \quad R_i(\epsilon) = 0$, P_π-a.s., and

$$R_i(\epsilon) \leq 2 \sup_{\theta \in \Theta} |\gamma_\theta(\theta, Z_{i,\ell}, Z_{i+1})|.$$

The Dominated Convergence Theorem therefore implies that $\lim_{\epsilon \to 0} \; E_\pi \left(R_i(\epsilon) \right) = 0$. Thus,

$$\lim_{n \to \infty} \Gamma'_n(\xi_n) = E_\pi[\gamma_\theta(\theta^*, Z_{i,\ell}, Z_{i+1})] = a, \quad P_\pi\text{−a.s.}$$

Also,

$$n^{1/2}\Gamma_n(\theta^*) \Rightarrow N(0, \beta^2) \qquad P_\pi\text{−weakly},$$

as $n \to \infty$, under the Martingale Central Limit Theorem (see Billingsley [1968], p. 206), where $\beta^2 = E_\pi[\gamma^2(\theta^*, Z_{i,\ell}, Z_{i+1})]/a^2$. The proof is then complete upon application of the converging-together proposition.

Remark: This particular proof becomes harder when Θ is multi-dimensional. In particular, one needs to show that there exists θ_n such that $\Gamma_n(\theta_n) = 0$, in order that the current argument goes through. For a more general approach, see Hansen [1982].

With γ defined by (2.11), we would compute the partial derivative $\partial_\theta \gamma = \frac{\partial \gamma}{\partial \theta}$ by calculating $\partial_\theta f$, $\partial_\theta \lambda$, and $\partial_\theta g$. While g and λ are given functions, computation of $\partial_\theta f$ from the definition (2.7) of f calls for differentiation of $\mathcal{A}_\theta g^{(\theta,x)}(y)$ with respect to θ. Under technical regularity,

$$\partial_\theta[\mathcal{A}_\theta g^{(\theta,x)}(\,\cdot\,)] = (\partial_\theta \mathcal{A}_\theta) g^{(\theta,x)}(\,\cdot\,) + \mathcal{A}_\theta \partial_\theta g^{(\theta,x)}(\,\cdot\,),$$

where, for G in $D(\mathcal{A}_\theta)$, provided the derivative exists, we have

$$[(\partial_\theta \mathcal{A}_\theta)G](y) = \partial_\theta[\mathcal{A}_\theta G](y),$$

a vector in $\mathbb{R}^d$. For the case in which X satisfies the stochastic differential equation (2.5), under smoothness conditions on μ and σ given in Nualart [1995], Chapter 2, the solution X^θ of (2.5) is differentiable with respect[16] to θ, and for a C^2 function G in $D(\mathcal{A}_\theta)$, we have

$$[(\partial_\theta \mathcal{A}_\theta)G](x) = \partial_x G(x) \partial_\theta \mu(\theta, x) + \frac{1}{2} \sum_{i,j} \left[\partial_{x_i} \partial_{x_j} G(x) \partial_\theta C(\theta, x) \right],$$

[16] For this purpose, we can treat θ as a (degenerate) part of the state vector.

where $C(\theta, x) = \sigma(\theta, x)\sigma(\theta, x)^\top$. For other examples given in Section 6, one can also explicitly compute $\partial_\theta \mathcal{A}_\theta$ under technical regularity, provided the parameter-dependent functions defining $\mathcal{A}_\theta$ are smooth with respect to θ. Thus computation of "optimal" weighting matrices can be carried out relatively explicitly in many applications.

C Maximum-Likelihood Estimation

We take the case in which sampling times are observable.[17] A similar result holds when observation times are not observable.

Our data at time T_i is $Z_i = (Y(T_i), T_i)$, with typical outcome denoted $(y_i, t_i) \in S$. We first calculate the relevant likelihood function. We fix some θ in Θ and some initial state $(Y(T_0), 0) = x$. Suppose, for each t, that the transition measure $\mathcal{P}_\theta(t, x, \cdot\,)$ has a density[18] denoted $p(\theta, t, x, \cdot\,)$. Under $\mathcal{P}_\theta$, the joint density of $(Y(T_1), T_1, Y(T_2), T_2, \ldots, Y(T_n), T_n)$, evaluated at some $(y_1, t_1, y_2, t_2, \ldots, y_n, t_n)$, is denoted $L_n(y_1, t_1, y_2, t_2, \ldots, y_n, t_n \,|\, \theta)$. For notational ease, we let

$$
\begin{aligned}
Y^{(n)} &= (Y(T_1), \ldots, Y(T_n)) \\
T^{(n)} &= (T_1, \ldots, T_n) \\
y^{(n)} &= (y_1, \ldots, y_n) \\
t^{(n)} &= (t_1, \ldots, t_n) \\
dt^{(n)} &= dt_1\, dt_2 \cdots dt_n \\
dy^{(n)} &= dy_1\, dy_2 \cdots dy_n \\
\delta_i &= \exp\left(- \int_{t_{i-1}}^{t_i} \lambda(\theta, Y(u))\, du \right).
\end{aligned}
$$

[17] Maximum-likelihood estimation, with integer sampling times, is treated by Aït-Sahalia [2002], Chuang [1997], Clement [1995], Durham and Gallant [2002], Göing [1996], Lo [1988], Pedersen [1995a], Pedersen [1995b], Brandt and Santa-Clara [2002], and Sørensen [2001], using various numerical procedures to esimate the likelihood function.

[18] For the stochastic differential equation (2.5), conditions for existence of the density function p are given in Nualart [1995].

Using the usual abuse of notation for measures with a density,

$$\begin{aligned}
&L(y_1, t_1, y_2, t_2, \dots, y_n, t_n \,|\, \theta)\, dy^{(n)} dt^{(n)} \\
&= P_x^\theta \left(Y^{(n)} \in dy^{(n)},\, T^{(n)} \in dt^{(n)}\right) \\
&= E_x^\theta \left(P_x^\theta \left[Y^{(n)} \in dy^{(n)}, T^{(n)} \in dt^{(n)} \;\middle|\; \{Y(t) : t \geq 0\}\right]\right) \\
&= E_x^\theta \left(E_x^\theta \left[I\left(Y^{(n)} \in dy^{(n)}\right) P_x^\theta \left[T^{(n)} \in dt^{(n)} \;\middle|\; \{Y(t) : t \geq 0\}\right]\right]\right) \\
&= E_x^\theta \left[I\left(Y^{(n)} \in dy^{(n)}\right) \prod_{i=1}^n \lambda(\theta, Y(t_i))\delta_i \, dt^{(n)}\right] \\
&= E_x^\theta \left[E_x^\theta \left(I\left(Y^{(n)} \in dy^{(n)}\right) \prod_{i=1}^n \lambda(\theta, Y(t_i))\delta_i \;\middle|\; Y^{(n)}\right)\right] dt^{(n)} \\
&= \prod_{i=1}^n \lambda(\theta, y_i) E_x^\theta \left(I\left(Y^{(n)} \in dy^{(n)}\right) E_x^\theta \left[\prod_{i=1}^n \delta_i \;\middle|\; Y^{(n)}\right]\right) dt^{(n)} \\
&= \prod_{i=1}^n \lambda(\theta, y_i) p(\theta, t_i - t_{i-1}, y_{i-1}, y_i) \beta(\theta, t_i - t_{i-1}, y_{i-1}, y_i)\, dy^{(n)}\, dt^{(n)},
\end{aligned}$$

where $y_0 = x$ and

$$\beta(\theta, t, x, y) = E_x^\theta \left[\exp\left(-\int_0^t \lambda(\theta, Y(u))\, du\right) \;\middle|\; Y(t) = y\right].$$

(The event "$Y(t) = y$" may have zero probability, but we use this conditioning notation informally, in the usual sense.)

At points of strict positivity of the likelihood, the logarithm of the likelihood is then

$$\log L_n(y_1, t_1, y_2, t_2, \dots, y_n, t_n \,|\, \theta) = \sum_{i=1}^n \xi(\theta, \overline{\tau}_i, y_{i-1}, y_i),$$

where $\overline{\tau}_i = t_i - t_{i-1}$ and

$$\xi(\theta, \overline{\tau}_i, y_{i-1}, y_i) = \log \lambda(\theta, y_i) + \log p(\theta, \overline{\tau}_i, y_i, y_{i-1}) + \log \beta(\theta, \overline{\tau}_i, y_{i-1}, y_i).$$

Assuming differentiability[19] of λ, β, and p with respect to θ, the first-order

[19]For the case of the stochastic differential equation (2.5), smoothness conditions on μ and σ given in Nualart [1995], Chapter 2, ensure the existence and smoothness of the transition density with respect to both x and θ, using the theory of stochastic flows. One can apply Nualart's Lemma 2.1.5 and integration by parts, treating θ as a (degenerate) part of the state vector.

necessary condition for the maximum-likelihood estimator θ_n is

$$0 = \sum_{i=1}^{n} F(\theta_n, \tau_i, Y(T_{i-1}), Y(T_i)),$$

where

$$F_i(\theta, \overline{\tau}_i, y_{i-1}, y_i) = \frac{\frac{\partial}{\partial \theta_i}\lambda(\theta, y_i)}{\lambda(\theta, y_i)} + \frac{\frac{\partial}{\partial \theta_i}\beta(\theta, \overline{\tau}_i, y_{i-1}, y_i)}{\beta(\theta, \overline{\tau}_i, y_{i-1}, y_i)} + \frac{\frac{\partial}{\partial \theta_i}p(\theta, \overline{\tau}_i, y_{i-1}, y_i)}{p(\theta, \overline{\tau}_i, y_{i-1}, y_i)}.$$

Recalling that that the i-th state observation is $X(T_i) = (Y(T_i), T_i - T_{i-1})$, with generic outcome denoted $x_i = (y_i, \overline{\tau}_i) \in S$, we define $f : \Theta \times S \times S \to \mathbb{R}^d$ from $F(\,\cdot\,)$ by

$$f(\theta, x_{i-1}, x_i) = F(\theta, \overline{\tau}_i, y_{i-1}, y_i). \tag{C.1}$$

Maximum-likelihood estimation is then obtained within our general set of moment conditions with the test function $g : \Theta \times S \times S \to \mathbb{R}^d$ defined by

$$\mathcal{A}g(\theta, x_{i-1}, x_i) - \lambda(\theta, x_{i-1})g(\theta, x_{i-1}, x_i) - f(\theta, x_{i-1}, x_i) = 0. \tag{C.2}$$

Indeed, under technical integrability conditions, the fact that

$$E^{\theta}[\xi(\theta, T_1, Y(T_0), Y(T_1)) \,|\, Y(T_0)] = 1$$

for all θ allows one to differentiate through the expectation with respect to θ and find, as usual for maximum-likelihood estimation, that $g(\theta, x, x) = 0$, so for maximum-likelihood estimation we can ignore the role of g in our generic moment condition (2.9).

D Integrability Assumptions

The integrability assumptions (B.1) and (B.6) can be based on primitive assumptions on H, g, and f. For example, with $m = 1$ and differentiability of f, g, and H with respect to θ, consider

Condition 3.

(i) $E_{\pi}\left[H(\theta, Z_{i,\ell})^4 + f\big(\theta, Z_i\big)^4 + g\big(\theta, Z_i\big)^4\right] < \infty, \quad \theta \in \Theta.$

(ii) $E_\pi\left[\sup_{\theta\in\Theta}\left(H_\theta(\theta, Z_{i,\ell})^2\right) + \sup_{\theta\in\Theta}\left(f_\theta(\theta, Z_i)^2\right) + \sup_{\theta\in\Theta}\left(g_\theta(\theta, Z_i)^2\right)\right]$ is finite.

Assumption (B.1) is implied by Condition 3 *(i)* and Cauchy-Schwarz. For (B.6), we have

$$\gamma_\theta(\theta, z, z_1) = H_\theta(\theta, z)[f(\theta, z_1) - g(\theta, z)] + H(\theta, z)\left[f_\theta(\theta, z_1) - g_\theta(\theta, z)\right].$$

Also, for some "intermediate" parameter ζ, taking $\Theta = [\underline{\theta}, \overline{\theta}]$, we have

$$\begin{aligned}
\sup_{\theta\in\Theta}\left|\, H_\theta(\theta, z)\, f(\theta, z_1)\,\right| &= \sup_{\theta\in\Theta}\left|\, H_\theta(\theta, z)\ \left[f(\theta^*, z_1) + (\theta - \theta^*)\ f_\theta(\zeta, z_1)\right]\,\right| \\
&\leq \sup_{\theta\in\Theta}\left|\, H_\theta(\theta, z)\,\right| \cdot \left|\, f(\theta^*, z_1)\,\right| \\
&\qquad + \left|\, \overline{\theta} - \underline{\theta}\,\right| \sup_{\theta\in\Theta}\left|\, H_\theta(\theta, z)\,\right| \ \sup_{\theta\in\Theta}\left|\, f_\theta(\theta, z_1)\,\right|,
\end{aligned}$$

and likewise for $\sup_{\theta\in\Theta}\left|\, H_\theta(\theta, z)\, g(\theta, z_0)\,\right|$.

We now apply Cauchy-Schwarz and Minkowski, as well as Condition 3 *(i)* and *(ii)*, to obtain the finiteness of $E_\pi\left[\sup_{\theta\in\Theta}\left|\, H_\theta(\theta, Z_{i,\ell}) f(\theta, Z_1)\right|\right]$, and similarly for the other terms in (B.6).

The integrability assumptions of Condition 3 are stated purely in terms of the stationary distribution of Z. It turns out that we need not compute the transition kernel of Z in order to verify Condition 3. Lyapunov methods exist for this [see Meyn and Tweedie [1993a]]. For example, consider Condition 3 *(ii)* for the case in which $\ell = 1$.

Set

$$r(x) = \sup_{\theta\in\Theta} H_\theta(\theta, x)^2 + \sup_{\theta\in\Theta} f_\theta(\theta, x)^2 + \sup_{\theta\in\Theta} g_\theta(\theta, x)^2.$$

Proposition 2 ensures the finiteness of $E_\pi[r(Z_0)]$ provided that we establish

$$\overline{r} = E_\eta[r(X(0))\lambda(X(0))] < \infty.$$

The Lyapunov criterion for proving finiteness of $\overline{r}$ basically comes down to finding a non-negative k in $D(\mathcal{A})$ and a compact $A \subset S$ such that

$$\mathcal{A}k(x) \leq -r(x)\lambda(x), \quad x \in A^c.$$

(See Theorem 4.2 of Meyn and Tweedie [1993b] for details.)

For the case of the Cox, Ingersoll, and Ross [1985] model of the short interest rate, a process introduced by Feller [1951], we have

$$dX_t = \kappa\big(\overline{x} - X(t)\big)\, dt + \sigma\sqrt{X(t)}\, dB_t,$$

where B is a standard Brownian motion and κ, $\overline{x}$, and σ are positive scalar parameters. In this case,

$$\mathcal{A} = \kappa(\overline{x} - y)\frac{d}{dy} + \frac{\sigma^2 y}{2}\frac{d^2}{dy^2}.$$

If we set $k(y) = \exp(\epsilon y)$, then there exists $\delta > 0$ such that for y off a compact set,

$$(\mathcal{A}k)(y) = \kappa(\overline{x} - y)\epsilon e^{\epsilon y} + \frac{\sigma^2 y}{2}\epsilon^2 e^{\epsilon y} \leq -\delta y e^{\epsilon y},$$

provided we choose ϵ small enough that $\kappa > \sigma^2\epsilon/2$. Hence, according to Lyapunov theory, we can expect that $E_\pi\left[X(0)\exp\big(\epsilon X(0)\big)\right] < \infty$, for such an ϵ. Thus, so long as

$$\sup_{\theta\in\Theta} H_\theta(\theta, x)^2 + \sup_{\theta\in\Theta} f_\theta(\theta, x)^2 = \mathrm{O}\big(\exp(\epsilon x)\big),$$

we can expect Condition 3*(ii)* to be in force. Similarly, we can verify Condition 3*(i)*.

E Hansen-Scheinkman Estimators

Hansen and Scheinkman [1995] base an estimator on observation of X at integer times $1, 2, \ldots$ and on an assumption that X is ergodic, with stationary distribution η. (Their work is extended by Hansen, Scheinkman, and Touzi [1998], allowing random sampling time intevals, and by Conley, Hansen, Luttmer, and Scheinkman [1997].) Hansen and Scheinkman use the fact that, for any f such that $\mathcal{A}f(X(t))$ is well defined and integrable,

$$E_\eta[\mathcal{A}f(X(t))] = 0, \tag{E.3}$$

where E_η denotes expectation under the stationary distribution η of X. This relies on the simple fact that, under η, the rate of change of the expectation of *any* well-behaved function of the sample paths of X must be zero. This leads, under technical conditions, to a family of moment conditions that assist in estimating θ^*.

The Hansen-Scheinkman (HS) class of estimators, with Poisson inter-arrival times, can be recovered from ours, asymptotically, as follows. Suppose that $g(\theta, x, y) = g(y)$ and that we assume Poisson sampling, for some constant intensity $\overline{\lambda} > 0$. In this case, from telescopic cancellation in (2.10),

$$\Gamma_n(\theta) = \frac{1}{n}\left[f(\theta, Z_n) - g(Z_0)\right] - \frac{1}{n\overline{\lambda}} \sum_{i=1}^{n-1} \mathcal{A}_\theta g(Z_i), \tag{E.4}$$

which corresponds asymptotically to the moment condition (E.3).

Hansen and Scheinkman [1995] also use "reverse-time" moment conditions.[20] Reverse-time versions of our moment conditions can be developed analogously.

Because $\mathcal{A}f(X(1)), \mathcal{A}f(X(2)), \ldots$ are generally correlated, computation of the asymptotic standard errors associated with the HS moment conditions (E.3)-(E.4) may be relatively intractable. On the other hand, because of (2.8), the terms $\gamma(\theta^*, Z_i, Z_{i+1})$ of the criterion proposed here are first differences of a martingale, and therefore uncorrelated. This makes computation of asymptotic standard errors relatively tractable. (See Theorem 3 below.) As remarked above, however, with Poisson sampling times, the HS estimators can be viewed asymptotically as special cases of, and have the same asymptotic behavior as, the estimators presented here, including easily computed asymptotic standard errors.

As opposed to the class of estimators proposed here, HS estimators do not generally offer identification, as explained by Hansen and Scheinkman [1995]. HS estimators do, however, have the advantage that they can be

[20] That is, the reverse-time process Y, defined by $Y(t) = X(-t)$, has its own generator $\mathcal{A}^*$, and under technical regularity we have a condition analogous to (E.3): $E_\eta[\mathcal{A}^* f(X(t))] = 0$. Combining this condition with (E.3), and applying the definition of the adjoint, Hansen and Scheinkman propose a general moment condition of the form

$$E_\eta[g(X(t))\mathcal{A}f(X(t+1)) + f(X(t+1))\mathcal{A}^* g(X(t))] = 0.$$

In a general setting, computation of $\mathcal{A}^*$ is difficult. Under reversibility conditions on X [see for example Kent [1978] and Millet, Nualart, and Sanz [1989]], however, the infinitesimal generator $\mathcal{A}^*$ can be computed. The reversibility conditions are easily satisfied for solutions of 1-dimensional versions of a stochastic differential equation of the form of (2.5). Reversibility is generally a strong condition, however, for multivariate processes.

based on both deterministic and random sampling time schemes, under conditions. Section 7 discusses cases in which a modification of the moment conditions proposed in this paper could be considered with deterministic sampling schemes, after a time change.

The class of estimators associated with (2.10) can, in principle, be applied to arbitrary Markov processes, whether or not recurrent. In particular, the methodology can be used to estimate parameters for transient and null-recurrent Markov processes. The main difficulty is establishing the law of large numbers, in the absence of ergodicity, for purposes of a proof of consistency. In contrast, the HS moment conditions are meaningful only for positive-recurrent processes.

References

Aït-Sahalia, Y. (1996). Nonparametric Pricing of Interest Rate Derivative Securities. *Econometrica* **9**, 385–426.

Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica* **70**, 223–262.

Bally, V. and D. Talay (1996). The Law of The Euler Scheme for Stochastic Differential Equations: II. Convergence Rate of the Density. *Monte Carlo Methods and Applications* **2**, 93–128.

Banon, G. (1978). Nonparametric Identification for Diffusion Processes. *SIAM J. Control and Optimization* **16**, 380–395.

Barndorff-Nielsen, O., J. Jensen, and M. Sorensen (1998). Some Stationary Processes in Discrete and Continuous Time. *Advances in Applied Probability* **30**, 989–1007.

Bergstrom, A. (1990). *Continuous Time Econometric Modelling*. Oxford University Press.

Bibby, B. and M. Sorensen (1995). Martingale Estimation Functions for Discretely Observed Diffusion Processes. *Bernoulli* **1**, 17–39.

Bibby, B. and M. Sorensen (1997). Estimating functions for discretely observed diffusions: A review. In I. V. Basawa, V. Godambe, and R. Taylor (Eds.), *Selected Proceedings of the Symposium on Estimating Functions*, Volume 32 of *IMS Lecture Notes - Monograph Series*, pp. 305–325.

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

Brandt, M. and P. Santa-Clara (2002). Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets. *Journal of Financial Economics* **63**, 161–210.

Brémaud, P. (1981). *Point Processes and Queues – Martingale Dynamics*. New York: Springer-Verlag.

Broze, L., O. Scaillet, and J.-M. Zakoïan (1998). Quasi-Indirect Inference for Diffusion Processes. *Econometric Theory* **14**, 161–18.

Chuang, C. (1997). Approximate Computation of Transition Densities of Continuous-Time Markov Processes. Working Paper, Department of Statistics, University of Wisconsin.

Clark, P. (1973). A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica* **41**, 135–156.

Clement, E. (1995). Modelisation Statistique en Finance et Estimation de Processus de Diffusion. Working Paper, U.F.R. Mathematiques de la Decision, Université de Paris IX, Dauphine.

Conley, T., L. Hansen, E. Luttmer, and J. Scheinkman (1997). Short-Term Interest Rates as Subordinated Diffusions. *Review of Financial Studies* **10**, 525–577.

Corradi, V. and N. Swanson (2001). Bootstrap Specification Tests with Dependent Observations and Parameter Estimation Error. Working Paper, University of Exeter.

Cox, J., J. Ingersoll, and S. Ross (1985). A Theory of the Term Structure of Interest Rates. *Econometrica* **53**, 385–407.

Dai, Q., D. Duffie, and P. Glynn (1997). Asymptotic Efficiency of Hansen-Scheinkman and Resolvent Estimators of Ornstein-Uhlenbeck Process. Working Paper, Graduate School of Business, Stanford University.

Dai, Q. and K. Singleton (2000). Specification Analysis of Affine Term Structure Models. *Journal of Finance* **55**, 1943–1978.

Daley, D. and D. Vere-Jones (1988). *An Introduction to the Theory of Point Processes.* New York: Springer-Verlag.

Dellacherie, C. and P.-A. Meyer (1988). *Probailities and Potential C: Potential Theory for Discrete and Continuous Semigroups.* Amsterdam: North-Holland.

Duffie, D. and P. Glynn (1995). Efficient Monte Carlo Estimation of Security Prices. *Annals of Applied Probability* **5**, 897–905.

Duffie, D. and P. Glynn (1997). Inference for Markov Processes with Asynchronously Observed Coordinate Processes. Working Paper, Research Notes, Stanford University.

Duffie, D. and R. Kan (1996). A Yield-Factor Model of Interest Rates. *Mathematical Finance* **6**, 379–406.

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

Durham, G. and R. Gallant (2002). Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes, forthcoming. *Journal of Business and Economic Statistics* **20**, 297–316.

Ethier, S. and T. Kurtz (1986). *Markov Processes: Characterization and Convergence.* New York: Wiley.

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

Froot, K. and M. Obstfeld (1991). Exchange-Rate Dynamics under Stochastic Regime Shifts, a Unified Approach. *Journal of International Economics* **31**, 203–229.

Gallant, R. and J. Long (1997). Estimation of Stochastic Differential Equations Efficiently by Minimum Chi-Square. *Biometrica* **84**, 125–141.

Gallant, R. and G. Tauchen (1996). Which Moments to Match? *Econometric Theory* **12**, 657–681.

Gallant, R. and G. Tauchen (1997). Estimation of Continuous Time Models for Stock Returns and Interest Rates. *Macroeconomic Dynamics* **1**, 135–168.

Geman, H., D. Madan, and M. Yor (2001). Time Chnages for Lévy Processes. *Mathematical Finance* **11**, 79–96.

Ghysels, E., C. Gourerioux, and J. Jasiak (1997). Market Time and Asset Price Movements: Theory and Estimation. In D. Hand and S. Jacka (Eds.), *Statistics in Finance*, pp. 307–332. Edward Arnold, London.

Ghysels, E., C. Gourerioux, and J. Jasiak (2004). Stochastic Volatility Duration Models. *Journal of Econometrics* **119**, 413–435.

Ghysels, E. and J. Jasiak (1994). Stochastic Volatility and Time Deformation: an Application of Trading Volume and Leverage Effects. Working Paper, CRDE Université de Montréal and CIRANO.

Gihman, I. and A. Skorohod (1972). *Stochastic Differential Equations*. Berlin: Springer-Verlag.

Glynn, P. (1994). Some Topics in Regenerative Steady-State Simulation. *Acta Applicandae Mathematicae* **34**, 225–236.

Glynn, P. and S. Meyn (1995). A Lyapunov Bound for Solutions of Poisson's Equation. *Annals of Probability* **24**, 916–931.

Glynn, P. and K. Sigman (1998). Independent Sampling of a Stochastic Process. *Stochastic Processes and Their Applications* **74**, 151–164.

Göing, A. (1996). Estimation in Financial Models. Working Paper, Department of Mathematics, ETH Zurich.

Gourieroux, C., J. Jasiak, and G. L. Fol (1999). Intra-Day Market Activity. *Journal of Financial Markets* **2**, 193–226.

Gourieroux, C. and A. Monfort (1995). *Statistics and Econometric Models (English Translation)*. Cambridge University Press.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect Inference. *Journal of Applied Econometrics* **8**, S85–S118.

Hansen, L. (1982). Large Sample Propreties of Generalized Method of Moments Estimators. *Econometrica* **50**, 1029–1286.

Hansen, L. and J. Scheinkman (1995). Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes. *Econometrica* **63**, 767–804.

Hansen, L., J. Scheinkman, and N. Touzi (1998). Spectral Methods for Identifying Scalar Diffusions. *Journal of Econometrics* **86**, 1–32.

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

Has'minskiĭ, R. (1980). *Stochastic Stability of Differential Equations*. Sijthoff and Noordhoff.

He, H. (1990). Moment Approximation and Estimation of Diffusion Models of Asset Prices. Working Paper, University of California, Berkeley.

Hildenbrand, W. (1974). *Core and Equilibria of a Large Economy*. Princeton University Press.

Ingram, B. and B.-S. Lee (1991). Simulation Estimation of Time Series Models. *Journal of Econometrics* **47**, 197–250.

Jasiak, J. (1998). Persistence in Intertrade Durations. *Finance* **19**, 165–195.

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

Kent, J. (1978). Time-Reversible Diffusions. *Advanced Applied Probability* **10**, 819–835.

Kessler, M. (1996). Martingale Estimating Functions for a Markov Chain. Working Paper, Laboratoire de Probabilités, Université Pierre et Marie Curie.

Kessler, M. (1997). Estimation of an Ergodic Diffusion from Discrete Observations. *Skandinavian Journal of Statistics* **24**, 211–229.

Kessler, M. (2000). Simple and Explicit Estimating Functions for a Discretely Observed Diffusion. *Scandinavian Journal of Statistics* **27**, 65–82.

Kessler, M. and M. Sorensen (1999). Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion Process. *Bernoulli* **5**, 299–314.

Klöden, P. and E. Platen (1992). *Numerical Solutions of Stochastic Differential Equations*. Berlin: Springer-Verlag.

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

Lesne, J.-P. and E. Renault (1995). Indirect Inference Estimation of Yield Curve Factor Models. Working Paper, Université de Cergy-Pontoise, THEMA, France.

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

Mandelbrot, B. and H. Taylor (1967). On the Distribution of Stock Price Differences. *Operations Research* **15**, 1057–1062.

Masry, E. (1983). Spectral and Probability Density Estimation from Irregularly Observed Data. In E. Parzen (Ed.), *The Statistical Analysis of Irregularly Observed Time Series in Statistics, Lecture Notes in Statistics*, pp. 224–250. New York: Springer Verlag.

Meyer, P.-A. (1966). *Probability and Potentials*. Waltham, Massachusetts: Blaisdell Publishing Company.

Meyn, S. and R. Tweedie (1993a). *Markov Chains and Stochastic Stability*. London: Springer-Verlag.

Meyn, S. and R. Tweedie (1993b). Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continous-Time Processes. *Advances in Applied Probability* **25**, 518–548.

Meyn, S. and R. Tweedie (1994). Computable Bounds for Geometric Convergence Rates of Markov Chains. *Annals of Applied Probability* **4**, 981–1011.

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

Milshtein, G. N. (1978). A Method of Second Order Accuracy Integration of Stochastic Differential Equations. *Theory of Probability and Applications* **3**, 557–562.

Milshtein, G. N. (1985). Weak Approximation of Solutions of Systems of Stochastic Differential Equations. *Theory of Probability and Applications* **30**, 75–766.

Newey, W. and K. West (1987). A Simple Positive Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* **55**, 703–708.

Nualart, D. (1995). *The Malliavin Calculus and Related Topics*. New York: Springer-Verlag.

Pedersen, A. (1995a). Consistency and Asymptotic Normality of an Approximate Maximum Likelihood Estimator for Discretely Observed Diffusion Processes. *Bernoulli* **1**, 257–279.

Pedersen, A. (1995b). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics* **22**, 55–71.

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

Protter, P. (1990). *Stochastic Integration and Differential Equations*. New York: Springer-Verlag.

Protter, P. and D. Talay (1997). The Euler Scheme for Lévy Driven Stochastic Differential Equations. *Annals of Probability* **25**, 393–423.

Redekop, J. (1995). Extreme-Value Distributions for Generalizations of Brownian Motion. Working Paper, Department of Economics, Waterloo University, Ontario, Canada.

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

Russel, J. and R. Engle (1998). Econometric Analysis of Discrete Valued, Irregularly Spaced Financial Transaction Data Using a New Autoregressive Conditional Multinomial Model. *Econometrica* **66**, 1127–1162.

Sigman, K. (1990). One-Dependent Regenerative Processes and Queues in Continuous Time. *Mathematics of Operations Research* **15**, 175–189.

Solo, V. (1983). Some Aspects of Continuous-Discrete Time Series Modelling. In E. Parzen (Ed.), *The Statistical Analysis of Irregularly Observed Time Series in Statistics, Lecture Notes in Statistics*, pp. 326–345. New York: Springer Verlag.

Sørensen, H. (2001). Simulated Likelihood Approximations for Stochastic Volatility Models. Working Paper, Center for Analytic Finance, University of Aarhus.

Stanton, R. (1997). A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. *Journal of Finance* **52**, 1973–2002.

Talay, D. and L. Tubaro (1990). Expansion of the Global Error for Numerical Schemes Solving Stochastic Differential Equations. *Stochastic Analysis and Applications* **8**, 94–120.