Econometrica, Vol. 71, No. 2 (March, 2003), 483–549

THE EFFECTS OF RANDOM AND DISCRETE SAMPLING WHEN ESTIMATING CONTINUOUS-TIME DIFFUSIONS

BY YACINE AÏT-SAHALIA AND PER A. MYKLAND[1]

High-frequency financial data are not only discretely sampled in time but the time separating successive observations is often random. We analyze the consequences of this dual feature of the data when estimating a continuous-time model. In particular, we measure the additional effects of the randomness of the sampling intervals over and beyond those due to the discreteness of the data. We also examine the effect of simply ignoring the sampling randomness. We find that in many situations the randomness of the sampling has a larger impact than the discreteness of the data.

KEYWORDS: Likelihood, diffusion, discrete sampling, random sampling, cost of discreteness, cost of randomness.

1. INTRODUCTION

DIFFUSION MODELS, and their extensions such as jump-diffusions and Markov models driven by Lévy processes, are essential tools for much of theoretical asset pricing. Estimating these models from discrete time observations has become in recent years an active area of research in econometrics and statistics. Beyond the choice of inference strategy, an important debate in this area concerns the question of what sampling scheme to use, if a choice is available, and in any event what to do with the sampling times. The most straightforward thing to do, in accordance with the usual low-frequency data collection procedures in finance, is to view the sampling as occurring at fixed discrete time intervals, such as a day or a week. In many circumstances, however, this is not realistic. In fact, all transaction-level data are available at irregularly and randomly spaced times. For example, Figure 1 shows a histogram distribution of the time between trades for the Nokia shares traded on the New York Stock Exchange during the month of January 2000. A continuous exponential distribution fitted to the same data is also plotted.[2]

Not only are the data randomly spaced in time, but whenever a theoretical model is spelled out in continuous time, its estimation necessarily relies on

[1] We are grateful to a co-editor and two anonymous referees as well as to Oliver Linton, Marcelo Fernandes, and many seminar participants for their comments and suggestions. Financial support from an Alfred P. Sloan Research Fellowship (Aït-Sahalia) and the NSF under Grants SBR-9996023 and SBR-0111140 (Aït-Sahalia) and DMS-9626266 and DMS-9971738 (Mykland) is also gratefully acknowledged.

[2] The data come from the NYSE's Trades and Quotes (TAQ) database. Nokia shares are American Depository Shares (ADS). We aggregated trades that do not result in a price change for the purpose of defining the times between trades. The histogram in Figure 1 is based on 46,063 such trades, with a mean time between trades of 10.7 seconds.

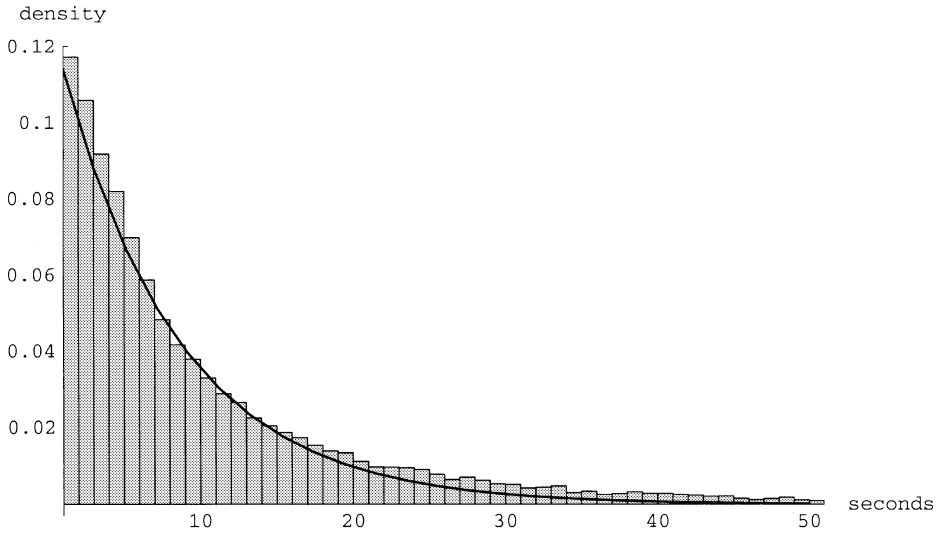


FIGURE 1.—The distribution of the sampling intervals for Nokia, January 2000.

discretely sampled data. In fact, is there any other kind of data? By now, we have a good understanding of the implications of sampling discreteness, and how to design estimation methods that correct for it (see Hansen and Scheinkman (1995), Aït-Sahalia (1996), and Aït-Sahalia (2002)). Our objective in this paper is to understand the *additional* effect that the randomness of the sampling intervals might have when estimating a continuous-time model with discrete data. Specifically, we aim to disentangle the effect of the sampling *randomness* from the effect of the sampling *discreteness*, and to compare their relative magnitudes. We also examine the effect of simply ignoring the sampling randomness. We achieve this by comparing the properties of three likelihood-based estimators, which make different use of the observations on the state process and the times at which these observations have been recorded. We design these estimators in such a way that each one of them is subject to a specific subset of the effects we wish to measure. As a result, the differences in their properties allow us to zero in and isolate these different effects.

Our main conclusion is that the loss from not observing, or not using, the sampling intervals, will be at least as great, and often substantially greater, than the loss due to the fact that the data are discrete rather than continuous. While correcting for the latter effect has been the main focus of the literature in recent years, our results suggest however that empirical researchers using randomly spaced data should pay as much attention, if not more, to sampling randomness as they do to sampling discreteness. A further contribution of this paper is the development of a set of tools that allows these calculations to be performed in closed form.

The paper is organized as follows. Section 2 sets up the model and provides an intuitive description of our approach and results. Section 3 formally lays out our assumptions and presents the theoretical results that underpin our methodology. In Section 4, we propose three types of likelihood-based inference, making different use of the sampling intervals. By comparing these different likelihood-based estimators, we can assess the relative costs attributable to the discreteness of the sampling and to its randomness. We carry out this analysis in Section 5, before analyzing in Section 6 the cost of ignoring the sampling randomness: what happens if one behaves as if the sampling times are not random when in fact they are? Section 7 shows what our general expressions yield in a specific example and reports evidence from simulations. Section 8 concludes. All proofs are in the Appendix.

2. DESCRIPTION OF THE APPROACH

Before delving into formal assumptions and theorems, let us provide a heuristic description of our setup, the estimators and techniques we use, and finally our results.

2.1. *The Setup*

The basic setup is as follows. We assume that the discrete data we observe have been generated by a time-homogeneous diffusion on the real line

$$dX_t = \mu(X_t; \theta)\, dt + \sigma\, dW_t \tag{2.1}$$

where W_t is a Brownian motion, $\mu(.,.)$ is the drift function, σ^2 the diffusion coefficient, and θ the drift parameters, $\theta \in \Theta$ and $\sigma > 0$. We will show that the properties of estimators vary widely depending upon whether only the drift or the diffusion parameters, or both together, are estimated. Hence we consider the three cases of estimating $\beta = (\theta, \sigma^2)$ jointly, estimating $\beta = \theta$ with $\sigma^2 = \sigma_0^2$ known or estimating $\beta = \sigma^2$ with $\theta = \theta_0$ known. As usual, β_0, θ_0, and σ_0^2 denote the true values of the parameters. The parameter space for the d-dimensional vector β is an open and bounded set.

We will assume that the process (2.1) is stationary. The parameter vector β is to be estimated at time T on the basis of $N_T + 1$ discrete observations recorded at times $\tau_0 = 0, \tau_1, \tau_2, \ldots, \tau_{N_T}$, where N_T is the smallest integer such that $\tau_{N_T+1} > T$. We let $Y_n = X_{\tau_n}$ denote the observed values of the state variable. Because we are interested in studying the effect on the estimators of sampling that is not only discrete but also random, we let the sampling intervals $\Delta_n = \tau_n - \tau_{n-1}$ be random variables, in which case the number $N_T + 1$ of observations recorded by time T will also be random.

Although this is not an issue for univariate processes, it is shown in Masry (1978), Masry (1983), and Robinson (1977) that aliasing problems can sometimes be avoided by sampling at random times. A particular feature of this literature,

however, is that inference is often carried out without using the times of observation τ_n of the process. This is also the case for the Duffie and Glynn (1997) method of moments estimator. In the empirical market microstructure literature in finance, it is typical to set up the model to be estimated in discrete time, thereby effectively discarding the sampling intervals (see, e.g., Hasbrouck (1991), Hasbrouck (1996)).

More recently, the literature has focused on the informational content of the times between trades. For example, Engle and Russell (1998) propose a model for the conditional distribution of the time between trades given past information, while Rydberg and Shephard (1999) report different models and empirical evidence on the number of trades recorded in a given time interval. The primary objective of this literature is to build realistic models for the distribution of times between trades in discrete time, whereas we are concerned with understanding the effect that the very existence of such a distribution (as opposed to having nonrandom times between trades) would have on the behavior of estimators of continuous-time models for the price process (2.1), as well as with the interesting issues arising at the interface of the continuous and discrete time domains.

2.2. *The Estimators: FIML, IOML, and PFML*

We consider three possible estimators $\hat{\beta}$ of β. We design these estimators in such a way that each one of them is subject to a specific subset of the costs we wish to measure. The first estimator we consider is the Full Information Maximum Likelihood (FIML) estimator, using the bivariate observations (Y_n, Δ_n); the second is the partial information maximum likelihood estimator using only the state observations Y_n, with the sampling intervals integrated out (IOML for Integrated Out Maximum Likelihood); the third is the pseudo maximum likelihood estimator pretending that the observation times are fixed (PFML for Pretend Fixed Maximum Likelihood). Not surprisingly, the first estimator, FIML, is asymptotically efficient, making the best possible use of the joint data (Y_n, Δ_n). The second estimator, IOML, corresponds to the asymptotically optimal choice if one recognizes that the sampling intervals Δ_n's are random but does not observe them. The third estimator, PFML, corresponds to the "head-in-the-sand" policy consisting of acting as if the sampling intervals are all identical (pretending that $\Delta_n = \overline{\Delta}$ for all n) when in fact they are random.

Both FIML and IOML confront the randomness issue head-on. FIML uses the recorded sampling times; IOML does not, but still recognizes their relevance by integrating them out in the absence of observations on them. Because the data are always discretely sampled, each estimator is subject to the "cost of discreteness," which we define to be the additional variance, or efficiency loss, relative to the variance of an asymptotically efficient estimator based on the full continuous-time sample path. It also represents the error that one would make if one were to use continuous-time asymptotics when the data are in fact discretely sampled.

However, FIML is only subject to the cost of discreteness, while IOML is penalized by both the fact that the data are discrete (the continuous-time sample

path is not observed) and randomly spaced in time (the sampling intervals are not observed). The additional variance of IOML over that of FIML will therefore be identified as the "cost of randomness," or the cost of not observing the randomly-spaced sampling intervals. But if in fact one had recorded the observation times but chosen not to use them in the empirical estimation phase, then what we call the cost of randomness can be interpreted as the cost of throwing away, or not using, these data. Note that this is a lower bound to the cost of throwing away the sampling intervals, since it assumes that one is still following the optimal likelihood strategy in that context.

By contrast, PFML does as if the sampling times were simply not randomly spaced. Comparing it to FIML will give rise to the notion of the "cost of ignoring the randomness." This is the cost imputable to *ignoring* the randomness of the sampling intervals, as opposed to the what we call the cost of randomness, which is the cost due to *not observing* the randomly-spaced sampling intervals. In the former case, one (mistakenly) uses PFML, while in the latter case one realizes that the intervals are informative but, in their absence, IOML is the best one can do. Different types of estimation strategies in empirical market microstructure that do not use the sampling intervals can be viewed as versions of either IOML or PFML, depending upon their treatment of the sampling intervals: throw them away, or ignore their randomness. They will often be suboptimal versions of these estimators because they are subject to an additional efficiency loss if they do not use maximum-likelihood.

All three estimators rely on maximizing a version of the likelihood function of the observations. Let $p(y_1|y_0, \delta, \beta)$ denote the transition function of the process X defined in (2.1), i.e., the density with respect to Lebesgue measure of $X_{t+\delta} = y_1$ given $X_t = y_0$ and given the time interval δ between $X_{t+\delta}$ and X_t and the parameter vector β. Because of the time homogeneity of the model, the transition function p depends only on δ and not on $(t, t+\delta)$ separately. All three estimators make use of some functional of the density p: namely, $p(Y_n|Y_{n-1}, \Delta_n, \beta)$ for FIML; the expectation $\tilde{p}(Y_n|Y_{n-1}, \beta)$ of $p(Y_n|Y_{n-1}, \Delta_n, \beta)$ over the law of $\Delta_n|Y_{n-1}$ for IOML; and $p(Y_n|Y_{n-1}, \overline{\Delta}, \beta)$ for PFML (i.e., like FIML except that $\overline{\Delta}$ is used in place of the actual Δ_n). In practice, even though most diffusion models do not admit closed-form transition densities, all three estimators can be calculated for any diffusion X using arbitrarily accurate closed-form approximations of the transition function p (see Aït-Sahalia (2002)). We also show that $\tilde{p}$ can be obtained in closed form.

2.3. *Asymptotic Properties and Taylor Expansions*

Clearly, if the sampling intervals are in fact fixed at some value $\overline{\Delta}$, then all three estimators would be identical and subject only to the effects of the sampling discreteness. The purpose of this paper is to figure out the relative effects of discrete and random sampling when estimating a continuous time model, which we achieve by comparing the properties of the three estimators FIML, IOML,

and PFML. We will show that each estimator has a probability limit $\bar{\beta}$, and $\sqrt{T}(\hat{\beta}-\bar{\beta})$ converges in law to $N(0, \Omega_\beta)$ as T tends to infinity, with

$$\Omega_\beta = \begin{pmatrix} \omega_\theta & \omega_{\theta\sigma^2} \\ \omega_{\theta\sigma^2} & \omega_{\sigma^2} \end{pmatrix} \tag{2.2}$$

if $\beta = (\theta, \sigma^2)$, $\Omega_\beta = \Omega_\theta$ if $\beta = \theta(\sigma^2 = \sigma_0^2$ known), and $\Omega_\beta = \Omega_{\sigma^2}$ if $\beta = \sigma^2$ ($\theta = \theta_0$ known). While FIML and IOML are always consistent estimators, i.e., $\bar{\beta} = \beta_0$, this is not the case for PFML when the sampling intervals are random.

The asymptotic bias $\bar{\beta} - \beta_0$, when present, and the asymptotic variances of the estimators all involve calculating expectations over the joint distribution of (Δ, Y_1, Y_0) of some functions (which vary by estimator) of these random variables. In general, these expectations cannot be obtained in closed form. They could of course be computed either by numerical integration or by simulation, but comparisons across estimators or sampling schemes and comparative statics would then be essentially infeasible. Our solution is instead to derive an explicit Taylor expansion for each one of these expectations.

For the purpose of calculating these Taylor expansions, let us assume that each sampling interval Δ_n is drawn conditionally upon Y_{n-1} from a common distribution. We write a random variable from that common distribution as $\Delta = \varepsilon\Delta_0$, where ε is deterministic and Δ_0 has a given finite distribution conditional on Y_0. We next figure out, for given fixed ε, the limit as T goes to infinity of the distribution of the estimator. This makes the relevant quantities of the distribution, such as the asymptotic (in T) variance Ω_β and probability limit (again in T) $\bar{\beta}$, functions of ε. To avoid overburdening the notation, we do not indicate this dependence formally as $\Omega_\beta(\varepsilon)$ and $\bar{\beta}(\varepsilon)$, but this is what they are. We then consider Taylor expansions in ε of these deterministic quantities, around $\varepsilon = 0$. In other words, the Taylor expansions have as their leading term the value of the quantity in the limiting case where the full continuous-time sample path is observable, i.e., when $\varepsilon = 0$. By adding higher order terms in ε, we correct this leading term for the discreteness of the sampling. Calculating these Taylor expansions in the case of random sampling required that we develop a new operator, which we call the *generalized infinitesimal generator* of the diffusion X. This is done as part of Theorem 1 below.

2.4. *The Results*

The Taylor expansions lead to the following results. We will show that the asymptotic variance matrices Ω_β can be expanded in ε in the following form:

$$\Omega_\beta = \Omega_\beta^{(0)} + \varepsilon\Omega_\beta^{(1)} + \varepsilon^2\Omega_\beta^{(2)} + O(\varepsilon^3), \tag{2.3}$$

and we will detail the effects of the choice of inference strategy as they show up in the functions $\Omega_\beta^{(i)}$, $i = 0, 1, \ldots$. The first thing that emerges is that for all three estimators, there is no uncertainty in the continuous-sampling limit as far as the estimation of σ^2 is concerned; that is, the order 0 coefficient of the term ω_{σ^2} of

Ω_β is $\omega^{(0)}_{\sigma^2} = 0$. This is another way of saying that the complete continuous-time sample path fully reveals (via the quadratic variation of the X process) the value of σ^2. The uncertainty reappears however as soon as we have discrete sampling ($\varepsilon > 0$) since $\omega^{(1)}_{\sigma^2} > 0$. For the drift, even the entire continuous-time sample path does not fully reveal the parameter θ: we have $\omega^{(0)}_\theta > 0$.

In addition, we will show that the bias of PFML, with the likelihood evaluated at $\bar{\Delta} = E[\Delta]$, is of the form

$$\bar{\beta} - \beta_0 = b_1\varepsilon + b_2\varepsilon^2 + O(\varepsilon^3). \tag{2.4}$$

As could be expected, the bias of PFML disappears in the continuous sampling limit (i.e., the coefficient of order 0 in ε is $b_0 = 0$) but not when the sampling is discrete ($b_1 \neq 0$, or if $b_1 = 0$ then $b_2 \neq 0$, so $\bar{\beta} \neq \beta_0$ when $\varepsilon > 0$).

The two equations (2.3)–(2.4), where the terms $\Omega^{(i)}_\beta$ and b_i each depend upon the distribution of the sampling intervals, can then be used to analyze the effects of any given sampling scheme on the three estimators. Our analysis first documents the efficiency loss that can be attributed to not observing the sampling intervals: we measure this by determining the first order i in ε at which the coefficient $\Omega^{(i)}_\beta$ for IOML differs from the corresponding coefficient $\Omega^{(i)}_\beta$ for FIML, and seeing how much bigger the IOML coefficient is compared to the FIML coefficient. In some cases, this difference will be present at the level of the first nonzero coefficient $\Omega^{(i)}_\beta$. We call the amount by which the variance of IOML exceeds that of FIML the cost of randomness of the sampling process.

We can then compare that efficiency loss to the one attributable to not observing the full continuous-time sample path: we measure the latter by the coefficient at the first order i in ε for which the FIML variance differs from its continuous-time limit $\Omega^{(0)}_\beta$ corresponding to $\varepsilon = 0$. This correction term $\Omega^{(i)}_\beta$ in the FIML variance is what we call the cost of discreteness of the sampling process. An alternative interpretation is that the term $\Omega^{(i)}_\beta\varepsilon^i$ represents the leading term of the error that one would make if one were to use continuous-time asymptotics (i.e., $\Omega^{(0)}_\beta$) instead of the full Ω_β when the data are in fact discretely sampled.

The comparison of the two losses, imputable respectively to the discreteness and the randomness of the sampling, reveals that the loss from not using the observation times will be at least as great, and often substantially greater than the loss from using discrete rather than continuous observations. Specifically, if one estimates the diffusion coefficient σ^2 in the presence of known drift parameters $\theta = \theta_0$, the costs incurred due to the discreteness and the randomness of time both appear in the $\Omega^{(1)}_\beta$ term, i.e., at the same order ε^1, while the limiting variance (the term of order ε^0) is zero if the full continuous time sample path is observed. As one considers different sampling schemes with the same average sampling rate, the loss due to not using sampling times can be an arbitrarily large *multiple* of the loss due to the discreteness of the data. However, if one estimates θ when $\sigma^2 = \sigma^2_0$ is known, then the cost of randomness is an order of magnitude in ε bigger (order ε^1) than the loss from observing the process discretely rather than continuously (order ε^2). However, both are small relative to the sampling

TABLE I
SUMMARY OF RESULTS: FIML VS. IOML

Parameter	Order in ε	FIML (Δ Included) Variance	IOML (Δ Integrated Out) Variance
Drift	0	$\sigma_0^2 E[\Delta_0]\big(E_{\Delta_0 Y_0}[\Delta_0(\partial\mu(Y_0,\theta_0)/\partial\theta)^2]\big)^{-1}$	$\sigma_0^2 E[\Delta_0]\big(E_{\Delta_0 Y_0}[\Delta_0(\partial\mu(Y_0,\theta_0)/\partial\theta)^2]\big)^{-1}$
θ	1	0	nonzero (can be made arbitrarily large)
	2	nonzero	nonzero
Diffusion	0	0	0
σ^2	1	$2\sigma_0^4 E[\Delta_0]$	$2\sigma_0^4 E[\Delta_0]\big(1-\frac{1}{2}E[\mathrm{var}[\chi^2\|\chi^2\Delta_0]]\big)^{-1}$ (can be made arbitrarily large)
	2	nonzero	nonzero

Notes: For each estimator and parameter to be observed, the term of order 0 in ε represents the variance due to the sampling noise that would be present even if the full continuous sample path were observed (i.e., in the limit $\varepsilon = 0$). This variance would be present even if the full continuous sampling path were observed. When estimating σ^2, the continuous-time sample path fully reveals the diffusion parameter; hence this variance is 0. The leading term in the variance is of order 1 in ε. This is not the case however for the drift, where even the complete sample path does not remove all uncertainty regarding θ. Relative to the limiting variance for continuous observation (the term of order 0 in ε), any correction term of higher order in ε in the variance of FIML represents what we call the cost of discreteness of the sampling intervals. Then, because FIML makes the most efficient use of the joint discrete observations on the state variable Y_n and the sampling intervals Δ_n, any further increase in the variance of IOML relative to the variance of FIML represents what we call the cost of randomness of the sampling intervals. It represents the price to pay, in variance terms, for not observing the Δ_n's. The conclusion from the table is that for the drift estimator, the cost of discreteness is an order of magnitude (ε^2) smaller than the cost of randomness (ε^1). For the diffusion estimator, they are both of the same order (ε^1), but the cost of randomness can be an arbitrarily large multiple of the cost of discreteness. For clarity, we do not include in this table the exact expressions for the higher-order terms that are marked "nonzero"; they are given later in the paper. In the IOML first-order variance for the diffusion estimator, χ^2 represents a chi-squared variable with one degree of freedom and independent of Δ_0.

variance that remains present even if the full continuous time sample path is observed (order ε^0). We summarize these results in Table I. The implication of these results for empirical researchers using randomly spaced data is that they should pay as much attention, if not more, to sampling randomness as they do to sampling discreteness.

Turning next to a comparison of PFML to FIML, we can quantify precisely the cost (both in bias and variance) imputable to ignoring the randomness of the sampling intervals, as opposed to what we call the cost of randomness, which is the cost due to not observing the randomly-spaced sampling intervals. In terms of root mean squared error comparisons, any biased estimator such as PFML will always do worse than an unbiased estimator such as FIML or IOML—irrespectively of how large their variances are—since the variance is of order $O(T^{-1})$ whereas the squared bias is of order $O(1)$ as T goes to infinity.

Given this heuristic description of our approach and results, we can now turn to a more precise description of our assumptions and methodology.

3. ASSUMPTIONS AND METHODOLOGY

3.1. *The Distribution of the Sampling Intervals*

To be able to give specific results on the effects of the sampling randomness on the estimation of β, and in particular to be able to compute a partial likelihood with the sampling intervals integrated out, we need to put some structure on

the time series dependence of the sampling intervals $\Delta_n = \tau_n - \tau_{n-1}$. Recall that the process X_t is sampled at times $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \cdots$, and we set $Y_n = X_{\tau_n}$. We assume the following regarding the data generating process for the sampling intervals:

ASSUMPTION 1 (Sampling): *The sampling intervals $\Delta_n = \tau_n - \tau_{n-1}$ are drawn at time τ_{n-1} so as to be conditionally independent of the future of the process X given the past data $(Y_{n-1}, \Delta_{n-1}, Y_{n-2}, \Delta_{n-2}, \ldots, Y_1, \Delta_1)$ as follows*:

(i) *When we study the FIML estimator, we assume that Δ_n is Markovian, i.e.*,

$$L(\Delta_n | Y_{n-1}, \Delta_{n-1}, Y_{n-2}, \Delta_{n-2}, \ldots, Y_1, \Delta_1) = L(\Delta_n | Y_{n-1}, \Delta_{n-1})$$

where L is the density of the listed arguments with respect to a reference measure; furthermore, $L(\Delta_n | Y_{n-1}, \Delta_{n-1})$ is independent of n and of the parameter β.

(ii) *When we study the IOML estimator, we further restrict the dependence structure of the sampling intervals by assuming that $L(\Delta_n | Y_{n-1}, \Delta_{n-1}) = L(\Delta_n | Y_{n-1})$.*

(iii) *When we study the PFML estimator, we add the further restriction that $L(\Delta_n | Y_{n-1}) = L(\Delta_n)$.*

We write $D(\cdot | Y_{n-1}, \Delta_{n-1})$ (respectively d) for the cumulative distribution function (respectively density) of the law of Δ_n given (Y_{n-1}, Δ_{n-1}). Under the further restrictions (ii) and (iii), we write respectively $D(\cdot | Y_{n-1})$ and $D(\cdot)$ and similarly for d. Under Assumption 1(iii), each Δ_n is simply drawn unconditionally from a density $d(\delta) = D'(\delta)$, in which case the sampling intervals are independent and identically distributed. A special case occurs when this density is a Dirac mass at some $\overline{\Delta}$, corresponding to sampling at the fixed deterministic interval $\overline{\Delta}$. But by letting Δ_n be drawn conditionally on Y_{n-1}, as allowed under Assumption 1(ii), and even more so under the even less restrictive Assumption 1(i), we can capture effects such as an increase in trading activity (i.e., smaller Δ_n's on average) after the asset return Y_{n-1} crosses specific boundaries as a result for instance of limit orders or stop-loss policies.

So we can model the dependence of the sampling interval on the just recorded return. For instance, the top graph in Figure 2 reports a kernel regression estimate of the conditional mean $E[\Delta_n | Y_{n-1}]$ for the same complete set of Nokia trades in January 2000 as in Figure 1, which reported the unconditional density of the Δ's. It is apparent from Figure 2 that large price changes in absolute value tend to be followed by shorter wait times to the next price change. We can also trivially extend the Markovian dependence structure to include additional lags. For instance, by letting the distribution of Δ_n depend on Y_{n-1} and further lags Y_{n-2}, Y_{n-3}, etc., we could allow for more sophisticated differential trading intensities as functions of the price path.

The assumption that d is independent of β is equivalent to saying that the sampling interval process is "sequentially exogenous" for the parameter β (see, e.g., Section 1.5.2, Definition 1.11 in Gouriéroux and Monfort (1995) for a definition of sequential exogeneity).[3] This being said, the distribution of the sampling

[3] This assumption can be relaxed: we would then obtain a partial likelihood (see the discussion in footnote 4).

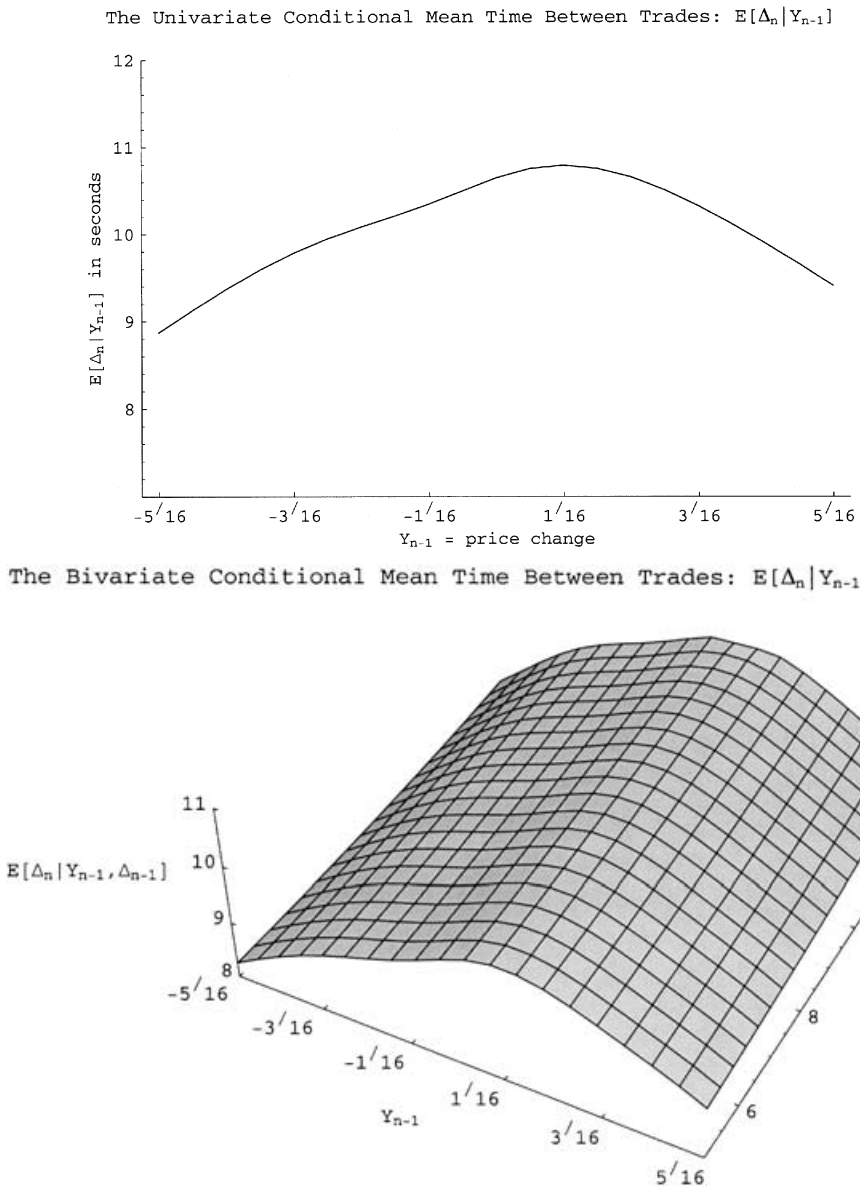


FIGURE 2.—The conditional distribution of sampling intervals for Nokia, January 2000.

intervals may only be known up to some parameters, so that d would involve unknown parameters. A typical example would be exponential arrival times with unknown intensity φ. The intensity parameter φ can also depend on y_0 in general. But we will not be concerned with drawing inference about the nuisance

parameters driving the sampling scheme, such as φ, only about the parameters β of the stochastic differential equation (2.1).

The particular distribution of the sampling scheme, i.e., the specification of d, will have important effects on the various estimators of β we consider, and their relative efficiency. While our results apply to any d, we will give specific examples in the rest of the paper. Specifically, we will consider in Section 5.4 three sampling schemes: (i) deterministic sampling intervals; (ii) uniformly distributed intervals; (iii) exponential arrivals. We will also discuss how to deal with a fourth sampling distribution, where the Δ_n's are generated by two regimes, one with high density ("trading day" or "New York open") and one with low density ("night" or "Singapore then London open"), the distributions of the sampling intervals being deterministic, uniform, or exponential with different expected values, reflecting the different sampling intensities during the trading day and night. Dealing with this fourth sampling scheme involves relaxing the identical distribution assumption for the sampling intervals. This can be done at the cost of complicating the notation. If in fact the Δ_n's do not have a common distribution, our results still go through with $E[\Delta]$ denoting the probability limit (assumed to exist) of $\sum_{i=1}^{n} \Delta_i/n$ as n tends to infinity. If, further, the Δ_n's are not random, but possibly irregularly spaced, then $E[\Delta]$ denotes the Cesaro limit of $\sum_{i=1}^{n} \Delta_i/n$.

But for now let us return to the situation where the Δ_n's are random and drawn from a common distribution D with density d. The restriction in Assumption 1(ii) that $L(\Delta_n|Y_{n-1}, \Delta_{n-1}) = L(\Delta_n|Y_{n-1})$ is essential if one is to compute a tractable partial likelihood, as required for the IOML estimator. The bottom graph in Figure 2 illustrates the content of this assumption, by reporting the bivariate kernel estimator of $E[\Delta_n|Y_n, \Delta_{n-1}]$. The kernel regression shows evidence of variability of Δ_n in the direction of Y_{n-1} (the inverse U shaped pattern whereby large price changes tend to be followed by short wait times to the next trade) of comparable if not larger magnitude than in the direction of Δ_{n-1} (where we find an increase, so that long wait times tend to be followed by another long wait time). Of course, the dependence structure in the full distribution does not reduce to that in the conditional mean, and we are only tracking one asset for one month. But the assumption we are making to study the IOML estimator, namely $L(\Delta_n|Y_{n-1}, \Delta_{n-1}) = L(\Delta_n|Y_{n-1})$, is a compromise between the most general dependence structure and what would be theoretically sufficient to generate our theoretical results (i.i.d. sampling, that is $L(\Delta_n|Y_{n-1}, \Delta_{n-1}) = L(\Delta_n)$), given the need to obtain a tractable partial likelihood. At least on the basis of that limited evidence, this compromise appears to be empirically plausible. As for Assumption 1(iii), it is not essential and we are making it only for ease of exposition of the PFML results.

Let us finally stress that our calculation of the properties of the FIML estimator is valid for a generic dependence structure in $L(\Delta_n|Y_{n-1}, \Delta_{n-1})$, i.e., under Assumption 1(i) only. The progressively more stringent restrictions in Assumption 1(ii) and (iii) are not required for the FIML estimator (more on that in Section 4.1).

Throughout the paper, we denote by Δ a generic random variable with the common distribution of the Δ_n's and, for the purpose of calculating Taylor expansions in ε, write

$$\Delta = \varepsilon \Delta_0 \tag{3.1}$$

where Δ_0 has a finite distribution and ε is deterministic. This Δ_0 is not to be confused with the sequence $\Delta_1, \ldots, \Delta_n$, which have the distribution of Δ, not Δ_0, the difference being the deterministic scaling factor ε. Under Assumption 1(ii), we denote by $d_0(\delta_0|y_0)$ the density that $\Delta_0 = \delta_0$ given that $Y_0 = y_0$. $E[\Delta_0|Y_0]$ denotes the conditional expectation, and $E[\Delta_0]$ the unconditional expectation of Δ_0. Under Assumption 1(iii), we write $d_0(\delta_0)$ for the unconditional density of Δ_0.

Assumption 1(ii), and of course Assumption 1(iii), since it is a special case of Assumption 1(ii), have implications in terms of causality (see Granger (1969) and Sims (1972) for the original definition of causality, and Florens and Fougere (1996) for a treatment in continuous time). In our context, we say that sampling intervals (the Δ's) do not cause future observed states (the Y's) if and only if, for all n, $(Y_k)_{k \geq n}$ is conditionally independent of $(Y_k, \Delta_k)_{k<n}$ given $(Y_k)_{k<n}$. This is indeed the case here:

PROPOSITION 1: *Assume that the solution* $\{X_t; t \geq 0\}$ *of the stochastic differential equation* (2.1) *is strong Markov and unique in law for all initial values* X_0. *Then, under Assumption* 1(ii), *the sampling intervals do not cause future observed states*.

Under assumptions on the stochastic differential equation (2.1) we make below in Section 3.2, the drift function is continuous (infinitely differentiable in fact), hence bounded on compacts; also, the solution of (2.1) is weakly unique. It then follows from Theorem 5.4.20 of Karatzas and Shreve (1991) that X is strong Markov and unique in law for each X_0, thereby satisfying the conditions of Proposition 1.

Note that this noncausality does not hold under the more general sampling generation assumed in Assumption 1(i). Indeed, consider the following example. Suppose that we draw Δ_n conditionally on (Y_{n-1}, Δ_{n-1}) as follows. If $\Delta_{n-1} > 1$, then $\Delta_n = 0$ with probability 1. Otherwise, Δ_n is drawn independently of Δ_{n-1}, for example using an exponential density with conditional mean $E[\Delta_n|Y_{n-1}]$. With this sampling process, knowing that $\Delta_{n-1} > 1$ tells us a lot about Y_n, namely that $Y_n = Y_{n-1}$ with probability one. So in this case

$$L(Y_n|Y_{n-1}, \Delta_{n-1}, Y_{n-2}, \Delta_{n-2}, \ldots, Y_1, \Delta_1) \neq L(Y_n|Y_{n-1}, Y_{n-2}, \ldots, Y_1)$$

and we no longer have noncausality.

3.2. *The Data Generating Process for the State Variable*

We let $\mathcal{S} = (\underline{x}, \bar{x})$ denote the domain of the diffusion X_t. In general, $\mathcal{S} = (-\infty, +\infty)$, but in many examples in finance, we consider the price of an asset with limited liability (stock, foreign currency, bond, etc.) or a nominal interest

rate, in which case $\mathcal{S} = (0, +\infty)$. For arithmetic rates of return, $\mathcal{S} = (-1, +\infty)$. We make the following assumptions:

ASSUMPTION 2 (Differentiability): *The function $\mu(x; \theta)$ is infinitely differentiable in $x \in \mathcal{S}$, and three times continuously differentiable in $\theta \in \Theta$.*

The differentiability assumption in x allows us to compute expansions of the transition density p, as well as expansions of the biases and variances. The differentiability assumption in θ is necessary for the computation of the maximum-likelihood estimator and its asymptotic properties.

ASSUMPTION 3 (Boundary Behavior): *Let* $\kappa(x, \theta, \sigma^2) = -(\mu^2(x; \theta)/\sigma^2 + \partial\mu(x; \theta)/\partial y)/(2\sigma^2)$. *We assume that* $\lim_{x \to \underline{x} \text{ or } x \to \bar{x}} \kappa(x, \theta, \sigma^2) < +\infty$ *for all values of the parameters. In addition, if the left boundary is* $\underline{x} = -\infty$, *then there exist constants* $E > 0$ *and* $K > 0$ *such that for all* $x < -E$ *and* $\theta \in \Theta$, $\mu(x; \theta) \geq Kx$. *If instead* $\underline{x} = 0^+$, *then there exist constants* $E > 0$ *and* $K \geq 0$ *such that for all* $0 < x < E$ *and* $\theta \in \Theta$, $\mu(x; \theta) \geq Kx^{-\alpha}$ *where either* $\alpha > 1$ *and* $K > 0$ *or* $\alpha = 1$ *and* $K \geq 1$. *Finally, for the right boundary* $\bar{x} = +\infty$, *there exist constants* $E > 0$ *and* $K > 0$ *such that for all* $x > E$ *and* $\theta \in \Theta$, $\mu(x; \theta) \leq Kx$.

The constraints on the behavior of the function μ are essentially the best possible. More specifically, if μ has the "wrong" sign near an infinity boundary, i.e., is negative near $-\infty$ or positive near $+\infty$, and grows faster than linearly, then X explodes in finite time. Near the zero boundary 0^+, if there exist $k_0 \geq 0$ and $\alpha_0 < 1$ such that $\mu(x; \theta) < Kx^{-\alpha}$, then 0 and negative values become attainable by X. Assumption 3 implies that the right boundary $\bar{x} = +\infty$ is natural if, in its neighborhood, $|\mu(x; \theta)| \leq K_0 x$, and entrance if $\mu(x; \theta) \leq -Kx^{\alpha}$ for some $\alpha > 1$. If the left boundary is $\underline{x} = -\infty$, then it is natural if, in its neighborhood, $|\mu(x; \theta)| \leq Kx$, and entrance if $\mu(x; \theta) \geq -K|x|^{\alpha}$ for some $\alpha > 1$. If the left boundary is $\underline{x} = 0^+$, then it is an entrance boundary. Both natural entrance boundaries are unattainable (see, e.g., Section 15.6 in Karlin and Taylor (1981)) so that the exit time from $\mathcal{S}$ will be infinite, namely $\Pr(T_{\mathcal{S}} = +\infty) = 1$ where $T_{\mathcal{S}} = \inf\{t \geq 0 | X_t \notin \mathcal{S}\}$. Natural boundaries can neither be reached in finite time nor can the diffusion be started there. Entrance boundaries such as 0^+, cannot be reached starting from an interior point in $\mathcal{S} = (0, +\infty)$, but it is possible for X to start there, in which case the process moves quickly away from 0 and never returns. We take μ to be such that the process is stationary and hence ergodic:

ASSUMPTION 4 (Stationarity): *Assume that* $\exp\{2\sigma^{-2} \int^x \mu(y; \theta)\, dy\}$ *is integrable in x at both boundaries of* $\mathcal{S}$. *Then let*

$$\pi(x, \beta) = \frac{\exp\{2\sigma^{-2} \int^x \mu(y; \theta)\, dy\}}{\int_{\underline{x}}^{\bar{x}} \exp\{2\sigma^{-2} \int^x \mu(y; \theta)\, dy\}\, dx} \tag{3.2}$$

and assume that the initial value of the process, X_0, *has density* π.

Under the integrability assumed in Assumption 4, $\pi(x, \beta)$ is the stationary density of the process X, i.e., the common marginal density of each X_t. In the financial examples that motivate our analysis, the stationarity assumption on X means that we will consider X to be the process for (instantaneous) continuously compounded asset returns or interest rates rather than asset prices.

Note that when both boundaries of the process are entrance boundaries, then the integrability condition in Assumption 4 is automatically satisfied. When at least one of the boundaries is natural, the integrability condition is neither precluded nor implied. But in any event, under Assumptions 2–3, the stochastic differential equation (2.1) admits a weak solution that is unique in probability law. The solution process $\{X_t, t \geq 0\}$ admits a transition density $p(y_1|y_0, \delta, \beta)$, which is continuously differentiable in δ, infinitely differentiable in y_1 and y_0 on $\mathcal{S}$, and three times continuously differentiable in β (see Aït-Sahalia (2002)).

For inference purposes, we assume the following.

ASSUMPTION 5 (Identification): *$\mu(x; \theta) = \mu(x; \tilde{\theta})$ for π-almost all x in $\mathcal{S}$ implies $\theta = \tilde{\theta}$.*

Finally, note that we might have started with the model

$$dX_t = \mu(X_t; \theta)\, dt + \sigma(X_t)\, dW_t \tag{3.3}$$

where $\sigma(\cdot)$ is any known positive function. But if we define

$$Y_t = \int^{X_t} \frac{1}{\sigma(x)}\, dx, \tag{3.4}$$

then Itô's Lemma applied to (3.3)–(3.4) allows us to reduce the analysis to the case (2.1) with $\sigma = 1$.

3.3. *Taylor Expansions for Expectations*

As will become apparent below, the asymptotic properties of the estimators (bias for PFML, and variance for all three) depend upon expectations of matrices of the form $E_{\Delta, Y_1, Y_0}[f(Y_1, Y_0, \Delta, \beta, \varepsilon)]$. Throughout the paper, we denote by E_{Δ, Y_1, Y_0} expectations taken with respect to the joint law of (Δ, Y_1, Y_0), and write E_{Δ, Y_1}, etc., for expectations taken from the appropriate marginal laws of (Δ, Y_1), etc. All these expectations are taken with respect to the true distribution of the process. We denote by $\{U_\delta; \delta > 0\}$ the semigroup of conditional expectation operators associated with the process X, defined for all f in L^2 (where L^2 refers to integration with respect to the y_1 variable, weighted by the stationary density $\pi(y_1; \beta_0)$) by its action

$$U_\delta \cdot f(y_0, \delta, \beta, \varepsilon) = E_{Y_1}[f(Y_1, Y_0, \Delta, \beta, \varepsilon)|Y_0 = y_0, \Delta = \delta]. \tag{3.5}$$

In what follows, the functions f will be the log-likelihood, or its derivatives, evaluated at different sampling intervals, or at the expected value of the sampling

interval. Because these expectations are not generally available in closed-form, our approach is based on calculating Taylor expansions in ε of these matrices—which happen to all be fully explicit.

To introduce the approach, let us first suppose for a moment that the sampling intervals are deterministic. In this case, Taylor expansions are intimately related to the infinitesimal generator A_{β_0} of the diffusion X, i.e., the operator that returns

$$A_{\beta_0} \cdot f(y_1, y_0, \delta, \beta, \varepsilon) = \frac{\partial f(y_1, y_0, \delta, \beta, \varepsilon)}{\partial \delta} + \mu(y_1, \beta_0) \frac{\partial f(y_1, y_0, \delta, \beta, \varepsilon)}{\partial y_1} + \frac{1}{2}\sigma_0^2 \frac{\partial^2 f(y_1, y_0, \delta, \beta, \varepsilon)}{\partial y_1^2}. \tag{3.6}$$

The domain of A_{β_0} is the set of functions f in L^2 that are continuously differentiable in δ (once) and in y_1 (twice) and for which the limit in the sense of L^2

$$\lim_{\delta \to 0} \frac{1}{\delta}\left[U_\delta \cdot f(y_0, \delta, \beta, \varepsilon) - f(y_0, \delta, \beta, \varepsilon)\right] = A_{\beta_0} \cdot f(y_0, y_0, 0, \beta_0, 0) \tag{3.7}$$

exists. A sufficient condition for the existence of the limit is given on page 10 of Hansen, Scheinkman, and Touzi (1998). The limit exists provided that $\partial f / \partial y_1$ and $A_{\beta_0} \cdot f$ are both also in L^2 and

$$\lim_{y_1 \to \underline{x}} \frac{\partial f / \partial y_1}{s(y_1; \beta)} = \lim_{y \to \bar{x}} \frac{\partial f / \partial y_1}{s(y_1; \beta)} = 0 \tag{3.8}$$

where s denotes the scale density of the process, defined as

$$s(x; \beta) \equiv \exp\left\{-2\sigma^{-2} \int^x \mu(y; \theta)\, dy\right\} \tag{3.9}$$

for a given lower bound of integration whose choice is irrelevant. Then we define $\mathcal{D}$ to be the set of functions f that are in the domain of the standard infinitesimal generator and also continuously differentiable in β and ε.

If the function f does not explicitly depend on β and ε, we can evaluate

$$U_\delta \cdot f(y_0, \delta, \beta, \varepsilon) = E_{Y_1}[f(Y_1, Y_0, \Delta) | Y_0 = y_0, \Delta = \delta],$$

which, by a direct application of the deterministic Taylor formula around $\delta = 0$, has the expansion

$$U_\delta \cdot f(y_0, \delta) = \sum_{j=0}^{J} \frac{\delta^j}{j!} A_{\beta_0}^j \cdot f(y_1, y_0, \delta)_{|y_1 = y_0,\, \delta = 0} + O(\delta^{J+1}) \tag{3.10}$$

for all f in the set of functions with $J+2$ continuous derivatives in δ, $2(J+2)$ in y_1, such that f and its first $J+1$ iterates by repeated applications of A_{β_0}

(that is the functions $A^j_{\beta_0} \cdot f$, $j = 1, 2, \ldots, J+1$) all remain in $\mathcal{D}$. Let $\mathcal{D}^J$ denote the set of functions f that satisfy this property and additionally have $J+1$ continuous derivatives in β and ε. Note that (3.10) is not surprising if (3.7) is viewed as generating U_δ from A_{β_0} (instead of the other way around); in that case, one can think of (3.7) in the exponential formula form $U_\delta = \exp(A_{\beta_0}\delta)$.

Note that in general the Taylor series on the right-hand side of (3.10) need not converge as J goes to infinity, as $U_\delta \cdot f$ need not be analytic in δ at $\delta = 0$. There are sufficient conditions, however, that will insure this: see Proposition 4 in Aït-Sahalia (2002). In the present context, we need not worry about the convergence of the series as $J \to \infty$. We are merely interpreting the truncated series at some J as a (local) Taylor expansion near $\delta = 0$ and, for that, analyticity of the function being Taylor-expanded is not necessary.

As we will show in Theorem 1 below, the analogous expansion for $E_{Y_1}[f(Y_1, Y_0, \Delta, \beta, \varepsilon)|Y_0, \Delta]$ when the sampling intervals are random and f can depend on (β, ε) in addition to (Y_1, Y_0, Δ), involves replacing the operator A_{β_0} by a generalized version of it—an operator Γ_{β_0} which we call the *generalized infinitesimal generator* of X. Given an expansion for the condition expectation $E_{Y_1}[f(Y_1, Y_0, \Delta, \beta, \varepsilon)|Y_0, \Delta]$, we can then calculate Taylor expansions for the unconditional expectation of f by applying the law of iterated expectations first to

$$E_{\Delta, Y_1, Y_0}[f(Y_1, Y_0, \Delta, \beta, \varepsilon)] = E_\Delta\big[E_{Y_1, Y_0}[f(Y_1, Y_0, \Delta, \beta, \varepsilon)|\Delta]\big]$$

and then to $E_{Y_1, Y_0}[f(Y_1, Y_0, \Delta, \beta, \varepsilon)|\Delta] = E_{Y_0}[E_{Y_1}[f(Y_1, Y_0, \Delta, \beta, \varepsilon)|Y_0, \Delta]]$.

What we will show in later theorems is that the asymptotic distributions of biased estimators such as PFML require us to evaluate these expectations not at $\beta = \beta_0$ but at some $\beta = \bar{\beta}$ such that its limit as ε tends to 0 is β_0. $\bar{\beta}$ will represent the probability limit of the estimator calculated from discrete data. For now, we will write a general result allowing for the possibility that β depends on ε, in which case it satisfies $\bar{\beta} \to \beta_0$ as $\varepsilon \to 0$. If $\beta = \beta_0$ always (i.e., not just when $\varepsilon \to 0$), then we have an obvious special case.

THEOREM 1 (Taylor Expansions for Expectations): *Define the generalized infinitesimal operator Γ_{β_0} for the process X in* (2.1) *as the operator that applies to functions f in $\mathcal{D}$ and returns*

$$\Gamma_{\beta_0} \cdot f(y_1, y_0, \delta, \beta, \varepsilon) = \Delta_0 A_{\beta_0} \cdot f(y_1, y_0, \delta, \beta, \varepsilon) + \frac{\partial f}{\partial \varepsilon}(y_1, y_0, \delta, \beta, \varepsilon) + \frac{\partial f}{\partial \beta}(y_1, y_0, \delta, \beta, \varepsilon)\frac{\partial \beta}{\partial \varepsilon}(\beta, \varepsilon). \tag{3.11}$$

Note that the operator Γ_{β_0} is random in that it depends on Δ_0. Then under Assumptions 1–4 *if the function f is in $\mathcal{D}^J$, a Taylor expansion in ε around $\varepsilon = 0$ of the conditional expectation of f, evaluated at $\beta = \bar{\beta}$, with respect to the law*

of $Y_1|(Y_0, \Delta)$ *is given by*:

$$E_{Y_1}\left[f(Y_1, Y_0, \Delta, \bar{\beta}, \varepsilon)|Y_0, \Delta\right] = \sum_{j=0}^{J} \frac{\varepsilon^j}{j!} \Gamma_{\beta_0}^j \cdot f(Y_1, Y_0, \Delta, \beta, \varepsilon)_{|Y_1=Y_0,\, \Delta=\varepsilon=0,\, \beta=\beta_0} + O_p(\varepsilon^{J+1}). \tag{3.12}$$

The unconditional expectation with respect to the joint law of (Δ, Y_1, Y_0) *is given by*:

$$E_{\Delta, Y_1, Y_0}[f(Y_1, Y_0, \Delta, \bar{\beta}, \varepsilon)] = \sum_{j=0}^{J} \frac{\varepsilon^j}{j!} E_{\Delta, Y_0}\left[\Gamma_{\beta_0}^j \cdot f(Y_1, Y_0, \Delta, \beta, \varepsilon)_{|Y_1=Y_0,\, \Delta=\varepsilon=0,\, \beta=\beta_0}\right] + O(\varepsilon^{J+1}). \tag{3.13}$$

When we specify an estimation strategy and a sampling scheme, we will be presented with a variety of functions f to which Theorem 1 can be applied. We now examine the various estimation strategies and sampling schemes.

4. LIKELIHOOD-BASED INFERENCE MAKING DIFFERENT USE OF THE SAMPLING INTERVALS

4.1. *FIML: Full-Information Likelihood Inference*

The first estimator involves using both sets of data (Y_n, Δ_n) and maximizing their joint likelihood. Recall that we write $p(y_1|y_0, \delta, \beta)$ for the transition function of the process X, i.e., the density that $Y_n = y_1$ given that $Y_{n-1} = y_0$ and that $\Delta_n = \delta$. We also defined $\pi(x, \beta)$ to be the common marginal density of X_t for all t. Let us also define $p(y_1, y_0, \delta, \beta) = p(y_1|y_0, \delta, \beta)\pi(y_0, \beta)$ to be the joint density of the $(X_{t+\delta}, X_t)$. Recall also that $d(\delta_1|y_0)$ denotes the conditional density of the sampling interval, i.e., the density that $\Delta_n = \delta_1$ given $Y_{n-1} = y_0$. We now wish to write the likelihood function for the observations $\{(Y_n, \Delta_n) : n = 1, \ldots, N_T\}$. By Bayes' Rule and the assumed Markovian structure of the (Y_n, Δ_n)'s we have

$$\begin{aligned} &L(Y_{N_T}, \Delta_{N_T}, \ldots, Y_n, \Delta_n, Y_{n-1}, \Delta_{n-1}, \ldots, Y_1, \Delta_1|X_0; \beta) \\ &= \prod_{n=1}^{N_T} L(Y_n, \Delta_n|Y_{n-1}, \Delta_{n-1}, \ldots, Y_1, \Delta_1, X_0; \beta) \\ &= \prod_{n=1}^{N_T} L(Y_n, \Delta_n|Y_{n-1}, \Delta_{n-1}; \beta) \end{aligned}$$

where the notation L stands for the density of the listed arguments. As is usual in likelihood inference for Markov processes, we condition on the first observation X_0.

One further application of Bayes' Rule yields

$$L(Y_n, \Delta_n|Y_{n-1}, \Delta_{n-1}; \beta) = L(Y_n|Y_{n-1}, \Delta_n, \Delta_{n-1}; \beta) \times L(\Delta_n|Y_{n-1}, \Delta_{n-1}).$$

Then Assumption 1 implies that

$$L(Y_n|Y_{n-1}, \Delta_n, \Delta_{n-1}; \beta) = L(Y_n|Y_{n-1}, \Delta_n; \beta) = p(Y_n|Y_{n-1}, \Delta_n, \beta)$$

so that the likelihood function using all the available information is given by

$$\begin{aligned} (4.1) \qquad & L(Y_{N_T}, \Delta_{N_T}, \ldots, Y_n, \Delta_n, Y_{n-1}, \Delta_{n-1}, \ldots, Y_1, \Delta_1|X_0; \beta) \\ & = \prod_{n=1}^{N_T} p(Y_n|Y_{n-1}, \Delta_n, \beta) L(\Delta_n|Y_{n-1}, \Delta_{n-1}) \end{aligned}$$

and the corresponding log-likelihood function using all the available information is

$$(4.2) \qquad \sum_{n=1}^{N_T-1} l(Y_n|Y_{n-1}, \Delta_n, \beta) + \sum_{n=1}^{N_T-1} c(\Delta_n|Y_{n-1}, \Delta_{n-1})$$

where

$$(4.3) \qquad l(y_1|y_0, \delta, \beta) \equiv \ln(p(y_1|y_0, \delta, \beta))$$

and $c(\delta_1|y_0, \delta_0) \equiv \ln(L(\delta_1|y_0, \delta_0))$.

Recall now that we assumed that the conditional sampling density $L(\delta_1|y_0, \delta_0)$ may have its own (nuisance) parameters, such as an unknown arrival rate, but does not depend on the parameters β of the state process.[4] Since we only care about β, we will then define

$$(4.4) \qquad l_T(\beta) = \sum_{n=1}^{N_T-1} l(Y_n|Y_{n-1}, \Delta_n, \beta)$$

and the FIML estimator $\hat{\beta}$ as a solution of $\partial l_T(\beta)/\partial\beta = 0$. We will use the notation $\dot{f}$ to denote differentiation of the function f with respect to the parameters, i.e., for instance $\dot{l} = \partial l/\partial\beta$ and $\ddot{l} = \partial^2 l/\partial\beta\partial\beta'$.

It is clear from the calculations that precede that the only part of Assumption 1 that we used so far is the Markovian nature of the sampling intervals and the fact that the distribution of $L(\Delta_n|Y_{n-1}, \Delta_{n-1})$ is independent of β, i.e., Assumption 1(i). As discussed in Section 3.1, the further restriction contained in Assumption 1(ii) and the noncausality it implies (recall Proposition 1) are not necessary for the study of the FIML estimator. Assumption 1(ii) would simply add the fact that $L(\Delta_n|Y_{n-1}, \Delta_{n-1}) = L(\Delta_n|Y_{n-1}) = d(\Delta_n|Y_{n-1})$, but that part of the likelihood (the function c) is already irrelevant in (4.2). Of course, the further

[4] An extension of our theory would cover the case where $d(\delta_1|y_0)$ depends on β; in this case, one could take the first term in (4.2) to be a partial likelihood (see, e.g., Cox (1975) and Wong (1986)). This is still covered by our general theory.

restriction contained in Assumption 1(iii) is even more unnecessary for the study of FIML.

However, when we turn to IOML and the integration of the full information likelihood (4.1) over the Δ_n's, then Assumption 1(ii) implies that for each n the corresponding Δ_n appears only in the single product $p(Y_n|Y_{n-1}, \Delta_n, \beta)$ $L(\Delta_n|Y_{n-1})$ and the multidimensional integration over the Δ_n's can be performed for each one of these terms at a time, resulting in a product of one dimensional integrals. In fact, we will show below how this integration can be performed in closed form.

Returning now to FIML, the function l, hence the criterion function $l_T(\beta)$, can be evaluated in closed-form using the expansion of the transition density $p(Y_1|Y_0, \Delta, \beta)$ of an arbitrary diffusion X given in Aït-Sahalia (2002). With the correction terms gathered in increasing order of Δ, the Hermite-based closed-form expansion is

$$p^{(J)}(Y_1|Y_0, \Delta, \beta) = \frac{1}{\sqrt{2\pi\Delta}\sigma} \exp\left(-\frac{(Y_1 - Y_0)^2}{2\Delta\sigma^2} + \int_{Y_0}^{Y_1} \frac{\mu(w, \theta)}{\sigma^2}\, dw\right) \times \sum_{j=0}^{J} c_j\left(\frac{Y_1}{\sigma}\middle|\frac{Y_0}{\sigma}, \beta\right)\frac{\Delta^j}{j!} \tag{4.5}$$

where the coefficient functions c_j are given recursively by $c_0 = 1$ and

$$c_{j+1}(z_1|z_0, \beta) = j(z_1 - z_0)^{-j}\int_{z_0}^{z_1}(w - z_0)^{j-1} \times \left\{\kappa(w, \theta, \sigma^2)c_{j-1}(w|z_0; \theta) + (1/2)\partial^2 c_{j-1}(w|z_0; \theta)/\partial w^2\right\} dw \tag{4.6}$$

for $j \geq 1$, with $\kappa(w, \theta, \sigma^2) = -(\mu^2(w; \theta) + \partial\mu(w; \theta)/\partial w)/2$. The function $l(y_1|y_0, \delta, \beta)$ can also be expressed as a series in Δ by Taylor-expanding $\ln(\sum_{j=0}^{J} c_j(Y_1/\sigma|Y_0/\sigma, \beta)(\Delta^j/j!))$ in Δ. With $l^{(J)} = \ln p^{(J)}$, we have

$$\left|l(y_1|y_0, \beta, \varepsilon\delta_0) - l^{(J)}(y_1|y_0, \beta, \varepsilon\delta_0)\right| = O(\varepsilon^J) \tag{4.7}$$

and similarly for their first three partial derivatives with respect to β, uniformly.

In order to be able to carry maximum likelihood inference about β, note that Assumption 5 yields that

$$0 < E_{Y_1, Y_0}\left[\dot{l}(y|x, \delta, \beta)\dot{l}(y|x, \delta, \beta)'\right] < \infty.$$

This means that none of the parameters are redundant. From standard likelihood considerations (see, e.g., Billingsley (1961)) and the regularity properties of the log-likelihood function given in Aït-Sahalia (2002), we obtain that $\sqrt{N_T}(\hat{\beta} - \beta_0)$ is asymptotically normal with information $E_{Y_1, Y_0}[-\ddot{l}(Y_1|Y_0, \Delta, \beta_0)]$, i.e., that

$$\sqrt{T}(\hat{\beta} - \beta_0) \underset{T\to\infty}{\longrightarrow} N(0, V_\beta) \tag{4.8}$$

in distribution, where

$$V_\beta^{-1} = (E[\Delta])^{-1} E_{\Delta, Y_1, Y_0}\left[-\ddot{l}(Y_1|Y_0, \Delta, \beta_0)\right] \tag{4.9}$$

since T/N_T converges in probability to $E[\Delta]$.[5]

V_β is the FIML version of what is denoted by Ω_β in (2.2). We use a special symbol in this case so as to be able to refer later to the FIML asymptotic variance separately from the symbol Ω_β which we use as generic notation for any one estimator. By applying Theorem 1 to the particular function f given by $-\ddot{l}(Y_1|Y_0, \Delta, \beta_0)$, we will be able to derive the Taylor expansion of V_β: see Section 5.

4.2. *IOML: Partial Information Likelihood Inference Integrating Out the Observation Times*

Suppose now that the observation intervals Δ_n are either not observable or discarded prior to conducting inference on β. When the Δ_n's are, in fact, random, and we are willing to account for their randomness as described in Assumption 1(ii), one can still integrate them out before doing the analysis. This results in a proper likelihood framework for inference on β based on using the Y_n data only provided that inference is based on the IOML likelihood ratio $\widetilde{LR}$:

$$\widetilde{LR}(Y_n|Y_{n-1}) = E_{\Delta_n}[LR(Y_n|Y_{n-1}, \Delta_n)|Y_n, Y_{n-1}] \tag{4.10}$$

where LR is the FIML likelihood ratio

$$LR(Y_n|Y_{n-1}, \Delta_n) \equiv \frac{p(Y_n|Y_{n-1}, \Delta_n, \beta)}{p(Y_n|Y_{n-1}, \Delta_n, \beta_0)}. \tag{4.11}$$

PROPOSITION 2: *The IOML likelihood ratio $\widetilde{LR}$ is given by*

$$\widetilde{LR}(y_1|y_0) = \frac{\tilde{p}(y_1|y_0, \beta)}{\tilde{p}(y_1|y_0, \beta_0)} \tag{4.12}$$

where

$$\tilde{p}(y_1|y_0, \beta) = E_\Delta[p(y_1|Y_0, \Delta, \beta)|Y_0 = y_0]. \tag{4.13}$$

As the notation suggests, $\tilde{p}(Y_1|Y_0, \beta)$ is the conditional density of Y_1 given Y_0.

[5] When scaling the difference $\hat{\beta} - \beta_0$, we use throughout the square root of the length of time T as opposed to the number of observations N_T actually recorded between time 0 and time T. When the sampling intervals are random, N_T is also random. However, T/N_T tends in probability to $E[\Delta]$, the unconditional expected value of the sampling intervals so that the only difference involved between (deterministic) scaling by $\sqrt{T}$ and (random) scaling by $\sqrt{N_T}$ is a factor $E[\Delta]$ in the asymptotic variance.

4.2.1. *Calculating the Integrated-Out Likelihood Function*

Among other examples, when Δ has an exponential distribution, the integration in (4.13) can be done in closed-form for the (expansion of the) transition density p given in equation (4.5). The fact that an expansion of the integrated-out conditional density $\tilde{p}$ is available in closed-form makes the computation of the IOML estimator just as straightforward as that of FIML. We have the following proposition.

PROPOSITION 3: *If the density d of Δ given Y_0 is the exponential density with conditional mean $E[\Delta|Y_0]$, that is,*

$$d(\Delta|Y_0) = \exp(-\Delta/E[\Delta|Y_0])/E[\Delta|Y_0], \tag{4.14}$$

then the corresponding expansion of order J for the density $\tilde{p}$, that is,

$$\tilde{p}^{(J)}(Y_1|Y_0,\beta) = \int_0^{+\infty} p^{(J)}(Y_1|Y_0,\delta,\beta)\, d(\delta|Y_0)\, d\delta,$$

is also explicit and is given by

$$\begin{aligned}\tilde{p}^{(J)}(Y_1|Y_0,\beta) = \exp\left(\int_{Y_0}^{Y_1} \frac{\mu(w,\theta)}{\sigma^2}\, dw\right) \\ \times \left\{\sum_{j=0}^{J} c_j\left(\frac{Y_1}{\sigma}\Big|\frac{Y_0}{\sigma},\beta\right) \times B_{j+(1/2)}\left(\frac{\sqrt{2}|Y_1-Y_0|}{\sqrt{E}[\Delta|Y_0]\sigma}\right)\right. \\ \left. \times |Y_1-Y_0|^{(2j+1)/2} \frac{2^{(1-2j)/4} E[\Delta|Y_0]^{(2j-3)/4}}{j!\sqrt{\pi}\sigma^{j+(3/2)}}\right\}\end{aligned} \tag{4.15}$$

where $B_{j+(1/2)}(z)$ is the Bessel K function of half-integer order $j+(1/2)$.

The Bessel K functions of half-integer order appearing in Proposition 3 are in closed-form for any j; hence (4.15) is indeed a closed-form expression. For instance, the first three $B_{j+(1/2)}$ functions are given by

$$B_{1/2}(z) = \exp(-z)\sqrt{\frac{\pi}{2z}}, \qquad B_{3/2}(z) = \exp(-z)\sqrt{\frac{\pi}{2z}}\left(1+\frac{1}{z}\right),$$

$$B_{5/2}(z) = \exp(-z)\sqrt{\frac{\pi}{2z}}\left(1+\frac{3}{z}+\frac{3}{z^2}\right).$$

4.2.2. *Inference with the Integrated-Out Likelihood Function*

Coming back to the inference problem for IOML, instead of (4.4), the part of the log-likelihood function dependent on β is now

$$\lambda_T(\beta) = \sum_{n=1}^{N_T-1} \lambda(Y_n|Y_{n-1},\beta) \tag{4.16}$$

where

$$\lambda(y_1|y_0,\beta) = \ln(\tilde{p}(y_1|y_0,\beta)) \tag{4.17}$$

is the log of the integrated-out density. We call $\tilde{\beta}$ the IOML estimator obtained as a solution of $\partial\lambda_T(\beta)/\partial\beta = 0$. As in the FIML case of Section 4.1, $\sqrt{T}(\tilde{\beta}-\beta_0)$ converges in law to $N(0,\widetilde{V}_\beta)$, but now

$$\widetilde{V}_\beta^{-1} = (E[\Delta])^{-1}E_{Y_1,Y_0}[-\ddot{\lambda}(Y_1|Y_0,\beta_0)] \tag{4.18}$$

where as always E_{Y_1,Y_0} denotes an expectation taken with respect to the joint stationary density of (Y_1, Y_0). Computing a Taylor expansion for $\widetilde{V}_\beta$ involves calculations that are substantially more involved than those for V_β, although they still follow from Theorem 1 applied to the particular function f given by $-\ddot{\lambda}(Y_1|Y_0,\beta_0)$.

4.2.3. *Relationship Between the Full and Partial Information Likelihoods*

To determine the relationship between the partial and full information inference strategies, we need to express the integrated out likelihood function (4.16) in terms of the full information one, (4.4). The result is Proposition 4.

PROPOSITION 4 (FIML and IOML Likelihoods): *The score of the integrated-out likelihood is given by*

$$\dot{\lambda}(y_1|y_0,\beta_0) = E_\Delta\big[\dot{l}(Y_1|Y_0,\Delta,\beta_0) \mid Y_0 = y_0, Y_1 = y_1\big] \tag{4.19}$$

while the information follows from

$$\begin{aligned}\bar{\lambda}(y_1|y_0,\beta_0) = {} & E_\Delta\big[\ddot{l}(Y_1|Y_0,\Delta,\beta_0)|Y_0 = y_0, Y_1 = y\big] \\ & + \mathrm{var}_\Delta\big[\dot{l}(Y_1|Y_0,\Delta,\beta_0)|Y_0 = y_0, Y_1 = y_1\big].\end{aligned} \tag{4.20}$$

These two equations are conditional versions of the Bartlett or "exlog" identities (see Bartlett (1953), or Barndorff-Nielsen and Cox (1989)).[6] Equation (4.20) is a critical element in our analysis of the differences between the two estimators FIML and IOML in light of (4.18). Namely, (4.20), jointly with (4.9) and (4.18), imply that

$$\widetilde{V}_\beta^{-1} = V_\beta^{-1} - (E[\Delta])^{-1}E_{Y_1,Y_0}\big[\mathrm{var}_\Delta[\dot{l}(Y_1|Y_0,\Delta,\beta_0)|Y_0,Y_1]\big] \tag{4.21}$$

from which the asymptotic information loss from not using the observation times, $E[\Delta](V_\beta^{-1} - \widetilde{V}_\beta^{-1})$, can be read. Not surprisingly, it follows immediately from

[6] For a different use of these identities in the econometric literature, see Gouriéroux et al. (1987).

(4.21) that $\widetilde{V}_\beta - V_\beta$ is positive definite; in other words, FIML is more efficient than IOML. The interesting issue is of course how much more efficient FIML is, i.e., how large is the loss of information from FIML to IOML, and also how that depends on the particular distribution of the sampling intervals. We will address these two questions in Section 5 using Taylor expansions of the two asymptotic variance matrices and relating them to moments of the distribution of Δ_0.

4.3. *PFML: Pseudo-Likelihood Inference Pretending that the Observation Times are Fixed*

A third inference strategy consists in ignoring the variations in the sampling intervals and fixing them at a common value. For instance, in the presence of daily data, one could ignore the presence of weekends, holidays, etc., and fix the sampling interval throughout the sample at $\overline{\Delta} = 1$ day. Or the sampling intervals could be truly random, as in transaction data, but one decides to act as if the observations were all equally spaced at a given sampling interval $\overline{\Delta} = E[\Delta] = \varepsilon E[\Delta_0]$—for example assuming that transaction data are sampled at 30 second intervals, or whatever the average time between observations is. The reason for considering this inference scheme, which would be asymptotically optimal if the sampling intervals were in fact fixed, is to measure the cost, in terms of bias and variance, of ignoring the randomness of the Δ_n's. This cost is different from the cost represented by the increase in variance between IOML and FIML, which measures the cost of not observing the sampling intervals while being aware of their randomness.

Another interpretation of PFML relates it to the common practice of setting up the theoretical model in a continuous-time model, but estimating a discrete time version. If the model to be estimated were specified in discrete time, as is most often the case in the empirical microstructure literature, then by design one is pretending that the sampling intervals are identical at $\overline{\Delta}$, with data $Y_n = X_{n\overline{\Delta}}, n = 0, 1, \ldots, N$. By setting up the model to be estimated in discrete time, as long as $\overline{\Delta} > 0$, one introduces an additional bias, due to the discretization, that would be present even if the data were in fact equally spaced in time. PFML would then correct for the discretization bias, but not the randomness bias should the data be in fact randomly spaced in time.

We are being generous with regard to the PFML estimation strategy by letting the fixed value at which the likelihood is evaluated be the true unconditional mean of the sampling interval $E[\Delta]$: by generous, we mean that the unconditional mean is indeed the best fixed value that one can choose (see Proposition 6 below). We are also assuming that no specification error is made other than ignoring the randomness of the sampling intervals. In particular, the transition function p is the same correct density as in the FIML case.

So this inference scheme is based on maximizing

$$\check{l}_T(\beta) = \sum_{n=1}^{N_T - 1} l(Y_n | Y_{n-1}, \overline{\Delta}, \beta) \tag{4.22}$$

where l is the full-information likelihood defined in Section 4.1 above. The PFML estimator $\check{\beta}$ is a solution of $\partial \check{l}_T(\beta)/\partial\beta = 0$. In this case, its asymptotic distribution follows from using the pseudo-score vector

$$h(y_1, y_0, \beta, \varepsilon) = \begin{bmatrix} h_1(y_1, y_0, \beta, \varepsilon) \\ h_2(y_1, y_0, \beta, \varepsilon) \end{bmatrix} = \begin{bmatrix} \dot{l}_\theta(y_1|y_0, \varepsilon E[\Delta_0], \beta) \\ \dot{l}_{\sigma^2}(y_1|y_0, \varepsilon E[\Delta_0], \beta) \end{bmatrix}. \tag{4.23}$$

The pseudo-scores given by the two components of the vector h are not in general martingales. In fact, they do not even have mean zero unconditionally. As a result of $E_{\Delta, Y_1, Y_0}[h(Y_1, Y_0, \beta_0, \varepsilon)] \neq 0$, the estimator is not consistent. Instead, the probability limit of $\hat{\beta}$ is the solution $\bar{\beta}$ of the equations

$$E_{\Delta, Y_1, Y_0}[h(Y_1, Y_0, \bar{\beta}, \varepsilon)] = 0. \tag{4.24}$$

In the limit where sampling is continuous ($\varepsilon \to 0$), this sampling scheme becomes FIML. Therefore, we should expect that the asymptotic bias of the estimator, $\bar{\beta} - \beta_0$, will be of order greater or equal to 1 in ε, that is: $\bar{\beta} - \beta_0 = b_1\varepsilon + b_2\varepsilon^2 + O(\varepsilon^3)$. We will show that this is indeed the case: the biases $\bar{\theta} - \theta_0$ and $\bar{\sigma}^2 - \sigma_0^2$ are both of order $O(\varepsilon)$, except if σ^2 is known, in which case the bias $\bar{\theta} - \theta_0$ is of order $O(\varepsilon^2)$. Thus, if one is prepared to live with a small amount of bias (small in the sense that $\bar{\beta} - \beta_0$ is small when ε is small), one can still use this estimator. It follows from standard arguments, subject to regularity conditions (see, e.g., Hansen (1982)) that the asymptotic distribution of $\sqrt{T}(\check{\beta} - \bar{\beta})$ is $N(0, \check{V}_\beta)$, where

$$\check{V}_\beta^{-1} = (E[\Delta])^{-1} D_\beta' S_\beta^{-1} D_\beta \tag{4.25}$$

with the matrices $D_\beta \equiv E_{\Delta, Y_1, Y_0}[\dot{h}(Y_1, Y_0, \bar{\beta}, \varepsilon)]$, $S_{\beta, j} \equiv E_{\Delta, Y_{1+j}, Y_j, Y_1, Y_0}[h(Y_{1+j}, Y_j, \bar{\beta}, \varepsilon)h(Y_1, Y_0, \bar{\beta}, \varepsilon)']$, and $S_\beta \equiv \sum_{j=-\infty}^{+\infty} S_{\beta, j}$. The variance $\check{V}_\beta$ is the PFML version of what is denoted by Ω_β in (2.2). Theorem 1 can be applied to each one of the functions $\dot{h}$ and $h \times h$ (with conditioning on (Y_1, Y_0) first for $S_{\beta, j}$, $j \neq 0$) to obtain a Taylor expansion for $\check{V}_\beta$.

5. COMPARING THE COST OF DISCRETENESS TO THE COST OF RANDOMNESS

Now that we have defined precisely the three estimators FIML, IOML, and PFML, and analyzed their respective asymptotic distributions, we can use the differences between their asymptotic properties to identify the different costs imposed by not being able to see all the data (a discrete sample path instead of the entire continuous-time one, and/or not observing the sampling times) or by ignoring some of their features (such as the randomness of the sample intervals).

5.1. *Structure of the Information Matrices of FIML and IOML*

In order to measure the information loss associated with not observing the sampling intervals, we zero in on the difference between the variances of FIML and IOML. We now detail the asymptotic variances as a function of the drift

and diffusion functions specified originally in equation (2.1). Consider the three information matrices

$$I = E_{\Delta, Y_1, Y_0}[-\ddot{l}(Y_1|Y_0, \Delta, \beta)], \quad J = E_{Y_1, Y_0}\big[\mathrm{var}_{\Delta}[\dot{l}(Y_1|Y_0, \Delta, \beta)|Y_0, Y_1]\big],$$
$$\widetilde{I} = I - J.$$

From formula (4.9) above, $V_\beta = \varepsilon E[\Delta_0] I^{-1}$ and from (4.21) and Proposition 4, we have that $\widetilde{V}_\beta = \varepsilon E[\Delta_0]\widetilde{I}^{-1}$. For simplicity of notation only, we will write the results below as if θ were one-dimensional—the extension to multivariate θ involves partitioning matrices, while gaining little additional insight into the behavior of different estimators. We obtain the following theorem.

THEOREM 2 (FIML): *The Taylor expansion of the FIML Information Matrix is given by*

$$I = \begin{pmatrix} \varepsilon i^{(1)}_{\theta\theta} + \varepsilon^3 i^{(3)}_{\theta\theta} + O(\varepsilon^4) & \varepsilon i^{(1)}_{\theta\sigma^2} + \varepsilon^2 i^{(2)}_{\theta\sigma^2} + O(\varepsilon^3) \\ \varepsilon i^{(1)}_{\theta\sigma^2} + \varepsilon^2 i^{(2)}_{\theta\sigma^2} + O(\varepsilon^3) & i^{(0)}_{\sigma^2\sigma^2} + \varepsilon^2 i^{(2)}_{\sigma^2\sigma^2} + O(\varepsilon^3) \end{pmatrix} \tag{5.1}$$

where

$$i^{(1)}_{\theta\theta} = \frac{1}{\sigma_0^2} E_{\Delta_0, Y_0}\left[\Delta_0 \left(\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\right)^2\right],$$
$$i^{(1)}_{\theta\sigma^2} = \frac{1}{2\sigma_0^2} E_{\Delta_0, Y_0}\left[\Delta_0 \left(\frac{\partial^2 \mu}{\partial y \partial \theta}(Y_0, \theta_0)\right)\right],$$
$$i^{(0)}_{\sigma^2\sigma^2} = \frac{1}{2\sigma_0^4},$$

$$\begin{aligned} i^{(3)}_{\theta\theta} = E_{\Delta_0, Y_0}\Bigg[\Delta_0^3 \Bigg\{ & \frac{1}{8}\frac{\partial \mu}{\partial y}(Y_0, \theta_0)\left(\frac{\partial^2 \mu}{\partial y \partial \theta}(Y_0, \theta_0)\right)^2 \\ & - \frac{1}{12\sigma_0^2}\left(\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\right)^2 \left(\frac{\partial \mu}{\partial y}(Y_0, \theta_0)\right)^2 \\ & + \frac{1}{12}\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\frac{\partial^2 \mu}{\partial x \partial \theta}(Y_0, \theta_0)\frac{\partial^2 \mu}{\partial x^2}(Y_0, \theta_0) \\ & - \frac{\sigma_0^2}{48}\left(\frac{\partial^3 \mu}{\partial y^2 \partial \theta}(Y_0, \theta_0)\right)^2 \Bigg\}\Bigg], \end{aligned} \tag{5.2}$$

$$\begin{aligned} i^{(2)}_{\theta\sigma^2} = \frac{1}{6\sigma_0^2} E_{\Delta_0, Y_0}\Bigg[\Delta_0^2 \Bigg\{ & \frac{\partial \mu}{\partial y}(Y_0, \theta_0)\frac{\partial^2 \mu}{\partial y \partial \theta}(Y_0, \theta_0) \\ & + \frac{1}{2}\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\frac{\partial^2 \mu}{\partial y^2}(Y_0, \theta_0)\Bigg\}\Bigg], \end{aligned}$$

$$i^{(2)}_{\sigma^2\sigma^2} = \frac{1}{12\sigma_0^2} E_{\Delta_0, Y_0}\left[\Delta_0^2 \left(\frac{\partial^3 \mu}{\partial y^3}(Y_0, \theta_0)\right)\right].$$

A notable fact in (5.1) is that $i^{(2)}_{\theta\theta} = 0$ while $i^{(0)}_{\sigma^2\sigma^2} > 0$, which is another way of saying that, as we go to a continuous sampling limit, the sample path contains an order of magnitude in ε less information about θ than about σ^2. To determine the effect of not observing the observation times, we obtain the following result.

THEOREM 3 (IOML): *The Taylor expansion of the information loss due to not observing the sampling intervals is given by*

$$J = \begin{pmatrix} j^{(2)}_{\theta\theta}\varepsilon^2 + O(\varepsilon^3) & O(\varepsilon^2) \\ O(\varepsilon^2) & j^{(0)}_{\sigma^2\sigma^2} + O(\varepsilon) \end{pmatrix} \tag{5.3}$$

where

$$\begin{aligned} j^{(2)}_{\theta\theta} &= \frac{1}{4\sigma_0^2} E_{Y_0}\left[\left\{\sigma_0^2\left(\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\right)^2 - 2\left(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\right)^2\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\right\}\right. \\ &\qquad \left.\times \operatorname{var}[\Delta_0|\chi^2\Delta_0, Y_0]\right], \\ j^{(0)}_{\sigma^2\sigma^2} &= \frac{1}{4\sigma_0^4} E\big[\operatorname{var}[\chi^2|\chi^2\Delta_0, Y_0]\big], \end{aligned} \tag{5.4}$$

where χ^2 is a chi-squared random variable with one degree of freedom and is independent of Δ_0 and Y_0.

In Sections 5.2 (diffusion estimators) and 5.3 (drift estimators), we use these two theorems to derive expressions for the FIML (V_β) and IOML ($\widetilde{V}_\beta$) asymptotic variance matrices. They follow from inverting the matrices I and $\widetilde{I} = I - J$ respectively, calculating a Taylor expansion of the inverse, and multiplying the result by $\varepsilon E[\Delta_0]$. How the quantities $E[\operatorname{var}[\Delta_0|\chi^2\Delta_0, Y_0]]$ and $E[\operatorname{var}[\chi^2|\chi^2\Delta_0, Y_0]]$ in (5.4) depend upon the distribution of Δ_0 is the topic of Section 5.4. As is apparent from (5.4), these two quantities ultimately determine how large the information loss is when moving from FIML (where the randomly-spaced sampling intervals are used in the estimation) to IOML (where they are not).

5.2. *Behavior of the Diffusion Estimator*

By inverting the FIML information matrix, we now show that the FIML estimator $\hat{\sigma}^2$ has the asymptotic variance:

$$V_{\sigma^2} = V^{(1)}_{\sigma^2}\varepsilon + V^{(2)}_{\sigma^2}\varepsilon^2 + V^{(3)}_{\sigma^2}\varepsilon^3 + O(\varepsilon^4), \tag{5.5}$$

and similarly for $\widetilde{V}_{\sigma^2}$, v_{σ^2}, and $\tilde{v}_{\sigma^2}$. Recall that we use capital V to indicate that the coefficient is estimated alone (assume that the other is known), and lower case v to indicate the appropriate element in the variance matrix when the drift

and diffusion are estimated jointly. Recall also that we use V for FIML and $\widetilde{V}$ for IOML. The first thing to note is that the entire uncertainty when estimating σ^2 comes from the discretization: that is,

$$V^{(0)}_{\sigma^2} = \widetilde{V}^{(0)}_{\sigma^2} = v^{(0)}_{\sigma^2} = \tilde{v}^{(0)}_{\sigma^2} = 0.$$

Indeed, if the process (2.1) were observed continuously, then σ^2 would be the quadratic variation of X_t, and hence observable.

The first order terms in ε in the variances of the FIML and IOML estimators of σ^2 are the same whether θ is known or not, and they are given by the following corollary.

COROLLARY 1 (FIML vs. IOML: Leading Terms for σ^2):

$$V^{(1)}_{\sigma^2} = v^{(1)}_{\sigma^2} = \frac{E[\Delta_0]}{i^{(0)}_{\sigma^2\sigma^2}} = 2\sigma_0^4 E[\Delta_0], \tag{5.6}$$

$$\widetilde{V}^{(1)}_{\sigma^2} = \tilde{v}^{(1)}_{\sigma^2} = \frac{E[\Delta_0]}{i^{(0)}_{\sigma^2\sigma^2} - j^{(0)}_{\sigma^2\sigma^2}} = 2\sigma_0^4 E[\Delta_0]\left(1 - \frac{1}{2}E\big[\mathrm{var}[\chi^2|\chi^2\Delta_0, Y_0]\big]\right)^{-1}.$$

The consequence of this corollary is that the effect of throwing away the sampling times is to reduce the efficiency of $\tilde{\sigma}^2$ relative to $\hat{\sigma}^2$ at the level of the main order term: $\widetilde{V}^{(1)}_{\sigma^2} > V^{(1)}_{\sigma^2}$. Note that $\mathrm{var}[\chi^2] = 2$, and therefore

$$E\big[\mathrm{var}[\chi^2|\chi^2\Delta_0, Y_0]\big] = \mathrm{var}[\chi^2] - \mathrm{var}\big[E[\chi^2|\chi^2\Delta_0, Y_0]\big] \leq \mathrm{var}[\chi^2] = 2 \tag{5.7}$$

so $i^{(0)}_{\sigma^2\sigma^2} \geq j^{(0)}_{\sigma^2\sigma^2}$ in (5.6). The magnitude of the effect varies depending on the sampling pattern, i.e., the distribution of Δ_0; we study this in detail in Section 5.4 below. For now, it suffices to note that the loss of efficiency can get arbitrarily large. Consider, for instance, the sampling scheme independent of Y_0 given by a binomial distribution, $\Delta_0 = 0$ or $\Delta_0 = \bar{\delta}_0$ while keeping the average sampling interval $E[\Delta_0]$ fixed. In this case,

$$\mathrm{var}[\chi^2|\chi^2\Delta_0, Y_0] = \mathrm{var}[\chi^2]\mathbf{1}(\Delta_0 = 0) = 2 \times \mathbf{1}(\Delta_0 = 0)$$

where $\mathbf{1}(.)$ denotes the indicator function, and hence

$$\widetilde{V}^{(1)}_{\sigma^2} = \tilde{v}^{(1)}_{\sigma^2} = 2\sigma_0^4\bar{\delta}_0,$$

which can be made arbitrarily large by increasing the constant $\bar{\delta}_0$.

In other words, the variance increase imputable to the *discreteness* of the sampling is represented by the term $V^{(1)}_{\sigma^2} = 2\sigma^4 E[\Delta_0]$. The additional variance imputable to the *randomness* of the sampling, measured by the difference between $\widetilde{V}^{(1)}_{\sigma^2}$ and $V^{(1)}_{\sigma^2}$ can be made an *arbitrarily large* multiple of the discreteness-induced variance. Rather than the difference, we can consider the percentage increase in the variance of the IOML diffusion estimator compared

to the corresponding FIML variance. This is the percentage increase in variance that is attributable to not observing, or throwing away as the case may be, the sampling times. Based on equations (5.6), this is given by the term

$$\frac{\widetilde{V}_{\sigma^2}^{(1)} - \widehat{V}_{\sigma^2}^{(1)}}{\widehat{V}_{\sigma^2}^{(1)}} = \left(1 - \frac{1}{2}E\big[\mathrm{var}[\chi^2|\chi^2\Delta_0, Y_0]\big]\right)^{-1} - 1. \tag{5.8}$$

An interesting side aspect of the result is that this percentage increase is independent of the particular diffusion model (2.1) specified for the dynamics of the state variable X, that is, μ and σ^2 do not enter equation (5.8). The percentage increase only involves the distribution of the sampling times.

5.3. *Behavior of the Drift Estimator*

Estimating the drift parameter θ is fundamentally different from estimating σ^2 in that there would be nonzero asymptotic variance even if we observed the full continuous sample path. The asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta_0)$ takes the form

$$V_\theta = V_\theta^{(0)} + V_\theta^{(1)}\varepsilon + V_\theta^{(2)}\varepsilon^2 + O(\varepsilon^3) \tag{5.9}$$

with $V_\theta^{(0)}$ positive definite and similarly for $\widetilde{V}_\theta$, v_θ, and $\tilde{v}_\theta$. The main order effects are again the same as under continuous observation in all four cases:

COROLLARY 2 (FIML vs. IOML: Leading Terms for θ): *The leading term of the asymptotic variances for estimating θ is identical for FIML and IOML, and does not depend on whether σ^2 is estimated or known. It is given by*

$$\begin{aligned} V_\theta^{(0)} = \widetilde{V}_\theta^{(0)} = v_\theta^{(0)} = \tilde{v}_\theta^{(0)} &= \frac{E[\Delta_0]}{i_{\theta\theta}^{(1)}} \\ &= \sigma_0^2 E[\Delta_0]\left(E_{\Delta_0, Y_0}\left[\Delta_0\left(\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\right)^2\right]\right)^{-1}. \end{aligned} \tag{5.10}$$

The leading term (5.10) in the asymptotic variance of the θ estimator is the "inescapable" variance present even if the full continuous-time sample path were observable. It has a particularly nice interpretation, akin to the asymptotic variance in the nonlinear least squares situation. The variance is proportional to the inverse of the expected value of the squared gradient $(\partial\mu/\partial\theta)^2$ of the drift with respect to the parameter. Intuitively, the variance will therefore be high when the drift is relatively flat as a function of the parameter (i.e., $\partial\mu/\partial\theta$ close to 0). In that case, precise identification of θ is difficult. A contrario, when the drift function μ changes significantly in response to changes in the parameter θ (i.e., $\partial\mu/\partial\theta$ large), the drift will be estimated precisely.

Differences between FIML and IOML show up, however, in higher order terms. The most striking case occurs when σ^2 is known. In this instance, $V_\theta^{(1)} = 0$, while the corresponding term $\widetilde{V}_\theta^{(1)}$ is nonzero. In other words, not observing

(or not using) the sampling times is one order of magnitude more costly than the effect of discretization itself! The specific expressions for the terms of order ε^1 are given by the following corollary.

COROLLARY 3 (FIML vs. IOML: Higher Order Term for θ When $\sigma^2 = \sigma_0^2$ Is Known):

$$V_\theta^{(1)} = 0, \tag{5.11}$$

$$\tilde{V}_\theta^{(1)} = E[\Delta_0]\frac{j_{\theta\theta}^{(2)}}{\left(i_{\theta\theta}^{(1)}\right)^2}.$$

In the case where σ^2 is unknown, both $v_\theta^{(1)}$ and $\tilde{v}_\theta^{(1)}$ are nonzero. The FIML variance term $v_\theta^{(1)}$ is nonzero due to the contamination of the estimation of θ by the estimation of σ^2. In this case, however, there is a similar phenomenon to the one that occurs in the estimation of σ^2: the IOML variance $\tilde{v}_\theta^{(1)}$ can be made arbitrarily large by choosing specific sampling schemes, whereas $v_\theta^{(1)}$ stays bounded. The specific expressions are given by Corollary 4.

COROLLARY 4 (Higher Order Term for θ, When σ^2 Is Unknown):

$$v_\theta^{(1)} = E[\Delta_0]\frac{\left(i_{\theta\sigma^2}^{(1)}\right)^2}{\left(i_{\theta\theta}^{(1)}\right)^2 i_{\sigma^2\sigma^2}^{(0)}} = \frac{E[\Delta_0]\sigma_0^4}{2}\left(\frac{E_{\Delta_0, Y_0}\left[\Delta_0\left(\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\right)\right]}{E_{\Delta_0, Y_0}\left[\Delta_0\left(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\right)^2\right]}\right)^2 \tag{5.12}$$

and

$$\begin{aligned}\tilde{v}_\theta^{(1)} &= E[\Delta_0]\frac{\left(i_{\theta\sigma^2}^{(1)}\right)^2 + j_{\theta\theta}^{(2)}\left(i_{\sigma^2\sigma^2}^{(0)} - j_{\sigma^2\sigma^2}^{(0)}\right)}{\left(i_{\theta\theta}^{(1)}\right)^2\left(i_{\sigma^2\sigma^2}^{(0)} - j_{\sigma^2\sigma^2}^{(0)}\right)} \\ &= v_\theta^{(1)}\left(1 - \frac{1}{2}E\left[\operatorname{var}[\chi^2|\chi^2\Delta_0, Y_0]\right]\right)^{-1} + E[\Delta_0]\frac{j_{\theta\theta}^{(2)}}{\left(i_{\theta\theta}^{(1)}\right)^2}.\end{aligned} \tag{5.13}$$

As can be seen from these expressions, the first term on the right-hand side of (5.13), the IOML variance, contains the small ill-behaved expression, $(1 - \frac{1}{2}E[\operatorname{var}[\chi^2|\chi^2\Delta_0, Y_0]])^{-1}$, as the right-hand side of (5.6). Therefore $\tilde{v}_\theta^{(1)}$ blows up along with $\tilde{V}_{\sigma^2}^{(1)} = \tilde{v}_{\sigma^2}^{(1)}$ when the sampling scheme is such that $\frac{1}{2}E[\operatorname{var}[\chi^2|\,\chi^2\Delta_0, Y_0]]$ is close to one (recall from (5.7) that $\frac{1}{2}E[\operatorname{var}[\chi^2|\chi^2\Delta_0, Y_0]] \leq 1$), whereas from (5.12) the FIML variance $v_\theta^{(1)}$ remains bounded. In addition, the second term in (5.13) is of the form $c \times E[\operatorname{var}[\chi^2|\chi^2\Delta_0, Y_0]]/E[\Delta_0]$ where c does not depend on the sampling scheme. This term can therefore also be made arbitrarily large for fixed $E[\Delta_0]$ by selecting specific sampling schemes, as in Section 5.2.

5.4. *The Effect of the Sampling Distribution on the Loss of Efficiency*

More generally, the loss of efficiency associated with integrating out the sampling intervals is a function of $E[\operatorname{var}[\Delta_0|\chi^2\Delta_0, Y_0]]$ and $E[\operatorname{var}[\chi^2|\chi^2\Delta_0, Y_0]]$ as can be seen from Theorem 3, with equation (5.6) stating the implication for the

estimation of σ^2 and equations (5.12) and (5.13) for the estimation of θ. For clarity consider in this section the case where the density d_0 of Δ_0 is independent of Y_0 (that is, let us place ourselves under Assumption 1(iii)). Then the two quantities driving the efficiency loss of IOML relative to FIML are $E[\text{var}[\Delta_0|\chi^2\Delta_0]]$ and $E[\text{var}[\chi^2|\chi^2\Delta_0]]$. They depend upon the unconditional density $d_0(\delta_0)$ that $\Delta_0 = \delta_0$, and we now seek to get a better handle on that dependence. For that purpose, we define

$$m_q(b) = E_Z\left[Z^{-q}d_0\left(\frac{b}{Z^2}\right)\right], \tag{5.14}$$

where Z is $N(0,1)$. In the case where Δ_0 has a discrete distribution supported on $\delta_m, m = 1, \ldots, M$, we define instead

$$m_q(b) = \sum_{m=1}^{M}\left(\frac{b}{\delta_m}\right)^{2-\frac{q}{2}} g_{\chi^2}\left(\frac{b}{\delta_m}\right)P(\Delta_0 = \delta_m) \tag{5.15}$$

where g_{χ^2} is the $\chi^2[1]$ density function. As a function of the density d_0 from which the sampling intervals are drawn, the general formulae determining $E[\text{var}[\Delta_0|\chi^2\Delta_0]]$ and $E[\text{var}[\chi^2|\chi^2\Delta_0]]$ are given by the following.

PROPOSITION 5: *The conditional variances driving the efficiency loss in Theorem* 3 *and Corollaries* 1 *and* 4 *depend upon the distribution of the sampling intervals as follows*:

$$E[\text{var}[\chi^2|\chi^2\Delta_0]] = 3 - E\left[\left(\frac{m_0(\chi^2\Delta_0)}{m_2(\chi^2\Delta_0)}\right)^2\right] \tag{5.16}$$

and

$$E[\text{var}[\Delta_0|\chi^2\Delta_0]] = E[\Delta_0^2] - E\left[\left(\chi^2\Delta_0\frac{m_4(\chi^2\Delta_0)}{m_2(\chi^2\Delta_0)}\right)^2\right] \tag{5.17}$$

where χ^2 *and* Δ_0 *are independent random variables following respectively a* $\chi^2[1]$ *and a law with density* d_0.

We give the results in Table II for specific distributions of the sampling intervals. For the sampling schemes considered, the first four moments of Δ_0 and the functions $m_q, q = 0, 2, 4$ (hence all the expressions in Proposition 5, and consequently in Theorem 3 and Corollaries 1 and 4) are all in closed-form. We consider three sampling schemes, which, for comparison purposes, we parameterize as a function of their expected value $E[\Delta_0]$: (i) Δ_0 is nonrandom, i.e., had density given by a Dirac mass at $E[\Delta_0]$; (ii) Δ_0 is uniform on $(0, 2E[\Delta_0])$; (iii) Δ_0 has exponential distribution with arrival rate $1/E[\Delta_0]$, i.e., density $d_0(\delta) = (1/E[\Delta_0])\exp(-\delta/E[\Delta_0])$.

We can also consider a fourth scheme, corresponding to a trading day and night each with its own arrival distribution. Such a scheme would necessitate a simple relaxation of our stationarity assumption for the distribution of the sampling

TABLE II

MOMENTS OF THE SAMPLING PROCESS

Moment	$\Delta_0 \sim$ Uniform	$\Delta_0 \sim$ Exponential
$E[\Delta_0^2]$	$\frac{4}{3}E[\Delta_0]^2$	$2E[\Delta_0]^2$
$E[\Delta_0^3]$	$2E[\Delta_0]^3$	$6E[\Delta_0]^3$
$E[\Delta_0^4]$	$\frac{16}{5}E[\Delta_0]^4$	$24E[\Delta_0]^4$
$m_0(b)$	$\frac{1}{E[\Delta_0]}\left(1-\Phi\left(\sqrt{\frac{b}{2E[\Delta_0]}}\right)\right)$	$\frac{1}{E[\Delta_0]}e^{-\sqrt{2b/E[\Delta_0]}}$
$m_2(b)$	$\frac{1}{E[\Delta_0]}\left(\sqrt{\frac{E[\Delta_0]}{\pi b}}e^{-\frac{b}{4E[\Delta_0]}}+\Phi\left(\sqrt{\frac{b}{2E[\Delta_0]}}\right)-1\right)$	$\frac{1}{\sqrt{2bE[\Delta_0]}}e^{-\sqrt{2b/E[\Delta_0]}}$
$m_4(b)$	$\frac{1}{3E[\Delta_0]}\left(1+\left(\frac{2E[\Delta_0]}{b}-1\right)\sqrt{\frac{E[\Delta_0]}{\pi b}}e^{-\frac{b}{4E[\Delta_0]}}-\Phi\left(\sqrt{\frac{b}{2E[\Delta_0]}}\right)\right)$	$\frac{1}{2b}\left(1+\sqrt{\frac{E[\Delta_0]}{2b}}\right)e^{-\sqrt{2b/E[\Delta_0]}}$

Notes: For the sampling schemes considered, the first four moments of Δ and the functions m_q, $q=0,2,4$, in (5.16)–(5.17) are all explicit. By using these expressions in Proposition 5, the cost of randomness is obtained explicitly. Φ denotes the $N(0,1)$ c.d.f.

intervals. This straightforward extension is as follows. Assume that Δ_0 is generated by two sampling regimes, high density ("trading day" or "New-York open") and low density ("night" or "Singapore then London open"); the distributions of waiting times are $\Delta_1 = \varepsilon\Delta_{10}$ (day) and $\Delta = \varepsilon\Delta_{20}$ (night), with Δ_{10} and Δ_{20} being deterministic, uniform, or exponential (with different expected values, reflecting the different sampling intensities during the trading day and night). Suppose that α is the fraction of "day" out of total time (for example, $\alpha = 1/3$). Let τ_1 be the number of samples during the "day" and τ_2 during the night. Asymptotically as $\varepsilon \to 0$, τ_1 and τ_2 become nonrandom, with

$$\alpha = E[\Delta_{10}(1)+\cdots+\Delta_{10}(\tau)] = \tau_1\varepsilon E[\Delta_{10}],$$

and similarly $1-\alpha = \tau_2\varepsilon E[\Delta_{20}]$. Hence the asymptotic fraction of Δ_n's coming from the "day" distribution is

$$\phi = \frac{\tau_1}{\tau_1+\tau_2} = \frac{\alpha/E[\Delta_{10}]}{\alpha/E[\Delta_{10}]+(1-\alpha)/E[\Delta_{20}]}.$$

In this case, therefore,

$$E[\Delta_0^q] = \phi E[\Delta_{10}^q]+(1-\phi)E[\Delta_{20}^q], \tag{5.18}$$
$$m_q(b) = \phi m_{q,\Delta_1}(b)+(1-\phi)m_{q,\Delta_2}(b).$$

Note also that by combining these results, the expressions for additional sampling schemes can be straightforwardly obtained. For example, when Δ_0 has two regimes (for example a binomial distribution, or two regimes drawn from exponential distributions with different arrival rates), $E[\Delta_0^q]$ and $m_q(b)$ are given by (5.18) as a weighted average of the respective one-regime moments given in Table II.

Note also that when Δ_0 is exponentially distributed we obtain the particularly simple expressions

$$E\left[\left(\frac{m_0(\chi^2\Delta_0)}{m_2(\chi^2\Delta_0)}\right)^2\right] = 2,$$
$$E\left[\left(\chi^2\Delta_0\frac{m_4(\chi^2\Delta_0)}{m_2(\chi^2\Delta_0)}\right)^2\right] = \frac{5}{4}E[\Delta_0]^2 = \frac{5}{8}E[\Delta_0^2].$$

A consequence of this in equation (5.16) is that, for exponential sampling, $E[\mathrm{var}[\chi^2|\chi^2\Delta_0]] = 1$. Thus, from equation (5.8), the leading term in the IOML variance ($\widetilde{V}^{(1)}_{\sigma^2}$) is equal to twice the leading term in the FIML variance ($\widehat{V}^{(1)}_{\sigma^2}$). In other words, for the diffusion parameter σ^2, the cost of randomness (represented by the extra variance of IOML over FIML) is exactly equal, at first order in ε, to the cost of discreteness (represented by the FIML variance).

6. THE COST OF IGNORING THE SAMPLING RANDOMNESS

In the previous section, we examined the efficiency loss in FIML attributable to the sampling randomness and compared it to the one attributable to *not observing* the full continuous time sample path, the latter being represented by the extra variance of IOML relative to FIML. We now turn to a detailed examination of the cost (both in bias and variance) imputable to *ignoring* the randomness of the sampling intervals (by following what would be an optimal inference strategy, PFML, if the sampling intervals were all equal), as opposed to the cost of *not observing* the sampling intervals but being willing to account for their randomness (as in IOML).

We again distinguish between the three cases of estimating θ with $\sigma^2 = \sigma_0^2$ known, estimating σ^2 with $\theta = \theta_0$ known, and estimating both jointly. Recall also that we do the calculations in this section, for ease of exposition only, under the assumption that the density d_0 of Δ_0 is independent of Y_0, as was stated in Assumption 1(iii). The asymptotic properties of PFML are then given by $\sqrt{T}(\check{\beta} - \bar{\beta}) \to N(0, \check{V}_\beta)$, as follows.

THEOREM 4 (Bias and Variance for the PFML Estimator of θ When $\sigma^2 = \sigma_0^2$ Is Known): *The bias of the drift estimator is* $\bar{\theta} - \theta_0 = \varepsilon^2 b_2 + O(\varepsilon^3)$ *with*

$$b_2 = \frac{\sigma_0^2\,\mathrm{var}[\Delta_0]}{240E_{Y_0}[(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0))^2]}\left(20E_{Y_0}\left[\left(\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\right)^2\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y\partial\theta}\right]\right.$$
$$+10E_{Y_0}\left[\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\frac{\partial\mu(Y_0,\theta_0)}{\partial y}\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y^2}\right]$$
$$\left.-\sigma_0^2\left(11E_{Y_0}\left[\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y^2}\frac{\partial^3\mu(Y_0,\theta_0)}{\partial y^2\partial\theta}\right]\right.\right. \tag{6.1}$$

$$+4E_{Y_0}\left[\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y\partial\theta}\frac{\partial^3\mu(Y_0,\theta_0)}{\partial y^3}\right]+4E_{Y_0}\left[\frac{\partial\mu(Y_0,\theta_0)}{\partial y}\frac{\partial^4\mu(Y_0,\theta_0)}{\partial y^3\partial\theta}\right]$$
$$+E_{Y_0}\left[\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\frac{\partial^4\mu(Y_0,\theta_0)}{\partial y^4}\right]\Bigg)\Bigg)$$

and its asymptotic variance is

$$\check{V}_\theta=\check{V}_\theta^{(0)}+\check{V}_\theta^{(1)}\varepsilon+O(\varepsilon^2) \tag{6.2}$$

with $\check{V}_\theta^{(0)}=\sigma_0^2(E_{Y_0}[(\partial\mu(Y_0,\theta_0)/\partial\theta)^2])^{-1}$ *and*

$$\check{V}_\theta^{(1)}=\frac{\sigma_0^2E[\Delta_0^2]\big(\mathrm{var}[\Delta_0]\big(2E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\frac{\partial\mu(Y_0,\theta_0)}{\partial y}\big]-\sigma_0^2E_{Y_0}\big[\big(\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y\partial\theta}\big)^2\big]\big)+4\sigma_0^2T_\theta^{(2)}\big)}{4E[\Delta_0]E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\big]^3}$$

where $T_\theta^{(2)}=4E_{Y_0}[q_1^*(Y_0,\beta_0,0)G_1(Y_0,\beta_0)]$ *with* $G_1(Y_0,\beta_0)=\sigma_0^{-2}\int^{Y_0}\dot{\mu}(z,\theta_0)\,dz$ *and*

$$\begin{aligned}q_1^*(y,\beta_0,0)=-\frac{\mathrm{var}[\Delta_0]}{8\sigma_0^2}\Bigg(&4\mu(y,\theta_0)^2\frac{\partial^2\mu(y,\theta_0)}{\partial y\partial\theta}\\&+4\mu(y,\theta_0)\left(\frac{\partial\mu(y,\theta_0)}{\partial\theta}\frac{\partial\mu(y,\theta_0)}{\partial y}+\sigma_0^2\frac{\partial^3\mu(y,\theta_0)}{\partial y^2\partial\theta}\right)\\&+\sigma_0^2\Bigg(4\frac{\partial\mu(y,\theta_0)}{\partial y}\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y\partial\theta}+2\frac{\partial\mu(y,\theta_0)}{\partial\theta}\frac{\partial^2\mu(y,\theta_0)}{\partial y^2}\\&+\sigma_0^2\frac{\partial^4\mu(y,\theta_0)}{\partial y^3\partial\theta}\Bigg)\Bigg).\end{aligned} \tag{6.3}$$

Note that $\check{V}_\theta^{(0)}$ is the limiting term corresponding to a continuous record of observations, as in FIML and IOML. So the three estimators are equivalent at the level of the leading term for the variance of the drift estimator. Moreover note that, as expected, setting $\mathrm{var}[\Delta_0]=0$ in all the expressions involving the PFML estimator reduces its bias to zero and its variance to that of FIML (and IOML) since in that case all three estimators are identical. This will also be a feature of the results of the two theorems that follow.

THEOREM 5 (Bias and Variance for the PFML Estimator of σ^2 When $\theta=\theta_0$ Is Known): *The bias of the diffusion estimator is*

$$\bar{\sigma}^2-\sigma_0^2=\varepsilon\frac{\mathrm{var}[\Delta_0]\sigma_0^2}{2E[\Delta_0]}E_{Y_0}\left[\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\right]+O(\varepsilon^2) \tag{6.4}$$

and its asymptotic variance is

$$\check{V}_{\sigma^2}=\varepsilon\check{V}_{\sigma^2}^{(1)}+\varepsilon^2\check{V}_{\sigma^2}^{(2)}+O(\varepsilon^2) \tag{6.5}$$

with

$$\check{V}^{(1)}_{\sigma^2} = \frac{\sigma_0^4}{E[\Delta_0]}(3\,\mathrm{var}[\Delta_0] + 2E[\Delta_0]^2),$$

$$\check{V}^{(2)}_{\sigma^2} = \frac{\sigma_0^4}{E[\Delta_0]}\left(\left(3(E[\Delta_0^3] - E[\Delta_0]^3) - \mathrm{var}[\Delta_0]E[\Delta_0]\right)E_{Y_0}\left[\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\right] + 4\sigma_0^4 E[\Delta_0]^2 T^{(3)}_{\sigma^2}\right),$$

where

$$T^{(3)}_{\sigma^2} = \frac{2}{E[\Delta_0]}\left\{E_{Y_0}\left[q_2(Y_0,\beta_0,0)r_2(Y_0,\beta_0,0)\right] + \frac{3\,\mathrm{var}[\Delta_0^2]}{2\sigma_0^2 E[\Delta_0]}E_{Y_0}\left[\mu(Y_0,\theta_0)\frac{\partial r_2(Y_0,\beta_0,0)}{\partial y}\right] + \frac{(3E[\Delta_0^2] - E[\Delta_0]^2)}{4E[\Delta_0]}E_{Y_0}\left[\frac{\partial^2 r_2(Y_0,\beta_0,0)}{\partial y^2}\right]\right\}$$

with

$$q_2(y,\beta_0,0) = \frac{\mathrm{var}[\Delta_0]}{4E[\Delta_0]\sigma_0^4}\left(\sigma_0^2 E_{Y_0}\left[\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\right] - 2\left(\mu(y,\theta_0)^2 + \sigma_0^2\frac{\partial\mu}{\partial y}(y,\theta_0)\right)\right). \tag{6.6}$$

This theorem shows that there is a first order loss of efficiency in PFML represented in $\check{V}^{(1)}_{\sigma^2}$ by the additional term $3\sigma_0^4\,\mathrm{var}[\Delta_0]/E[\Delta_0]$, in addition to the cost of discreteness represented by the FIML leading term variance $V^{(1)}_{\sigma^2} = 2\sigma_0^4 E[\Delta_0]$ of Corollary 1.

Compared now to IOML, recall that the IOML estimator relies on standard likelihood inference using only the price data Y_n. The asymptotic efficiency of IOML suggests that its variance might be smaller or equal to the variance of the PFML which makes use of exactly the same data (the Y_n's but not the Δ_n's). However, because of the inconsistency of the PFML procedure, this is not assured. The inequality does, however, hold in the case of estimating σ^2 with $\theta = \theta_0$ known, since

$$\check{V}^{(1)}_{\sigma^2} = \left(1 + \frac{3}{2}\mathrm{var}[\Delta_0]\right)\widetilde{V}^{(1)}_{\sigma^2} \geq \widetilde{V}^{(1)}_{\sigma^2}. \tag{6.7}$$

THEOREM 6 (Bias and Variance for the Joint PFML Estimator of $\beta = (\theta, \sigma^2)'$): *The two components of the bias vector $\bar{\beta} - \beta_0$ are given by*

$$\bar{\theta} - \theta_0 = -\varepsilon\frac{\sigma_0^2\,\mathrm{var}[\Delta_0]}{4E[\Delta_0]E_{Y_0}\left[\left(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\right)^2\right]}E_{Y_0}\left[\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\right] \times E_{Y_0}\left[\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y\partial\theta}\right] + O(\varepsilon^2) \tag{6.8}$$

and (6.4) *respectively. The asymptotic variance of* $\hat{\beta}$ *is*

$$\check{V}_\beta = \begin{pmatrix} \check{v}_{\theta\theta} & \check{v}_{\theta\sigma^2} \\ \check{v}_{\sigma^2\theta} & \check{v}_{\sigma^2\sigma^2} \end{pmatrix} = \begin{pmatrix} \check{v}^{(0)}_{\theta\theta} + \varepsilon\check{v}^{(1)}_{\theta\theta} + O(\varepsilon^2) & \varepsilon\check{v}^{(1)}_{\theta\sigma^2} + O(\varepsilon^2) \\ \varepsilon\check{v}^{(1)}_{\sigma^2\theta} + O(\varepsilon^2) & \varepsilon\check{v}^{(1)}_{\sigma^2\sigma^2} + \varepsilon^{2\check{v}(2)}_{\sigma^2\sigma^2} + O(\varepsilon^3) \end{pmatrix} \tag{6.9}$$

where $\check{v}^{(0)}_{\theta\theta} = \check{V}^{(0)}_\theta$, $\check{v}^{(1)}_{\sigma^2\sigma^2} = \check{V}^{(1)}_{\sigma^2}$, *and*

$$\begin{aligned} \check{v}^{(1)}_{\theta\theta} = \check{V}^{(1)}_\theta &+ \frac{\sigma_0^4\big((2E[\Delta_0]^2 + 3\mathrm{var}[\Delta_0])E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\big] \times E_{Y_0}\big[\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\big]^2\big)}{4E[\Delta_0]E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\big]^3} \\ &+ \frac{\sigma_0^4\mathrm{var}[\Delta_0]E_{Y_0}\big[\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\big]\big(2E_{Y_0}\big[\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\frac{\partial^2\mu}{\partial\theta^2}(Y_0,\theta_0)\big] \times E_{Y_0}\big[\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\big] - E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\big]E_{Y_0}\big[\frac{\partial^3\mu}{\partial y\partial\theta^2}(Y_0,\theta_0)\big]\big)}{2E[\Delta_0]E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\big]^3} \\ &+ \frac{\sigma_0^4\big((t^{(2)}_{\theta\theta} - T^{(2)}_\theta) - 2\sigma_0^2E[\Delta_0]E_{Y_0}\big[\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\big]t^{(2)}_{\theta\sigma^2}\big)}{E[\Delta_0]E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\big]^2}, \end{aligned}$$

$$\check{v}^{(1)}_{\sigma^2\theta} = \check{v}^{(1)}_{\theta\sigma^2} = -\frac{\sigma_0^4\big((2E[\Delta_0]^2 + 3\mathrm{var}[\Delta_0])E_{Y_0}\big[\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\big] - 4\sigma_0^2E[\Delta_0]t^{(2)}_{\theta\sigma^2}\big)}{2E[\Delta_0]E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\big]},$$

$$\check{v}^{(2)}_{\sigma^2\sigma^2} = \check{V}^{(2)}_{\sigma^2} + \frac{4\sigma_0^6E_{Y_0}\big[\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\big]\big((E[\Delta_0]^2 + 3\mathrm{var}[\Delta_0])E_{Y_0}\big[\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\big] - 4\sigma_0^2E[\Delta_0]t^{(2)}_{\theta\sigma^2}\big)}{E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\big]},$$

with $t^{(2)}_{\theta\theta} = 4E_{Y_0}[G_1(Y_0,\beta_0)q_1(Y_0,\beta_0,0)]$, $t^{(2)}_{\theta\sigma^2} = 2E_{Y_0}[G_1(Y_0,\beta_0)q_2(Y_0,\beta_0,0)]$, *and* $t^{(3)}_{\sigma^2\sigma^2} = T^{(3)}_{\sigma^2}$ *where*

$$\begin{aligned} q_1(y,\beta_0,0) = q_1^*(y,\beta_0,0) &+ \frac{\mathrm{var}[\Delta_0]E_{Y_0}\big[\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\big]\big(E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\big]\frac{\partial^2\mu(y,\theta_0)}{\partial y\partial\theta} - E_{Y_0}\big[\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y\partial\theta}\big]\big(\frac{\partial\mu}{\partial\theta}(y,\theta_0)\big)^2\big)}{4E_{Y_0}\big[\big(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\big)^2\big]}, \end{aligned} \tag{6.10}$$

with q_1^* *denoting the function defined in* (6.3) *and* q_2 *the function defined in* (6.6).

This theorem shows that the additional variance over FIML is of the same order for both of these inference schemes, but of course the PFML estimator is biased unlike the IOML one. In terms of asymptotic root mean squared error

comparisons, any biased estimator such as PFML will always do worse than an unbiased estimator such as FIML or IOML, irrespectively of how large their variances are, since the variance is of order $O(T^{-1})$ whereas the squared bias is of order $O(1)$ as T goes to infinity.

This being said, from the above results, the PFML procedure is relatively much worse for estimating σ^2 than for estimating θ. The first order variance term is unaffected for the drift estimator, while for the diffusion estimator, it is larger than under IOML unless $\text{var}[\Delta_0] = 0$ (the deterministic sampling situation, in which case our three procedures are the same anyway). Also, the impact of bias is much greater when estimating σ^2 using PFML. To illustrate this, consider first the case of $\breve{\sigma}^2$, the PFML diffusion estimator. Suppose that ϵ is some fixed number small enough that the first order term in the bias (6.4) is a good approximation, i.e., $\breve{V}_{\sigma^2} \approx \epsilon \breve{V}^{(1)}_{\sigma^2}$, and similarly for the bias, say, $\bar{\sigma}^2 - \sigma_0^2 \approx \epsilon c$, where c is the expression for the first order term in (6.5). At the same time, let T be big enough that the asymptotic variances and biases are good approximations. The mean squared error MSE_{σ^2} is then given by

$$\text{MSE}_{\sigma^2} = E\left[\left(\breve{\sigma}^2 - \sigma_0^2\right)^2\right] = \text{var}\left[\breve{\sigma}^2 - \sigma_0^2\right] + \text{bias}^2 \approx \epsilon T^{-1} \breve{V}^{(1)}_{\sigma^2} + \epsilon^2 c^2.$$

In order for T to be big enough for the bias to not dominate MSE_{σ^2}, we therefore need, approximately, that ϵ be smaller than $c_1 T^{-1}$, where $c_1 = \breve{V}^{(1)}_{\sigma^2}/c^2$. On the other hand, for $\breve{\theta}$, one needs ϵ to be approximately smaller than $c_2 T^{-1/4}$ and $c_3 T^{-1/2}$ for the cases of Theorems 4 and 6, respectively, where c_2 and c_3 follow from these theorems. Hence, ϵ can be much bigger for $\breve{\theta}$ than for $\breve{\sigma}^2$ without having the bias dwarfing the variance for the PFML estimators.

One final remark. We wrote earlier that we were being "generous" with regard to the PFML inference strategy by evaluating the likelihood function at the unconditional mean of the sampling intervals, $E[\Delta] = \varepsilon E[\Delta_0]$. We now show that this is indeed the best one can do for the PFML scheme, in the sense that the choice of any other fixed value at which to evaluate the likelihood function would result in an estimator with even worse asymptotic properties. Suppose that we evaluate the likelihood function at $\varepsilon\omega_0$, where $\omega_0 \neq E[\Delta_0]$, so that our moment conditions become

$$h(y_1, y_0, \beta, \varepsilon) = \begin{bmatrix} h_1(y_1, y_0, \beta, \varepsilon) \\ h_2(y_1, y_0, \beta, \varepsilon) \end{bmatrix} = \begin{bmatrix} \dot{l}_\theta(y_1|y_0, \varepsilon\omega_0, \beta) \\ \dot{l}_{\sigma^2}(y_1|y_0, \varepsilon\omega_0, \beta) \end{bmatrix} \tag{6.11}$$

instead of (4.23). We then obtain the following result.

PROPOSITION 6 (PFML Estimator When the Likelihood Is Not Evaluated at $E[\Delta]$):

(i) *If we estimate θ with $\sigma^2 = \sigma_0^2$ known, the bias of the drift estimator is an order of magnitude in ε larger than when $\omega_0 = E[\Delta_0]$. Letting $\bar{\theta}(\omega_0)$ denote the* plim *of*

the estimator, we now have $\bar{\theta}(\omega_0) - \theta_0 = \varepsilon b_1(\omega_0) + O(\varepsilon^2)$ *with*

$$b_1(\omega_0) = (E[\Delta_0] - \omega_0) \times \frac{\sigma_0^2\left(2E_{Y_0}\left[\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y\partial\theta}\right] + E_{Y_0}\left[\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y^2}\right]\right)}{12\ E_{Y_0}\left[\left(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\right)^2\right]} \tag{6.12}$$

as opposed to (6.1). *The asymptotic variance of the estimator is of the same order as when* $\omega_0 = E[\Delta_0]$.

(ii) *If we attempt to estimate* σ^2 *with* $\theta = \theta_0$ *known, the bias of the PFML estimator does not even go to zero as* ϵ *goes to zero. Indeed, if* $\bar{\sigma}^2(\omega_0)$ *is the* plim *of the estimator, we have* $\bar{\sigma}^2(\omega_0) - \sigma_0^2 = O(1)$ *instead of the usual* $O(\varepsilon)$; *specifically,*

$$\bar{\sigma}^2(\omega_0) - \sigma_0^2 = \left(\frac{E[\Delta]_0}{\omega_0} - 1\right)\sigma_0^2 + O(\varepsilon). \tag{6.13}$$

As a sanity check, we note that the expressions given in (6.12) and (6.13) are, respectively, proportional to $E[\Delta_0] - \omega_0$ and $E[\Delta_0]/\omega_0 - 1$, so that they revert to those given in Theorems 4 and 5 respectively when the likelihood function is evaluated at the unconditional mean of the sampling intervals, i.e., when $\omega_0 = E[\Delta_0]$.

7. EXAMPLE: THE ORNSTEIN-UHLENBECK PROCESS

This section studies a specific example of model (2.1), the stationary ($\theta > 0$) Ornstein-Uhlenbeck process

$$dX_t = -\theta X_t\, dt + \sigma\, dW_t. \tag{7.1}$$

The transition density $l(y_1|y_0,\delta,\beta) = \ln(p(y_1|y_0,\delta,\beta))$ is a Gaussian density with expected value $e^{-\delta\theta}y_0$ and variance $(1-e^{-2\delta\theta})\sigma^2/2\theta$. The stationary density $\pi(y_0,\beta)$ is also Gaussian, with mean 0 and variance $\sigma^2/(2\theta)$. In Section 7.1, we determine how accurate our expansions are by comparing them to the exact expressions that can be obtained for this model. We also compare how the IOML and PFML fare relative to FIML in Section 7.2. Finally, we report in Section 7.3 the results of Monte Carlo simulations to address the issue of the adequacy of the asymptotic distributions in small samples.

7.1. *Accuracy of the Expansions*

We start by determining the accuracy of our expansions. Indeed, this model constitutes one of the rare instances where we can obtain exact (i.e., non-Taylor expanded) expressions for the asymptotic variance matrices V_β, $\tilde{V}_\beta$, and $\check{V}_\beta$ of the three estimators. We can therefore compare the exact expressions for these matrices to their Taylor expansions of different orders in ε.

The exact FIML expressions, obtained by a direct calculation using the explicit expressions of the transition and stationary densities, are derived from the information matrix

$$I_\beta = \begin{pmatrix} E\left[\frac{(-1+e^{2\varepsilon\Delta_0\theta_0})^2-4(-1+e^{2\varepsilon\Delta_0\theta_0})\varepsilon\Delta_0\theta_0+2(1+e^{2\varepsilon\Delta_0\theta_0})\varepsilon^2\Delta_0^2\theta_0^2}{2(-1+e^{2\varepsilon\Delta_0\theta_0})^2\theta_0^2}\right] & E\left[\frac{1-e^{2\varepsilon\Delta_0\theta_0}+2\varepsilon\Delta_0\theta_0}{-2\theta_0\sigma_0^2+2e^{2\varepsilon\Delta_0\theta_0}\theta_0\sigma_0^2}\right] \\ E\left[\frac{1-e^{2\varepsilon\Delta_0\theta_0}+2\varepsilon\Delta_0\theta_0}{-2\theta_0\sigma_0^2+2e^{2\varepsilon\Delta_0\theta_0}\theta_0\sigma_0^2}\right] & \frac{1}{2\sigma_0^4} \end{pmatrix} \tag{7.2}$$

where the expectation is taken over the distribution of Δ_0. The FIML asymptotic variance is $V_\beta = \varepsilon E[\Delta_0] I_\beta^{-1}$. Our corresponding expansion in ε for the information matrix I_β, obtained by applying Theorem 2 with $\mu(x,\theta) = -\theta x$, is

$$I_\beta = \begin{pmatrix} \varepsilon\frac{E[\Delta_0]}{2\theta_0} - \varepsilon^3\frac{E[\Delta_0^3]\theta_0}{6} + O(\varepsilon^4) & -\varepsilon\frac{E[\Delta_0]}{2\sigma_0^2} + \varepsilon^2\frac{E[\Delta_0^2]\theta_0}{6\sigma_0^2} + O(\varepsilon^3) \\ -\varepsilon\frac{E[\Delta_0]}{2\sigma_0^2} + \varepsilon^2\frac{E[\Delta_0^2]\theta_0}{6\sigma_0^2} + O(\varepsilon^3) & \frac{1}{2\sigma_0^4} \end{pmatrix}. \tag{7.3}$$

For the PFML estimator, the exact calculation of the time series term T_β is based on

$$T_\beta = \frac{2}{E[\Delta_0]} E_{\Delta, Y_1, Y_0}[h(Y_1, Y_0, \Delta, \bar\beta, \varepsilon) R(Y_1, \beta_0, \varepsilon)]$$

where $E_{\Delta, Y_1}[h(Y_0, Y_1, \Delta, \bar\beta, \varepsilon)|Y_0] = \varepsilon^\alpha q(Y_0, \beta_0, \varepsilon) \equiv Q(Y_0, \beta_0, \varepsilon)$ and

$$R(Y_1, \beta_0, \varepsilon) = E[\Delta_0] \sum_{k=1}^{\infty} E_{Y_k}[Q(Y_k, \beta_0, \varepsilon)|Y_1] = \varepsilon^{\alpha-1} r(Y_1, \beta_0, \varepsilon).$$

This last expression requires the calculation of $E[Y_k^2|Y_1]$. To this end, consider first the law of Y_k given Y_1 and $\Delta_2, \ldots, \Delta_k$. In this case, Y_k is conditionally Gaussian with mean $Y_1 \exp\{-\theta(\Delta_2 + \cdots + \Delta_k)\}$ and variance $((k-1) - \exp\{-2\theta(\Delta_2 + \cdots + \Delta_k)\})\sigma^2/(2\theta)$. Hence we obtain that

$$\begin{aligned} E\big[Y_k^2|Y_1\big] &= E\Big[Y_1^2 \exp\{-2\theta(\Delta_2 + \cdots + \Delta_k)\} \\ &\qquad + \frac{\sigma^2}{2\theta}((k-1) - \exp\{-2\theta(\Delta_2 + \cdots + \Delta_k)\})|Y_1\Big] \\ &= Y_1^2 E[\exp\{-2\theta\Delta\}]^{(k-1)} + \frac{\sigma^2}{2\theta}\Big((k-1) - E[\exp\{-2\theta\Delta\}]^{(k-1)}\Big). \end{aligned}$$

To save space, we only report results for the FIML estimator in Figure 3, and for the exponential sampling scheme described in Section 5.4. That is, we illustrate our results in the special case where Δ_0 is independent of Y_0 and has density $d_0(\delta_0) = \exp(-\delta_0/E[\Delta_0])/E[\Delta_0]$. The value of the average sampling interval $E[\Delta] = \varepsilon E[\Delta_0]$ on the graph ranges from 0 (the continuous-time limit) to 1/12 (one month). The true parameter values are $\theta_0 = 10$ and $\sigma_0^2 = 1$. We find that in all instances the approximation error introduced by using our expansion to evaluate the exact expression is much smaller than the magnitude of the estimator's

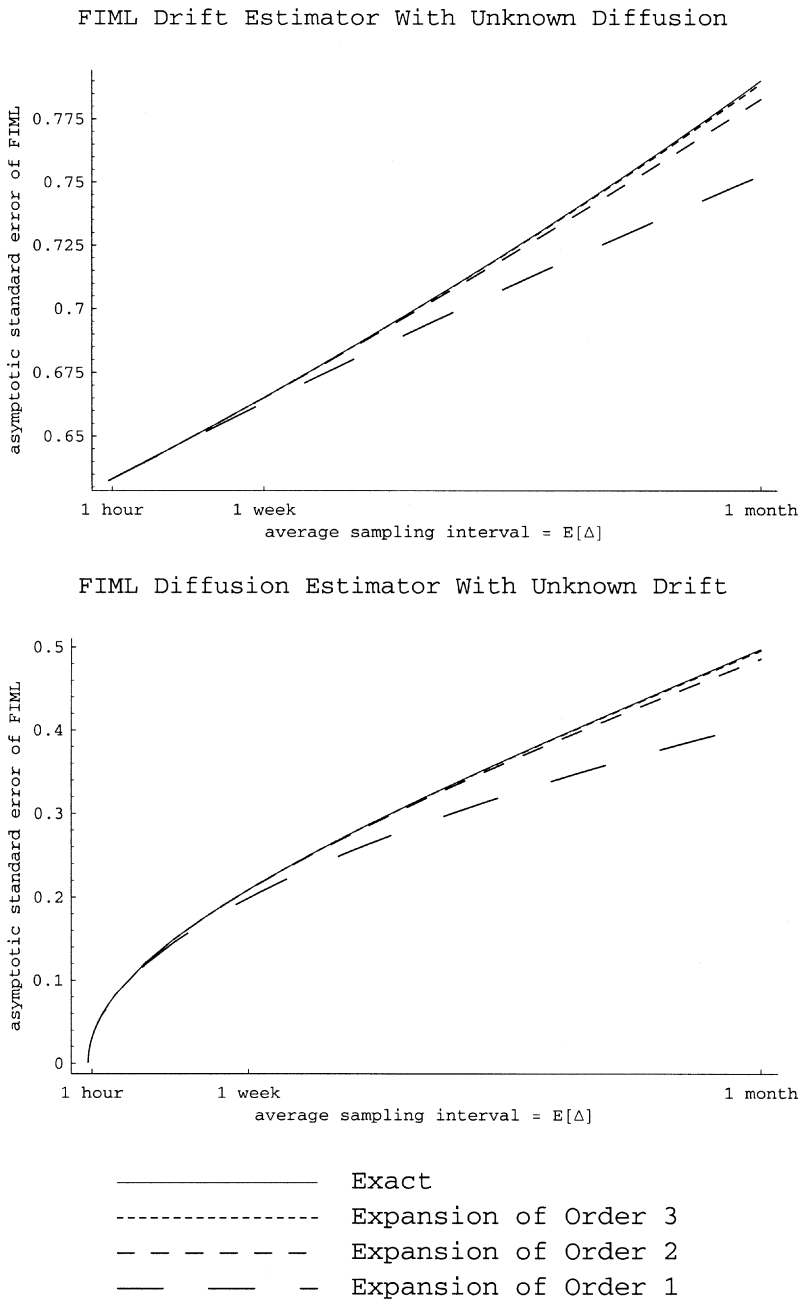


FIGURE 3.—Accuracy of the asymptotic variance expansions.

standard error. It is also the case that higher order Taylor expansions in ε are more accurate than low order expansions, when $E[\Delta]$ gets large. For high frequency data (small $E[\Delta]$, such as a day or less), expansions of order 1 in ε are amply sufficient. But even for monthly data, one can get very accurate results by using the expansion of order 3 in ε. Similar results hold for the other estimation methods. Therefore, in practical situations, our expansions can be taken as substitutes for the exact expressions (which we recall are not known in closed form for most models anyway).

7.2. *Comparison of the Different Methods*

We now compare the three estimators in terms of their asymptotic bias (if any) and standard error. We report in Figure 4 the results for the parameters estimated one at a time (i.e., θ knowing $\sigma^2 = \sigma_0^2$ and σ^2 knowing $\theta = \theta_0$). The complete results are given in full generality in the Theorems contained earlier in the paper, and specialized to the Ornstein-Uhlenbeck process in Tables III (FIML and IOML) and IV (PFML). The results in Figure 4 demonstrate that the differences in the asymptotic performance of the estimators can be substantial. The asymptotic standard error and bias are both scaled by the value of the true parameter.

In Figure 5, we plot the costs of discreteness and the cost of randomness for this model. Recall that the results coming out of the theory are as follows. For the drift parameters, the cost of discreteness (which is of order ε^1) is an order of magnitude in ε larger than the cost of randomness (which is of order ε^2); however, both are smaller than the inescapable variance that is present even if the full continuous time sample path is observed (which is of order ε^0). The difference in the orders of the two costs (linear in ε for the cost of randomness, quadratic for the cost of discreteness) when estimating θ is apparent in Figure 5 (first graph).

For the diffusion parameter, the limiting variance in the case of full continuous time sample path observation is zero. The cost of discreteness is of order ε^1. The cost of randomness is also of order ε^1; however, the latter can be made an arbitrary large multiple of the former. Recall also from the discussion in Section 5.4 that in the case of exponential sampling, the cost of randomness when estimating σ^2 is exactly equal to the cost of randomness. This is visible in Figure 5 (second graph).

7.3. *Simulation Evidence on the Small Sample Distributions*

Next, we carry out simulations under different values of T and $E[\Delta]$, again with exponential sampling. Figures 6, 7, and 8 plot the finite sample distributions of the estimators (histograms) and the corresponding asymptotic distribution (solid line). The expressions for the asymptotic bias and variances of the three estimators applied to this model are given in Tables III and IV.

Throughout the simulations, the true parameter values are $\theta_0 = 10$ and $\sigma_0^2 = 1$. To simulate the sample paths ($M = 5{,}000$ in each case), we first draw Δ_n from

the exponential distribution with parameter $E[\Delta]$; then, given Δ_n and Y_{n-1}, recursively draw Y_n conditionally on Y_{n-1} and Δ_n from a Gaussian distribution with conditional mean $\exp(-\theta_0\Delta_n)Y_{n-1}$ and variance $(1-\exp(-2\theta_0\Delta_n))\sigma_0^2/2\theta_0$. To start each simulated series, we draw Y_0 from the unconditional distribution of the process X. To compute the closed-form approximation of the integrated-out likelihood function $\tilde{p}(y_1|y_0,\beta)$, we apply Proposition 3, with the coefficients

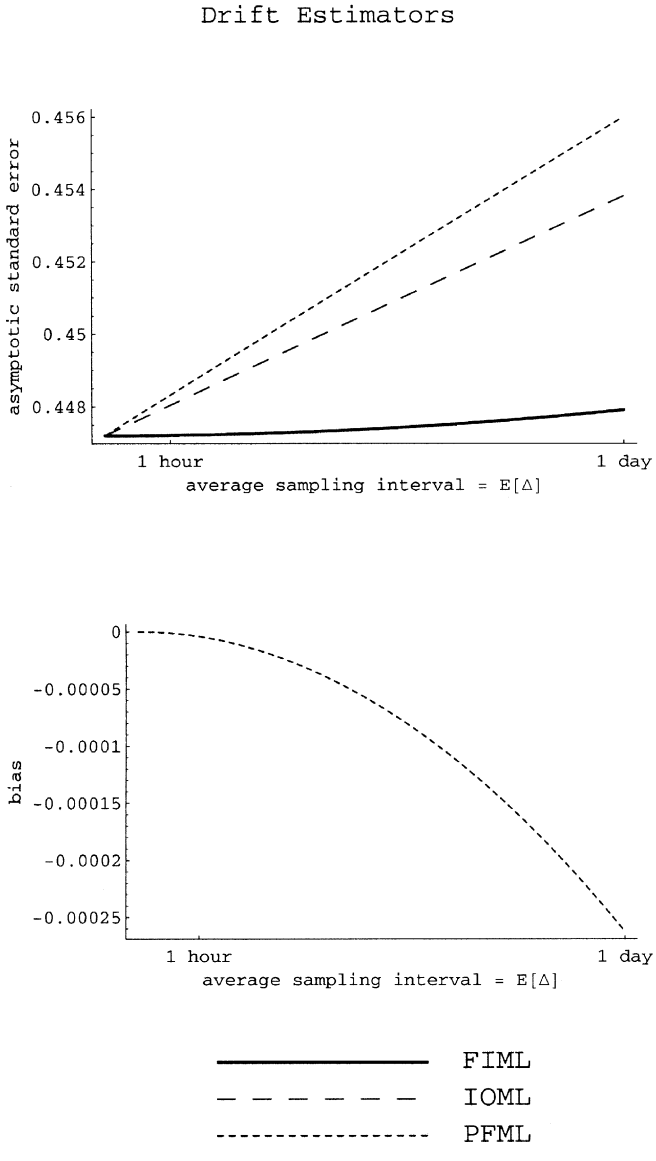


FIGURE 4.—Comparison of estimators when the sampling is exponential.

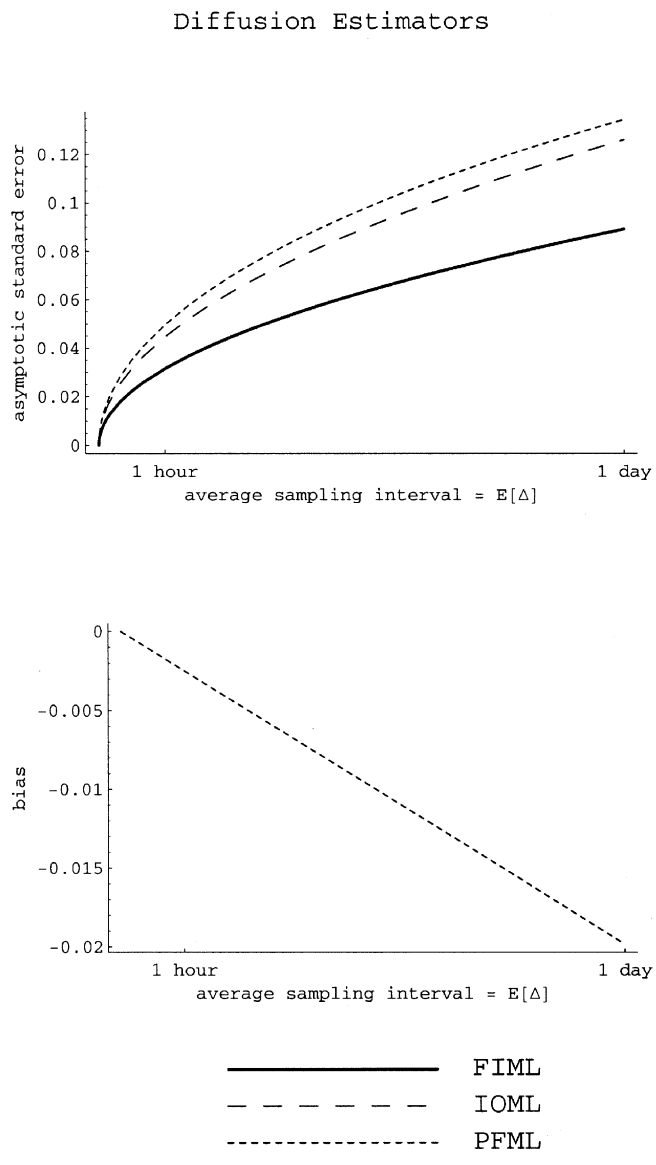


Figure 4.—*Continued*.

c_j, $j = 0, \ldots, 3$ of the explicit transition function given by equation (4.6). The expressions of the c_j coefficients are:

$$c_0(z_1|z_0,\beta) = 1,$$

$$c_1(z_1|z_0,\beta) = -\frac{\theta(-3\sigma^2+\theta z_0^2)}{6\sigma^2} - \frac{\theta^2 z_0 z_1}{6\sigma^2} - \frac{\theta^2 z_1^2}{6\sigma^2},$$

TABLE III

ORNSTEIN-UHLENBECK PROCESS ESTIMATED USING LIKELIHOOD METHODS

	FIML (Δ Included)	IOML (Δ Integrated Out)
D_θ	$-\varepsilon\left(\frac{E[\Delta_0]}{2\theta_0}\right)+\varepsilon^3\left(\frac{E\left[\Delta_0^3\right]\theta_0}{6}\right)+O(\varepsilon^4)$	$-\varepsilon\left(\frac{E[\Delta_0]}{2\theta_0}\right)+\varepsilon^2\left(\frac{E[\mathrm{var}[\Delta_0\vert\chi^2\Delta_0]]}{2}\right)+O(\varepsilon^3)$
$S_{\theta,0}$	$\varepsilon\left(\frac{E[\Delta_0]}{2\theta_0}\right)-\varepsilon^3\left(\frac{E\left[\Delta_0^3\right]\theta_0}{6}\right)+O(\varepsilon^4)$	$\varepsilon\left(\frac{E[\Delta_0]}{2\theta_0}\right)-\varepsilon^2\left(\frac{E[\mathrm{var}[\Delta_0\vert\chi^2\Delta_0]]}{2}\right)+O(\varepsilon^3)$
T_θ	0	0
Ω_θ	$2\theta_0+\varepsilon^2\left(\frac{2\theta_0^3E\left[\Delta_0^3\right]}{3E[\Delta_0]}\right)+O(\varepsilon^3)$	$2\theta_0+\varepsilon\left(\frac{2\theta_0^2E[\mathrm{var}[\Delta_0\vert\chi^2\Delta_0]]}{E[\Delta_0]}\right)+O(\varepsilon^2)$
$\bar\theta-\theta_0$	0	0
D_{σ^2}	$-\frac{1}{2\sigma_0^4}$	$-\frac{2-E[\mathrm{var}[\chi^2\vert\chi^2\Delta_0]]}{4\sigma_0^4}+O(\varepsilon)$
$S_{\sigma^2,0}$	$\frac{1}{2\sigma_0^4}$	$\frac{2-E[\mathrm{var}[\chi^2\vert\chi^2\Delta_0]]}{4\sigma_0^4}+O(\varepsilon)$
T_{σ^2}	0	0
Ω_{σ^2}	$\varepsilon(2\sigma_0^4E[\Delta_0])$	$\varepsilon\left(\frac{4\sigma_0^4E[\Delta_0]}{2-E[\mathrm{var}[\chi^2\vert\chi^2\Delta_0]]}\right)+O(\varepsilon^2)$
$\bar\sigma^2-\sigma_0^2$	0	0

Notes: These expressions follow from applying Theorems 2 (FIML) and 3 (IOML) to the Ornstein-Uhlenbeck process. Note that if Δ is actually not random, then integrating out Δ is identical to FIML. Our expressions in the second column indeed reduce to those of the FIML case (first column) when $E[\mathrm{var}[\Delta_0|\chi^2\Delta_0]]=E[\mathrm{var}[\chi^2|\chi^2\Delta_0]]=0$. Defining for the log-likelihood score vectors $h=l$ (FIML) or $h=\dot\lambda$ (IOML),

$$D_\beta\equiv E_{\Delta,Y_1,Y_0}[\dot h(Y_1,Y_0,\Delta,\tilde\beta)],$$
$$S_{\beta,0}\equiv E_{\Delta,Y_1,Y_0}[h(Y_1,Y_0,\Delta,\tilde\beta)h(Y_1,Y_0,\Delta,\tilde\beta)'],$$
$$S_\beta\equiv S_{\beta,0}+T_\beta,$$
$$\Omega_\beta\equiv\varepsilon E[\Delta_0]\left(D_\beta'S_\beta^{-1}D_\beta\right)^{-1},$$

we have for both FIML and IOML, $\tilde\beta-\beta_0=0$, $T_\beta=0$, and $D_\beta=-S_{\beta,0}$.

$$\begin{aligned}c_2(z_1|z_0,\beta)=&\frac{\theta^2\left(3\sigma^4-6\theta\sigma^2z_0^2+\theta^2z_0^4\right)}{72\sigma^4}+\frac{\theta^3z_0\left(-3\sigma^2+\theta z_0^2\right)z_1}{36\sigma^4}\\&+\frac{\theta^3\left(-2\sigma^2+\theta z_0^2\right)z_1^2}{24\sigma^4}+\frac{\theta^4z_0z_1^3}{36\sigma^4}+\frac{\theta^4z_1^4}{72\sigma^4},\\c_3(z_1|z_0,\beta)=&-\frac{\theta^3\left(135\sigma^6-27\theta\sigma^4z_0^2-45\theta^2\sigma^2z_0^4+5\theta^3z_0^6\right)}{6480\sigma^6}\\&-\frac{\theta^4\left(-27\sigma^4z_0-30\theta\sigma^2z_0^3+5\theta^2z_0^5\right)z_1}{2160\sigma^6}\\&-\frac{\theta^4\left(-9\sigma^4-45\theta\sigma^2z_0^2+10\theta^2z_0^4\right)z_1^2}{2160\sigma^6}-\frac{\theta^5z_0\left(-18\sigma^2+7\theta z_0^2\right)z_1^3}{1296\sigma^6}\\&-\frac{\theta^5\left(-3\sigma^2+2\theta z_0^2\right)z_1^4}{432\sigma^6}-\frac{\theta^6z_0z_1^5}{432\sigma^6}-\frac{\theta^6z_1^5}{1296\sigma^6}.\end{aligned}$$

TABLE IV

ORNSTEIN-UHLENBECK PROCESS ESTIMATED WHILE PRETENDING THAT Δ IS FIXED

	PFML (Δ Pretended Fixed)
D_θ	$-\varepsilon\left(\frac{E[\Delta_0]}{2\theta_0}\right)+O(\varepsilon^3)$
$S_{\theta,0}$	$\varepsilon\left(\frac{E[\Delta_0]}{2\theta_0}\right)-\varepsilon^2\left(\frac{\text{var}[\Delta_0]}{2}\right)+O(\varepsilon^3)$
T_θ	$\varepsilon^2(\text{var}[\Delta_0])+O(\varepsilon^3)$
Ω_θ	$2\theta_0+\varepsilon\left(\frac{2\theta_0^2\text{var}[\Delta_0]}{E[\Delta_0]}\right)+O(\varepsilon^2)$
$\bar\theta-\theta_0$	$-\varepsilon^2\left(\frac{\theta_0^3\text{var}[\Delta_0]}{6}\right)+O(\varepsilon^2)$
D_{σ^2}	$-\varepsilon\left(\frac{1}{2\sigma_0^4}\right)-\varepsilon^2\left(\frac{\theta_0\text{var}[\Delta_0]}{2\sigma_0^4E[\Delta_0]}\right)+O(\varepsilon^3)$
$S_{\sigma^2,0}$	$\varepsilon^2\left(\frac{2E[\Delta_0]^2+3\text{var}[\Delta_0]}{4\sigma_0^4E[\Delta_0]^2}\right)+\varepsilon^3\left(\frac{\theta_0(3E[\Delta_0]^4-3E[\Delta_0]E[\Delta_0^3]+5E[\Delta_0]^2\text{var}[\Delta_0]+6\text{var}[\Delta_0]^2)}{4\sigma_0^4E[\Delta_0]^3}\right)+O(\varepsilon^4)$
T_{σ^2}	$\varepsilon^3\left(\frac{\theta_0\text{var}[\Delta_0](4E[\Delta_0]^2+\text{var}[\Delta_0])}{8\sigma_0^4E[\Delta_0]^3}\right)+O(\varepsilon^4)$
Ω_{σ^2}	$\varepsilon\left(\frac{\sigma_0^4(4E[\Delta_0]^3+6E[\Delta_0]\text{var}[\Delta_0])}{2E[\Delta_0]^2}\right)+\varepsilon^2\left(\frac{\theta_0\sigma_0^4(6E[\Delta_0]^4-6E[\Delta_0]E[\Delta_0^3]+6E[\Delta_0]^2\text{var}[\Delta_0]+\text{var}[\Delta_0]^2)}{2E[\Delta_0]^2}\right)+O(\varepsilon^3)$
$\bar\sigma^2-\sigma_0^2$	$-\varepsilon\left(\frac{\theta_0\sigma_0^2\text{var}[\Delta_0]}{2E[\Delta_0]}\right)+O(\varepsilon^2)$

Notes: These expressions follow from applying Theorems 4 and 5 to the Ornstein-Uhlenbeck process. When estimating σ^2 with known θ using the Δ pretended fixed method, recall that h_2 is premultiplied by ε; hence h_2 is no longer a likelihood approximation, and D_{σ^2} and $S_{\sigma^2,0}$ are no longer of the same order, as seen in the table. This has no incidence on the asymptotic variance Ω_{σ^2} and the bias $\bar\sigma^2-\sigma_0^2$. Finally, if we use the PFML estimator, and Δ is actually not random, then this method reduces to FIML. This can be seen in our expressions, which indeed reduce to those of the FIML case when $\text{var}[\Delta_0]=0$ and $E[\Delta_0^i]=E[\Delta_0]^i=\Delta_0^i$ (see Table III).

When estimating θ, we consider a design where $T=25$ years and $E[\Delta]=1$ day (Figure 6). When estimating σ^2 we consider both a low frequency design where $T=10$ years and $E[\Delta]=1$ day (Figure 7) and a high frequency design where $T=1$ month and $E[\Delta]=1/100$ day (Figure 8). With high frequency (i.e., small $E[\Delta]$), small span (i.e., low T) data, we only consider the estimation of σ^2, since the drift parameter can only be estimated accurately over a long time interval T. In general, the convergence of the small sample distribution to its asymptotic limit is slower for θ than for σ^2. Table V reports the first four moments of the small sample distributions, compared to their asymptotic limits. Overall, the simulations show that the asymptotic distributions are a fairly accurate guide to the small sample distributions.

8. CONCLUSIONS

We have developed a set of tools for analyzing the effect on estimators of discretely-sampled continuous-time diffusions of random sampling. By Taylor-expanding the different matrices involved in the asymptotic distribution of the estimators, we are able to deliver fully explicit expressions of the variance, and, when applicable, the bias of the asymptotic properties of three likelihood-based

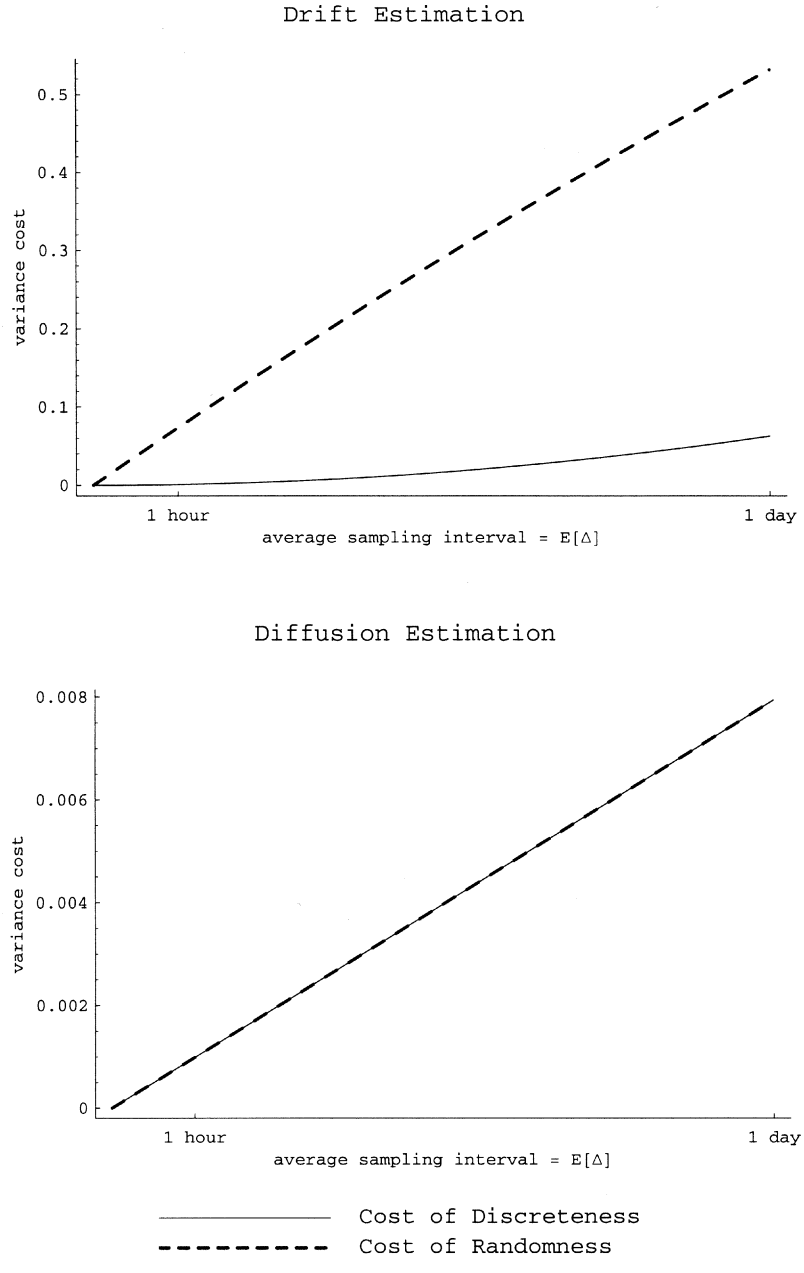


FIGURE 5.—The cost of discreteness compared to the cost of randomness.

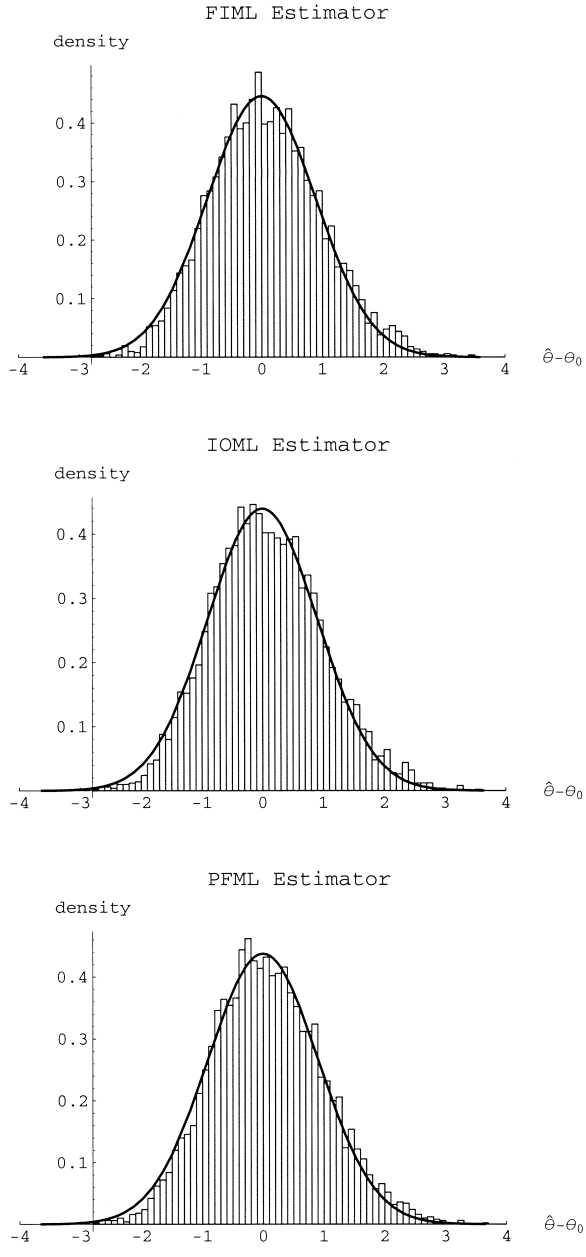


FIGURE 6.— Small sample and asymptotic distributions for the drift estimators, exponential sampling, $T = 25$ years, $E[\Delta] = 1$ day.

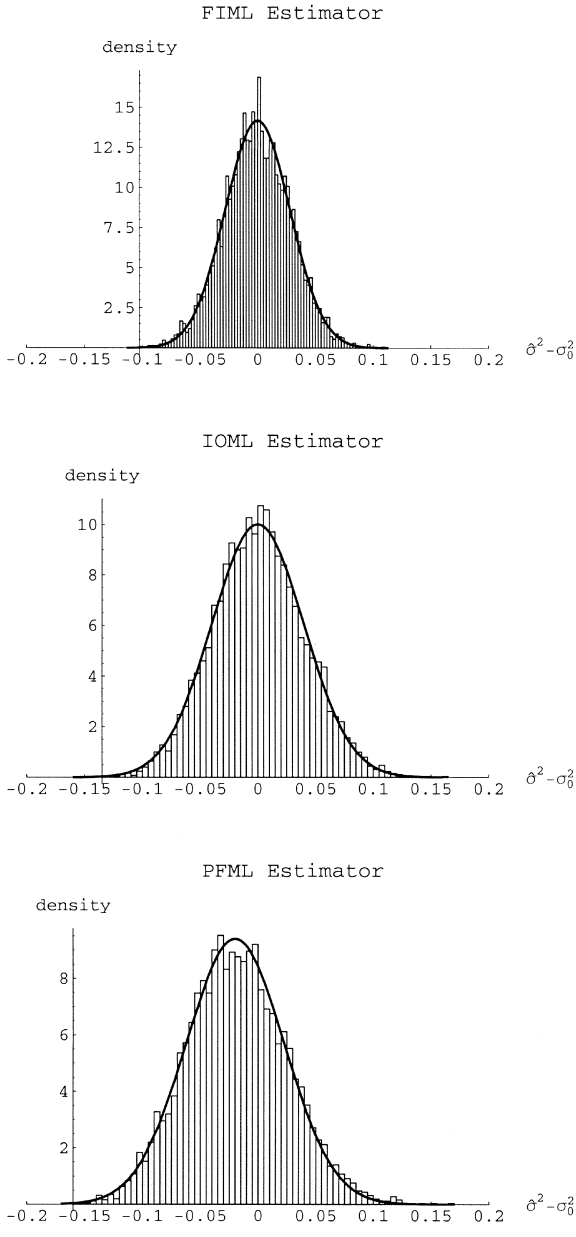


FIGURE 7.— Small sample and asymptotic distributions for the diffusion estimators, exponential sampling, $T = 10$ years, $E[\Delta] = 1$ day.

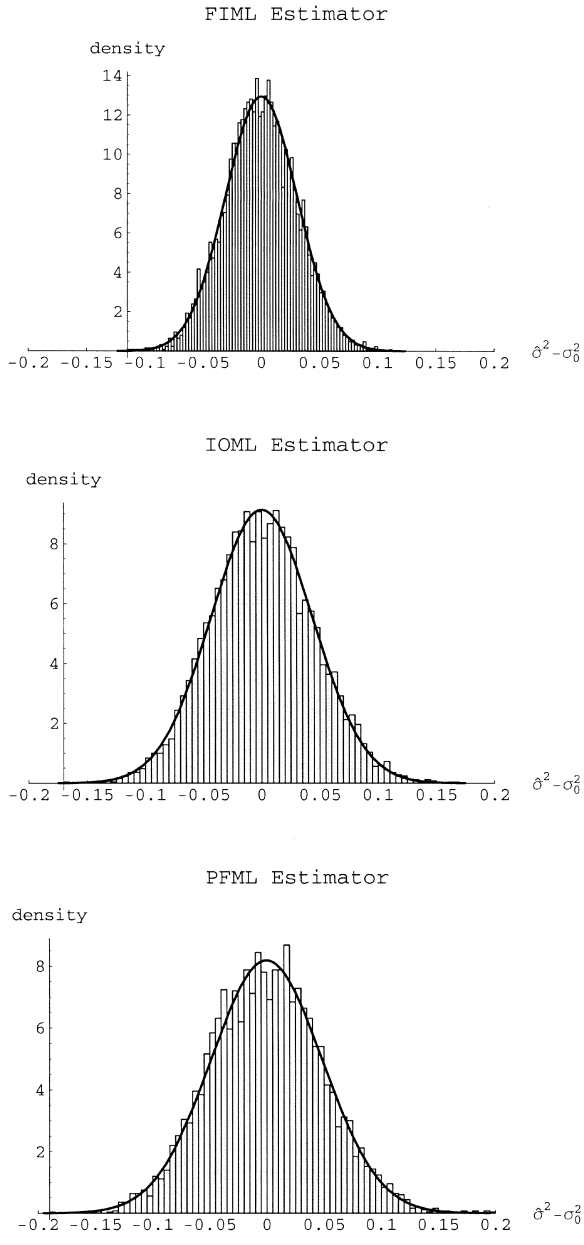


FIGURE 8.— Small sample and asymptotic distributions for the diffusion estimators, exponential sampling, $T = 1$ month, $E[\Delta] = 1/100$ day.

TABLE V

MONTE CARLO SIMULATIONS

Design	Parameter	Estimator	Mean	Std. Dev.	Skewness	Kurtosis
		FIML	0.08	0.91	0.26	3.13
$T = 25$			(0)	(0.90)	(0)	(3)
years						
	Drift	IOML	0.08	0.92	0.25	3.10
	$\hat{\theta} - \theta_0$		(0)	(0.91)	(0)	(3)
$E[\Delta] = 1$						
day		PFML	0.08	0.92	0.25	3.09
			(−0.003)	(0.91)	(0)	(3)
		FIML	−0.00014	0.0286	−0.0056	3.061
$T = 10$			(0)	(0.0282)	(0)	(3)
years						
	Diffusion	IOML	0.00081	0.0397	0.11	3.006
	$\hat{\sigma}^2 - \sigma_0^2$		(0)	(0.0398)	(0)	(3)
$E[\Delta] = 1$						
day		PFML	−0.019	0.0437	0.14	3.04
			(−0.018)	(0.0426)	(0)	(3)
		FIML	−0.0003	0.0308	0.02	3.03
$T = 1$			(0)	(0.0308)	(0)	(3)
month						
	Diffusion	IOML	0.0006	0.044	0.09	2.98
	$\hat{\sigma}^2 - \sigma_0^2$		(0)	(0.043)	(0)	(3)
$E[\Delta] = 1/100$						
day		PFML	−0.00006	0.049	0.13	3.08
			(−0.0002)	(0.049)	(0)	(3)

Notes: This table reports for the centered parameter estimates the first four moments of the small sample distribution and, below each small sample moment, in parentheses, its corresponding asymptotic value. The small sample moments are estimated by the sample statistics from the 5,000 Monte Carlo simulations. Closed-form expressions for the asymptotic moments are given in Tables III and IV. Std. Dev. refers to the standard deviation of the respective distributions.

estimators. By comparing the performance of these three estimators, we assessed the relative costs of ignoring the discreteness of the sampling (i.e., using the continuous-time asymptotics when the data are in fact discrete) versus those of not using the randomly-spaced sampling intervals, or simply ignoring their randomness (by pretending that the sampling is done at a fixed time interval). We found that estimators of the drift and diffusion coefficients not only behave very differently when the data are discretely sampled, but sampling randomness also affects these estimators differently.

We documented the relative magnitudes of the costs imputable to the sampling randomness and to the sampling discreteness respectively. The most common occurrence is one where the loss of efficiency due to not using the sampling times is of the same order in ε as the loss due to the discreteness of the data itself, but can be made an arbitrary large multiple of the cost of discreteness. The main implication of our results for empirical researchers using randomly spaced data is that they should pay as much attention, if not more, to sampling randomness as they do to sampling discreteness.

Market microstructure effects have the potential to make the univariate diffusion hypothesis unrealistic. Ideally, one should model separately the bid and ask prices, which cannot be done seriously without involving the quoted depth of these prices and the transaction volume as explanatory variables as well. Future price changes and durations then need to be related to the actual size of the last transaction. Finally, the arrival of new information could be better modeled in terms of jumps instead of, or in addition to, diffusive Brownian noise. The last area is one where our methodology could be most immediately extended. The notion of a generalized infinitesimal generator, which helped us obtain Taylor expansions for our expectations under random sampling, can be readily extended to other types of Markov processes, such as Lévy processes driven by a non-Brownian process, or in addition to the Brownian process, by replacing the diffusive standard generator in our generalized generator with its Lévy counterpart. The net result in the latter case would be to add a difference term (the expected value over the jump distribution of the difference between the function evaluated after and before the jump) to the differential terms already present in the generator. Our notion of generalized infinitesimal generator would be unchanged and the rest of our analysis would carry through with relatively minor modifications.

Department of Economics and Bendheim Center for Finance, Princeton University, Princeton, NJ 08544-1021, U.S.A., and N.B.E.R.; yacine@princeton.edu; http://www.princeton.edu/~yacine,

and

Department of Statistics, The University of Chicago, Chicago, IL 60637-1514, U.S.A.; mykland@galton.uchicago.edu; http://galton.uchicago.edu/~mykland.

Manuscript received October, 2000; final revision received June, 2002.

APPENDIX: PROOFS

A. *Proof of Proposition 1*

We want to prove that $(Y_k)_{k\geq n}$ is conditionally independent of $(Y_k, \Delta_k)_{k<n}$ given $(Y_k)_{k<n}$. Set

$$Z_{n-1} \equiv (Y_k, \Delta_k)_{k<n} = (Y_1, \ldots, Y_{n-1}, \Delta_1, \ldots, \Delta_{n-1}),$$

and consider first a one step prediction, in which case we need to prove that

$$\Pr(Y_n \in A | Z_{n-1}) = \Pr(Y_n \in A | Y_1, \ldots, Y_{n-1}).$$

X is strong Markov and unique in law. Because Δ_n is drawn conditionally upon the past data Z_{n-1}, as specified in Assumption 1, it follows that Δ_n is conditionally independent of $(X_{\tau_{n-1}+t}, t \geq 0)$ given the past data Z_{n-1}. Hence the distribution of the process $(X_{\tau_{n-1}+t})_{t\geq 0}$ given (Z_{n-1}, Δ_n) only depends on the value of $X_{\tau_{n-1}} = Y_{n-1}$. In particular, the distribution of $Y_n = X_{\tau_n} = X_{\tau_{n-1}+\Delta_n}$ given (Z_{n-1}, Δ_n) only depends on the values of Y_{n-1} and Δ_n:

$$\Pr(Y_n \in A | Z_{n-1}, \Delta_n) = \Pr(Y_n \in A | Y_{n-1}, \Delta_n).$$

Meanwhile, 1(ii) assures that the law of Δ_n given Z_{n-1} only depends on the value of Y_{n-1}. Therefore, if we take the conditional expectation of the above equation given Z_{n-1}, we get

$$\begin{aligned}\text{(A.1)}\qquad \Pr(Y_n \in A|Z_{n-1}) &= E[\Pr(Y_n \in A|Z_{n-1}, \Delta_n)|Z_{n-1}] \\ &= E[\Pr(Y_n \in A|Y_{n-1}, \Delta_n)|Z_{n-1}] \\ &= \Pr(Y_n \in A|Y_{n-1}).\end{aligned}$$

Again by iterated conditional expectation,

$$\begin{aligned}\text{(A.2)}\qquad \Pr(Y_n \in A|Y_1, \ldots, Y_{n-1}) &= E[\Pr(Y_n \in A|Z_{n-1})|Y_1, \ldots, Y_{n-1}] \\ &= E[\Pr(Y_n \in A|Y_{n-1})|Y_1, \ldots, Y_{n-1}] \\ &= \Pr(Y_n \in A|Y_{n-1}).\end{aligned}$$

By equating the right-hand sides of (A.1) and (A.2), we obtain

$$\Pr(Y_n \in A|Z_{n-1}) = \Pr(Y_n \in A|Y_1, \ldots, Y_{n-1}),$$

which is the desired result.

For the finite multi-step prediction problem, let g be a bounded measurable function. Then we have for $k \geq 2$,

$$\begin{aligned}E[g(Y_n, \ldots, Y_{n+k})|Z_{n-1}] &= E[E[g(Y_n, \ldots, Y_{n+k})|Z_{n+k-1}]|Z_{n-1}] \\ &= E[h(Y_n, \ldots, Y_{n+k-1})|Z_{n-1}],\end{aligned}$$

where $h(Y_n, \ldots, Y_{n+k-1}) \equiv E[g(Y_n, \ldots, Y_{n+k})|Y_n, \ldots, Y_{n+k-1}]$. This is the case since

$$E[g(Y_n, \ldots, Y_{n+k})|Z_{n+k-1}] = E[g(Y_n, \ldots, Y_{n+K})|Y_1, \ldots, Y_{n+k-1}]$$

from the one-step prediction result above. It then follows by induction that $Y_n, \ldots, Y_{n+k}$ is conditionally independent of Z_{n-1} given $Y_1, \ldots, Y_{n-1}$. Finally, by letting $k \to \infty$, the smallest σ-field generated by all Y_{n+k}, $k \geq 0$, satisfies the same conditional independence.

B. *Proof of Theorem 1*

We wish to prove that (3.13) is the Taylor expansion of the expectation of a function $f(Y_1, Y_0, \Delta, \beta, \varepsilon)$. All the expectations below are taken with respect to the law of the process at the true value β_0 so that

$$E_{Y_1}[f(Y_1, Y_0, \Delta, \beta, \varepsilon)|Y_0 = y_0, \Delta = \delta] = \int_{\underline{x}}^{\bar{x}} f(y_1, y_0, \delta, \beta, \varepsilon) p(y_1|y_0, \delta, \beta_0)\, dy_1.$$

By the standard (deterministic) Taylor formula applied to the function of ε defined by

$$\phi(\varepsilon) \equiv E_{\Delta, Y_1}[f(Y_1, Y_0, \varepsilon\Delta_0, \beta, \varepsilon)|Y_0 - y_0],$$

that is, $\phi(\varepsilon) = \phi(0) + \varepsilon\phi'(0) + O(\varepsilon^2)$, we have that

$$\begin{aligned}\text{(B.1)}\qquad & E_{\Delta, Y_1}[f(Y_1, Y_0, \Delta, \beta, \varepsilon)|Y_0 = y_0] \\ &= f(y_0, y_0, 0, \beta_0, 0) + \varepsilon\Bigg(E[\Delta_0|Y_0 = y_0]\big[A_{\beta_0} \cdot f(y_1, y_0, \delta, \beta, 0)\big]\big|_{y_1 = y_0,\, \delta = 0,\, \beta = \beta_0,\, \varepsilon = 0} \\ &\qquad + \frac{\partial f(y_1, y_0, \delta, \beta, \varepsilon)}{\partial \varepsilon}_{|y_1 = y_0,\, \delta = 0,\, \beta = \beta_0,\, \varepsilon = 0} \\ &\qquad + \frac{\partial f(y_1, y_0, \delta, \beta, \varepsilon)}{\partial \beta}_{|y_1 = y_0,\, \delta = 0,\, \beta = \beta_0,\, \varepsilon = 0} \frac{\partial \beta}{\partial \varepsilon}(\beta_0, 0)\Bigg) + O(\varepsilon^2).\end{aligned}$$

The last term allows for the fact that β may depend on ε (this would be the case for instance if f were computed at $\beta = \bar{\beta}$, where $\bar{\beta}$ is the probability limit of the estimator calculated from discrete data defined in (2.4)). Note that when β is a function of ε, its limit as ε tends to 0 is β_0.

We can rewrite (B.l) more compactly as

$$E_{\Delta, Y_1}[f(Y_1, Y_0, \Delta, \bar{\beta}, \varepsilon)|Y_0 = y_0] = f(y_0, y_0, 0, \beta_0, 0) + \varepsilon\big(\Gamma_{\beta_0} \cdot f\big)(Y_0, Y_0, 0, \beta_0, 0) + O(\varepsilon^2).$$

Indeed, we created the operator Γ_{β_0} in (3.11) in such a way that its iterated action will produce a Taylor expansion in ε of the function up to any order in ε as long as the partial derivatives of the function f exist. To see why, compute

$$\begin{aligned}
\Gamma_{\beta_0}^2 \cdot f &= \Gamma_{\beta_0} \cdot \big(\Gamma_{\beta_0} \cdot f\big) \\
&= \Gamma_{\beta_0} \cdot \left(\Delta_0 A_{\beta_0} \cdot f + \frac{\partial f}{\partial \varepsilon} + \frac{\partial f}{\partial \beta}\frac{\partial \beta}{\partial \varepsilon}\right) \\
&= \Delta_0^2 A_{\beta_0}^2 \cdot f + \frac{\partial^2 f}{\partial \varepsilon^2} + \frac{\partial f}{\partial \beta}\frac{\partial^2 \beta}{\partial \varepsilon^2} + \frac{\partial^2 f}{\partial \beta^2}\left(\frac{\partial \beta}{\partial \varepsilon}\right)^2 + 2\Delta_0 A_{\beta_0} \cdot \left(\frac{\partial f}{\partial \varepsilon} + \frac{\partial f}{\partial \beta}\frac{\partial \beta}{\partial \varepsilon}\right) + 2\frac{\partial^2 f}{\partial \beta \partial \varepsilon}\frac{\partial \beta}{\partial \varepsilon},
\end{aligned}$$

which is the full second order term in the Taylor expansion of f with respect to all its arguments $(y_1, y_0, \delta, \beta, \varepsilon)$, given that $\Delta_0^2 A_{\beta_0}^2 \cdot f$ is the second order term in the Taylor expansion of f with respect to its arguments (y_1, δ) as given in (3.10). Further iterations of the operator Γ_{β_0} will produce higher order Taylor expansions of f in ε, as long as the necessary derivatives of f exist. The unconditional expectation (3.13) follows from (B.1) by taking expectations with respect to Y_0 using the Law of Iterated Expectations.

C. *Proof of Proposition 2*

Consider the FIML likelihood ratio of β with respect to β_0 using both sets of data Y_n and Δ_n, that is,

$$\begin{aligned}
&LR\Big(Y_{N_T}, \Delta_{N_T}, \ldots, Y_n, \Delta_n, Y_{n-1}, \Delta_{n-1}, \ldots, Y_1, \Delta_1 | X_0\Big) \\
&\quad = \frac{L\Big(Y_{N_T}, \Delta_{N_T}, \ldots, Y_n, \Delta_n, Y_{n-1}, \Delta_{n-1}, \ldots, Y_1, \Delta_1 | X_0; \beta\Big)}{L\Big(Y_{N_T}, \Delta_{N_T}, \ldots, Y_n, \Delta_n, Y_{n-1}, \Delta_{n-1}, \ldots, Y_1, \Delta_1 | X_0; \beta\Big)} \\
&\quad = \frac{\prod_{n=1}^{N_T} p(Y_n | Y_{n-1}, \Delta_n, \beta) d(\Delta_n | Y_{n-1})}{\prod_{n=1}^{N_T} p(Y_n | Y_{n-1}, \Delta_n, \beta_0) d(\Delta_n | Y_{n-1})} \\
&\quad = \frac{\prod_{n=1}^{N_T} p(Y_n | Y_{n-1}, \Delta_n, \beta)}{\prod_{n=1}^{N_T} p(Y_n | Y_{n-1}, \Delta_n, \beta)}.
\end{aligned}$$

Recall that we use L as a generic notation for the density of the variables listed as its arguments. The second equality is due to (4.1), and the simplification in the third equality follows from the fact that the density d does not depend on the parameter β. With (4.11), we have

$$LR\Big(Y_{N_T}, \Delta_{N_T}, \ldots, Y_n, \Delta_n, Y_{n-1}, \Delta_{n-1}, \ldots, Y_1, \Delta_1 | X_0\Big) = \prod_{n=1}^{N_T} LR(Y_n | Y_{n-1}, \Delta_n).$$

Now consider the likelihood ratio using only the Y_n's, which by the martingale property of likelihood ratios is given by (4.10). Because Y_n is a Markov chain under both probabilities involved in the likelihood ratio, the likelihood ratio using only the Y_n's is the product of the pairwise conditional likelihood ratios

$$\text{(C.1)} \qquad \widetilde{LR}\Big(Y_{N_T}, \ldots, Y_1 | X_0\Big) = \prod_{n=1}^{N_T} \widetilde{LR}(Y_n | Y_{n-1}).$$

This (or its log) is the criterion function to be maximized over β to obtain the IOML estimator.

$\widetilde{LR}(Y_n|Y_{n-1})$ can be expressed as

$$\begin{aligned}
\widetilde{LR}(Y_n|Y_{n-1}) &= E_{\Delta_n}[LR(Y_n|Y_{n-1},\Delta_n)|Y_n, Y_{n-1}] \\
&= \int LR(Y_n|Y_{n-1},\Delta_n)L(\Delta_n|Y_n, Y_{n-1};\beta_0)\,d\Delta_n \\
&= \int \frac{p(Y_n|Y_{n-1},\Delta_n,\beta)}{p(Y_n|Y_{n-1},\Delta_n,\beta_0)} L(\Delta_n|Y_n, Y_{n-1};\beta_0)\,d\Delta_n \\
&= \int p(Y_n|Y_{n-1},\Delta_n,\beta)\frac{L(Y_{n-1},\Delta_n;\beta_0)}{L(Y_n, Y_{n-1},\Delta_n;\beta_0)}\frac{L(Y_n, Y_{n-1},\Delta_n;\beta_0)}{L(Y_n, Y_{n-1},\beta_0)}\,d\Delta_n \\
&= \int p(Y_n|Y_{n-1},\Delta_n,\beta)\frac{L(\Delta_n|Y_{n-1};\beta_0)L(Y_{n-1};\beta_0)}{L(Y_n|Y_{n-1};\beta_0)L(Y_{n-1};\beta_0)}\,d\Delta_n \\
&= \frac{1}{L(Y_n|Y_{n-1};\beta_0)}\int p(Y_n|Y_{n-1},\Delta_n,\beta)d(\Delta_n|Y_{n-1})\,d\Delta_n. \qquad \text{(C.2)}
\end{aligned}$$

Since $\tilde{p}$ is defined by the property (4.12), namely

$$\tilde{p}(Y_n|Y_{n-1},\beta) = \int p(Y_n|Y_{n-1},\Delta_n,\beta)d(\Delta_n|Y_{n-1})\,d\Delta_n,$$

one can take $\beta=\beta_0$ in $\widetilde{LR}(Y_n|Y_{n-1})$, which makes the ratio one by definition. This shows that

$$\tilde{p}(Y_n|Y_{n-1},\beta_0) = L(Y_n|Y_{n-1};\beta_0). \qquad \text{(C.3)}$$

However, there is nothing in the definition of $\widetilde{LR}$ that requires β_0 to be the true parameter. Hence (C.3) is in fact true for all β, not just the true parameter β_0. Therefore $\tilde{p}(Y_1|Y_0,\beta)$ is the conditional density of Y_1 given Y_0.

D. *Proof of Proposition 3*

We have that

$$\int_0^{+\infty} \exp\left(-\frac{\gamma_{-1}}{\Delta}+\gamma_0\Delta\right)\Delta^{j-1/2}\,d\Delta = 2\left(\frac{\gamma_{-1}}{\gamma_0}\right)^{(2j+1)/4} B_{j+(1/2)}(2\sqrt{\gamma_{-1}\gamma_0}) \qquad \text{(D.1)}$$

where γ_{-1} and γ_0 are positive constants. Then with

$$\gamma_{-1} = \frac{(Y_1-Y_0)^2}{2\sigma^2}, \qquad \gamma_0 = \frac{1}{E[\Delta|Y_0]},$$

the result for $\tilde{p}^{(J)}$ follows from an integration term by term of $\int_0^{+\infty} p^{(J)}(Y_1|Y_0,\delta,\beta)d(\delta|Y_0)\,d\delta$, with $p^{(J)}$ given by equation (4.5).

E. *Proof of Proposition 4*

Given (4.13), the IOML likelihood ratio

$$\widetilde{LR}(y_1|y_0) = \frac{\tilde{p}(y_1|y_0,\beta)}{\tilde{p}(y_1|y_0,\beta_0)} = \exp\{\lambda(y_1|y_0,\beta)-\lambda(y_1|y_0,\beta_0)\}$$

is related to the FIML likelihood ratio

$$LR(y_1|y_0,\delta) = \frac{p(y_1|y_0,\delta,\beta)}{p(y_1|y_0,\delta,\beta_0)} = \exp\{l(y_1|y_0,\delta,\beta)-l(y_1|y_0,\delta,\beta_0)\}$$

by equation (4.10), namely

$$\widetilde{LR}(Y_1|Y_0) = E_\Delta[LR(Y_1|Y_0,\Delta)|Y_1, Y_0]$$

with the expectation evaluated, as always, over the distribution of Δ given (Y_1, Y_0) evaluated at the true parameter values.

That is

$$\begin{aligned}\phi(\beta) &\equiv \exp\{\lambda(y_1|y_0,\beta) - \lambda(y_1|y_0,\beta)\}\\ &= E_\Delta[\exp\{l(Y_1|Y_0,\Delta,\beta) - l(Y_1|Y_0,\Delta,\beta_0)\}|Y_1 = y_1, Y_0 = y_0]\end{aligned}$$

which we first differentiate with respect to β to obtain

$$\begin{aligned}\dot\phi(\beta) &\equiv \dot\lambda(y_1|y_0,\beta)\phi(\beta)\\ &= E_\Delta\big[\dot l(Y_1|Y_0,\Delta,\beta)\exp\{l(Y_1|Y_0,\Delta,\beta) - l(Y_1|Y_0,\Delta,\beta_0)\}|Y_1 = y_1, Y_0 = y_0\big],\end{aligned}$$

then evaluate at $\beta = \beta_0$ to obtain

$$\dot\phi(\beta_0) \equiv \dot\lambda(y_1|y_0,\beta_0) = E_\Delta\big[\dot l(Y_1|Y_0,\Delta,\beta)|Y_1 = y_1, Y_0 = y_0\big],$$

which is (4.19).

A second differentiation with respect to β, followed by an evaluation of the result at $\beta = \beta_0$, yields

$$\begin{aligned}&\ddot\lambda(y_1|y_0,\beta_0) + \dot\lambda^2(y_1|y_0,\beta_0)\\ &\quad = E_\Delta\Big[\ddot l(Y_1|Y_0,\Delta,\beta_0) + \dot l(Y_1|Y_0,\Delta,\beta_0)\dot l(Y_1|Y_0,\Delta,\beta_0)'|Y_0 = y_0, Y_1 = y_1\Big]\end{aligned} \tag{E.1}$$

from which (4.20) follows.

F. *Proof of Theorem 2*

To calculate the I matrix, we apply Theorem 1 to the stochastic expansion of the log-likelihood in Δ, and calculate each of the terms. This expansion can be obtained in two ways. First, the Hermite expansion (4.5) for $p(Y_1|Y_0,\Delta,\beta)$, although valid for all Δ, can be interpreted as a Taylor expansion in Δ from which an expansion of its ln is obtained by expanding

$$\ln\left(1 + \sum_{j=1}^{J} c_j\left(\frac{Y_1}{\sigma}|\frac{Y_0}{\sigma},\beta\right)\frac{\Delta^j}{j!}\right)$$

in Δ around 0. The result is the Taylor expansion of the log of the transition function $l(Y_1|Y_0,\Delta,\beta) = \ln p(Y_1|Y_0,\Delta,\beta)$ given by

$$\begin{aligned}l^{(3)}(Y_1|Y_0,\Delta,\beta) = &-\frac{1}{2}\ln(2\pi\Delta\sigma^2) - \frac{(Y_1-Y_0)^2}{2\Delta\sigma^2} + \int_{Y_0}^{Y_1}\frac{\mu(z,\theta)}{\sigma^2}\,dz + f_1(Y_1,Y_0,\theta,\sigma)\sigma^2\Delta\\ &+ f_2(Y_1,Y_0,\theta,\sigma)\sigma^4\Delta^2 + f_3(Y_1,Y_0,\theta,\sigma)\sigma^6\Delta^3 + O_p(\varepsilon^4)\end{aligned} \tag{F.1}$$

where the coefficients f_j, $j = 1, 2, \ldots$ are functions of the c_j coefficients in (4.6). They are defined recursively and are all closed-form. Specifically, the first three coefficients, which are the only ones required for the order three Taylor expansions in ε we are calculating, are:

$$\begin{cases} f_1(Y_1,Y_0,\theta,\sigma) = (Y_1-Y_0)^{-1}\int_{Y_0}^{Y_1}\kappa(w,\theta,\sigma^2)\,dw,\\ f_2(Y_1,Y_0,\theta,\sigma) = \frac{f_1(Y_1,Y_0,\theta,\sigma) - f_1(Y_0,Y_0,\theta,\sigma)}{4(Y_1-Y_0)^2} + \frac{1}{4(Y_1-Y_0)}\frac{\partial f_1(Y_1,Y_0,\theta,\sigma)}{\partial Y_1},\\ f_3(Y_1,Y_0,\theta,\sigma) = \frac{1}{12(Y_1-Y_0)^3}\left\{\int_{Y_0}^{Y_1}(w-Y_0)^2\left(\left(\frac{\partial f_1(w,Y_0,\theta,\sigma)}{\partial w}\right)^2 + \frac{\partial^2 f_2(w,Y_0,\theta,\sigma)}{\partial w^2}\right)dw\right\}.\end{cases} \tag{F.2}$$

An alternative, provided by Dacunha-Castelle and Florens-Zmirou (1986), is to apply Girsanov's Theorem and expand the Radon-Nikodym derivative in Δ, with the result expressed as functionals of a Brownian Bridge process.

Functions f that are polynomial in y_1 near the boundaries of $\mathscr{S}$, and their iterates by repeated application of the generator, retain their polynomial growth characteristic near the boundaries; so they are all in L^2 and satisfy (3.8). This follows from the exponential divergence of $s(y_1;\beta)$ near both boundaries whereas polynomials and their iterates diverge at most polynomially under Assumption 3. Indeed, μ and its derivatives have at most polynomial growth; multiplying and adding functions with polynomial growth yields a function still with polynomial growth. The fact that the scale function diverges exponentially is a consequence of the fact that near an ∞ boundary $s(x;\beta) \geq \exp(k|x|^{\gamma})$ for some $k > 0$ and $\gamma \geq 1$, and near the 0^+ boundary $s(x;\beta) \geq \exp(k|x|^{-\gamma})$ again for some $k > 0$ and $\gamma \geq 1$. By the same method as Proposition 2 of Aït-Sahalia (2002), p and its β-derivatives have at most exponential growth and l and its β-derivatives have at most polynomial growth. Thus $\ddot{l}$ will be in $\mathscr{D}^K$.

The reason we can replace $l = \ln p$ by $\ln p^{(J)}$ (and similarly for the first three derivatives with respect to β) follows from (4.7), and the similar result for the β-derivatives. By applying Theorem 1, the asymptotic distribution of FIML can be obtained in Taylor series form starting with an expansion at order K of the inverse of

$$I = -E_{\Delta_0, Y_1, Y_0}\left[\ddot{l}(Y_1|Y_0, \varepsilon\Delta_0, \beta)\right] = I_\beta^{(0)} + I_\beta^{(1)}\varepsilon + \cdots + I_\beta^{(K-1)}\varepsilon^{K-1} + O(\varepsilon^K).$$

Suppose now that we apply the same calculation to the Taylor expansion of the likelihood $l^{(J)}$ (which is also in $\mathscr{D}^K$) instead of l, yielding

$$I^{(J)} = -E_{\Delta_0, Y_1, Y_0}\left[\ddot{l}^{(J)}(Y_1|Y_0, \varepsilon\Delta_0, \beta)\right] = I_\beta^{(J,0)} + I_\beta^{(J,1)}\varepsilon + \cdots + I_\beta^{(J,K-1)}\varepsilon^{K-1} + O(\varepsilon^K).$$

Then, from

$$\left|E_{\Delta_0, Y_1, Y_0}\left[\ddot{l}(Y_1|Y_0, \varepsilon\Delta_0, \beta)\right] - E_{\Delta_0, Y_1, Y_0}\left[\ddot{l}^{(J)}(Y_1|Y_0, \varepsilon\Delta_0, \beta)\right]\right| = O(\varepsilon^J),$$

it follows that if the order of the approximation of p is high enough ($J \geq K$), then $I^{(J,k)} = I^{(k)}$ for all $k = 0, 1, \ldots, K-1$. In other words, replacing p (resp. l) by $p^{(J)}$ (resp. $l^{(J)}$), or their first two derivatives, for the purpose of calculating a Taylor expansion of the FIML distribution at order K in ε is inconsequential.

From (F.1), we calculate $\dot{l}^{(3)}(Y_1|Y_0, \Delta, \beta)$ and then $\ddot{l}^{(3)}(Y_1|Y_0, \Delta, \beta)$. The two score components in the vector $\dot{l}^{(3)}$ are, for the drift parameters θ,

$$\begin{aligned} \dot{l}_\theta^{(3)}(Y_1, Y_0, \Delta, \beta) = &\int_{Y_0}^{Y_1} \frac{\dot{\mu}(z,\theta)}{\sigma^4}\,dz + \dot{f}_{1,\theta}(Y_1, Y_0, \theta, \sigma)\sigma^2\Delta \\ &+ \dot{f}_{2,\theta}(Y_1, Y_0, \theta, \sigma)\sigma^4\Delta^2 + \dot{f}_{3,\theta}(Y_1, Y_0, \theta, \sigma)\sigma^6\Delta^3 + O_p(\varepsilon^4), \end{aligned}$$

and, for the diffusion parameter σ^2,

$$\begin{aligned} \dot{l}_{\sigma^2}^{(3)}(Y_1, Y_0, \Delta, \beta) = &-\frac{1}{2\sigma^2} + \frac{(Y_1 - Y_0)^2}{2\Delta\sigma^2} - \int_{Y_0}^{Y_1} \frac{\mu(z,\theta)}{\sigma^4}\,dz \\ &+ \frac{\partial}{\partial\sigma^2}[f_1(Y_1, Y_0, \theta, \sigma)\sigma^2\Delta + f_2(Y_1, Y_0, \theta, \sigma)\sigma^4\Delta^2 \\ &+ f_3(Y_1, Y_0, \theta, \sigma)\sigma^6\Delta^3] + O_p(\varepsilon^4). \end{aligned}$$

Finally, the Taylor expansion for

$$I = -E_{\Delta, Y_1, Y_0}\left[\ddot{l}(Y_1|Y_0, \Delta, \beta)\right] = E_{\Delta, Y_1, Y_0}\left[\dot{l}(Y_1|Y_0, \Delta, \beta)\dot{l}(Y_1|Y_0, \Delta, \beta)'\right]$$

is calculated by applying Theorem 1 to the function $\dot{l}^{(3)} \times \dot{l}^{(3)\prime}$ (or equivalently to the function $\ddot{l}^{(3)}$) instead of $\dot{l} \times \dot{l}'$ (or $\ddot{l}$). The result of these lengthy computations is given in (5.1)–(5.2).

G. *Proof of Theorem 3*

We define J to be the expected covariance matrix

$$J = E_{Y_0, Y_1}\left[\mathrm{var}_\Delta\left[\dot{l}(Y_1|Y_0, \Delta, \beta)|Y_0, Y_1\right]\right]. \tag{G.1}$$

We will apply Theorem 1 to calculate the leading terms of a Taylor expansion of the matrix J, which requires an argument substantially more involved than that of the proof of Theorem 2. As we will show below, these terms are functions of $f_1(Y_1, Y_0, \theta, \sigma)$ in (F.1), which is given in (F.2). As in the proof of Theorem 2, the actual calculations are all performed by replacing l (resp. $\dot{l}$ and $\ddot{l}$) by $l^{(3)}$ (resp. $\dot{l}^{(3)}$ and $\ddot{l}^{(3)}$) but from now on we will not make a distinction between the true function and its order 3 Taylor expansion since as shown above the difference is irrelevant as far as computing expansions of order 3 or below of the moments we need.

Write $Z = (Y_1 - Y_0)/(\sigma\sqrt{\varepsilon})$, and note that

$$\dot{l}_\theta = \dot{f}_{1,\theta}\sigma^2\varepsilon\Delta_0 + \text{ terms that do not depend on } \Delta_0 + O_p(\varepsilon^2) \tag{G.2}$$

where $\dot{f}_{1,\theta} = \partial f_1(Y_0 + \sigma\sqrt{\varepsilon}Z, Y_0, \theta, \sigma^2)/\partial\theta$. Similarly,

$$\dot{l}_{\sigma^2} = \frac{Z^2}{2\sigma^2\Delta_0} + \text{ terms that do not depend on } \Delta_0 + O_p(\varepsilon). \tag{G.3}$$

It is apparent from (G.l) that we need to get expressions for $\mathrm{cov}(\Delta_0^p, \Delta_0^q|Y_0, Y_1) = \mathrm{cov}(\Delta_0^p, \Delta_0^q|Z, Y_0)$ in a form where the dependence on ε is explicit. We shall see that

$$\mathrm{cov}_\Delta\left(\Delta_0^p, \Delta_0^q|Y_0, Y_1\right) = \mathrm{cov}_Q\left(\Delta_0^p, \Delta_0^q|Z, Y_0\right) + O_p(\varepsilon) \tag{G.4}$$

where, under Q, the law of $Z|\Delta_0$ is $N(0, \Delta_0)$ and the law of Δ_0 is as under the original distribution. Under (G.4), the first term on the right-hand side in (G.3) is independent of ε and all other quantities in our system except the sampling scheme. It then follows that

$$J_{\sigma^2\sigma^2} = E_{Y_0, Y_1}\left[\mathrm{var}_\Delta\left[\dot{l}_{\sigma^2}|Y_0, Y_1\right]\right] = \frac{1}{4\sigma^4}E_Q\left[Z^4\mathrm{var}_Q\left[\frac{1}{\Delta_0}\middle|Z, Y_0\right]\right] + O_p(\varepsilon).$$

Similarly,

$$J_{\theta\theta} = E_{Y_0, Y_1}\left[\mathrm{var}_\Delta\left[\dot{l}_\theta|Y_0, Y_1\right]\right] = \varepsilon^2\sigma^4 E_{Y_0, Y_1}\left[\left(\dot{f}_{1,\theta}\right)^2\right] \times E_Q[\mathrm{var}_Q[\Delta_0|Z, Y_0]] + O_p(\varepsilon^3)$$

where the method of Section 3.3 yields

$$E_{Y_0, Y_1}\left[\left(\dot{f}_{1,\theta}\right)^2\right] = \frac{1}{4\sigma^6}E_{Y_0}\left[\sigma_0^2\left(\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0, \theta_0)\right)^2 - 2\left(\frac{\partial\mu}{\partial\theta}(Y_0, \theta_0)\right)^2\frac{\partial\mu}{\partial y}(Y_0, \theta_0)\right].$$

Finally,

$$\begin{aligned} J_{\theta\sigma^2} &= E_{Y_0, Y_1}\left[\mathrm{cov}_\Delta\left(\dot{l}_\theta, \dot{l}_{\sigma^2}|Y_0, Y_1\right)\right] = \frac{\varepsilon}{2}E_{Y_0, Y_1}\left[\dot{f}_{1,\theta}\right] \\ &\quad \times \mathrm{cov}_\Delta\left(\Delta_0, \frac{Z^2}{\Delta_0}|Y_0, Y_1\right) + O(\varepsilon^2) = O(\varepsilon^2) \end{aligned}$$

since we obtain that $E_{Y_0, Y_1}\left[\dot{f}_{1,\theta}\right] = O(\varepsilon^2)$ using again Theorem 1. It follows that the J matrix is of the form

$$J = \begin{pmatrix} j_{\theta\theta}^{(2)}\varepsilon^2 + O(\varepsilon^3) & j_{\theta\sigma^2}^{(2)}\varepsilon^2 + O(\varepsilon^3) \\ j_{\theta\sigma^2}^{(2)}\varepsilon^2 + O(\varepsilon^3) & j_{\sigma^2\sigma^2}^{(0)} + O(\varepsilon) \end{pmatrix}.$$

The expressions involving J that we give in (5.4) are obtained by setting $\chi^2 = Z^2/\Delta_0$ above. Obviously, $Z/\sqrt{\Delta_0}$ is standard normal and independent of Δ_0 and Y_0, and hence the distribution and properties of χ^2, a chi-squared variable with one degree of freedom, follow. In particular, $\text{var}[\chi^2] = 2$ and

$$\text{(G.5)} \qquad Z^4 \text{var}_Q\left[\Delta_0^{-1} | Z, Y_0\right] = \text{var}_Q\left[Z^2 \Delta_0^{-1} | Z, Y_0\right] = \text{var}\left[\chi^2 | \chi^2 \Delta_0, Y_0\right].$$

The rationale behind (G.4) is as follows. It is enough to show this for conditional moments, i.e.,

$$\text{(G.6)} \qquad E_\Delta\left[\Delta_0^p | Y_0, Y_1\right] = E_Q\left[\Delta_0^p | Z, Y_0\right] + O_p(\varepsilon)$$

where Q is the distribution under which X_t is a Brownian motion starting at Y_0 and with volatility σ, i.e., $X_t = Y_0 + \sigma \widetilde{W}_t$. To see this, observe that, by Girsanov's Theorem (see, e.g., Theorem III.21 in Protter (1992)), and in analogy with (G.2), the Radon-Nikodym derivative dP/dQ can be written

$$\begin{aligned}
\frac{dP}{dQ} &= E_Q\left[\exp\left\{\frac{1}{\sigma}\int_0^\Delta \mu(X_t, \theta)\, d\widetilde{W}_t - \frac{1}{2\sigma^2}\int_0^\Delta \mu(X_t, \theta)^2\, dt\right\} \middle| Y_0, Z, \Delta\right] \\
&= \exp\left\{\int_{Y_0}^{Y_0 + \sigma\sqrt{\varepsilon}Z} \frac{\mu(z, \theta)}{\sigma^2}\, dz + f_1(Y_0 + \sigma\sqrt{\varepsilon}Z, Y_0, \theta, \sigma)\sigma^2\Delta + O_p(\varepsilon^2)\right\} \\
&= \exp\{\text{terms that do not depend on } \Delta_0 + O_p(\varepsilon)\}.
\end{aligned}$$

Hence

$$\text{(G.7)} \qquad \begin{aligned}
E[\Delta_0^p | Y_0, Y_1] &= \frac{E_Q\left[\Delta_0^p \frac{dQ}{dP} | Y_0, Z\right]}{E_Q\left[\frac{dQ}{dP} | Y_0, Z\right]} = E_Q[\Delta_0^p | Y_0, Z](1 + O_p(\varepsilon)) \\
&= E_Q\left[\Delta_0^p | \Delta_0^{1/2} N(0,1), Y_0\right](1 + O_p(\varepsilon))
\end{aligned}$$

since, under Q, $Z/\sqrt{\Delta}$ is distributed as $N(0,1)$ where the normal variable is independent of Δ_0 and Y_0. This yields (G.4).

To see that the score with respect to σ^2 is Taylor-expandable in ε, note that from (G.7), $\dot{\lambda}_{\sigma^2}$ takes the form

$$\text{(G.8)} \qquad \begin{aligned}
\dot{\lambda}_{\sigma^2} &= E\left[\dot{l}_{\sigma^2} | Y_0, Y_1\right] = \frac{1}{2\sigma^2}\left(E_Q\left[\frac{Z^2}{\Delta_0} \middle| Z, Y_0\right] - 1\right) - \frac{1}{\sigma_0^4}(Y_1 - Y_0)\mu(Y_0, \theta_0) + O_p(\varepsilon) \\
&= \frac{1}{2\sigma_0^2}(E[\chi^2 | \chi^2 \Delta_0, Y_0] - 1) - \frac{1}{\sigma_0^4}(Y_1 - Y_0)\mu(Y_0, \theta_0) + O_p(\varepsilon),
\end{aligned}$$

where χ^2 is independent of Δ_0.

H. *Proof of Proposition 5*

With D_0 denoting the c.d.f. of the distribution of Δ_0,

$$P(\chi^2 \Delta_0 \le b | \chi^2) = D_0\left(\frac{b}{\chi^2}\right),$$

from which it follows that

$$f_{\chi^2 \Delta_0 | \chi^2}(b | \chi^2) = \frac{1}{\chi^2} d_0\left(\frac{b}{\chi^2}\right),$$

and hence

$$f_{\chi^2 \Delta_0, \chi^2}(b, \chi^2) = \frac{1}{\chi^2} d_0\left(\frac{b}{\chi^2}\right) f_{\chi^2}(\chi^2),$$

where $f_{\chi^2\Delta_0|\chi^2}$ is the conditional density of $\chi^2\Delta_0$ given χ^2, and $f_{\chi^2\Delta_0,\chi^2}$ their joint density. Hence

$$E[\chi^{2q}|\chi^2\Delta_0 = b] = \frac{m_{2-2q}(b)}{m_2(b)}. \tag{H.1}$$

Now since

$$E[\mathrm{var}[\chi^2|\chi^2\Delta_0]] = 3 - E[E[\chi^2|\chi^2\Delta_0]^2], \tag{H.2}$$

equation (5.16) follows. Meanwhile,

$$E[\Delta_0|\chi^2\Delta_0 = b] = bE[\chi^{-2}|\chi^2\Delta_0 = b] = b\frac{m_4(b)}{m_2(b)}, \tag{H.3}$$

from which equation (5.17) follows.

I. *Proof of Theorem 4*

Let h_1 denote the first component of the vector h in (4.23). We have for $\beta - \beta_0 = O_p(\varepsilon)$:

$$\begin{aligned} E_{\Delta,Y_1}[h_1(Y_1, Y_0, \beta, \varepsilon)|Y_0] &= E\big[\dot{l}_\theta(Y_1|Y_0, E[\Delta], \theta, \sigma^2)|Y_0\big] \\ &= E\big[\dot{l}_\theta(Y_1|Y_0, \Delta, \theta, \sigma^2)|Y_0\big] \\ &\quad - \varepsilon^2\sigma_0^4\mathrm{var}[\Delta_0]\left(\frac{1}{2\sigma_0^2}\frac{\partial^2\lambda(Y_0,\beta_0)}{\partial y\partial\theta}\mu(Y_0,\theta_0) + \frac{1}{4}\frac{\partial^3\lambda(Y_0,\beta_0)}{\partial Y_0^2\partial\theta}\right) \\ &\quad + O(\varepsilon^3). \end{aligned} \tag{I.1}$$

Note that the ε^2 term on the right-hand side of this equation integrates exactly to zero under the stationary distribution. For the bias computation, with $\bar{\beta} = (\bar{\theta}, \sigma_0^2)'$, apply the mean-value theorem to obtain

$$\bar{\beta} - \beta_0 = -\Big(E_{\Delta,Y_1,Y_0}\big[\dot{h}(Y_1, Y_0, \breve{\beta}, \varepsilon)\big]\Big)^{-1}E_{\Delta,Y_1,Y_0}[h(Y_1, Y_0, \beta_0, \varepsilon)] \tag{I.2}$$

where $\breve{\beta}$ denotes a value between β_0 and $\bar{\beta}$ and $\dot{h}$ denotes differentiation with respect to the parameter vector β. With

$$\bar{\theta} = \theta_0 + \varepsilon b_1 + \varepsilon^2 b_2 + O(\varepsilon^3),$$

recall that

$$E_{\Delta,Y_1,Y_0}[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon)] \equiv 0 \tag{I.3}$$

defines $\bar{\beta}$. We then apply (3.13) to the function $f = h_1$ and determine (b_1, b_2) by setting the first two terms of the Taylor expansion of h_1 to zero. We find in particular that $b_1 = 0$.

Then by applying Theorem 1 to Taylor-expand the quantities D_β and $S_{\beta,0}$ in (4.25), we obtain that

$$\begin{aligned} D_\theta &= -\varepsilon\frac{E[\Delta_0]}{\sigma_0^2}E_{Y_0}\left[\left(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\right)^2\right] + O(\varepsilon^3), \\ S_{\theta,0} &= \varepsilon\frac{E[\Delta_0]}{\sigma_0^2}E_{Y_0}\left[\left(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\right)^2\right] + \varepsilon^2\frac{\mathrm{var}[\Delta_0]}{4\sigma_0^2} \\ &\quad \times\left(2E_{Y_0}\left[\left(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\right)^2\frac{\partial\mu(Y_0,\theta_0)}{\partial y}\right] - \sigma_0^2E_{Y_0}\left[\left(\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y\partial\theta}\right)^2\right]\right) + O(\varepsilon^3), \end{aligned}$$

which we write as $D_\theta = \varepsilon D_\theta^{(1)} + O(\varepsilon^3)$ and $S_{\theta,0} = \varepsilon S_{\theta,0}^{(1)} + \varepsilon^2 S_{\theta,0}^{(2)} + O(\varepsilon^3)$.

This leaves the evaluation of the time series term $T_\theta = S_\theta - S_{\theta,0}$. For this purpose, let us now show that

$$E_{\Delta,Y_1}[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon)|Y_0] = \varepsilon^{\alpha_1} q_1^*(Y_0, \beta_0, 0) + o_p(\varepsilon^2), \tag{I.4}$$

where $\alpha_1 = 2$, and $q_1^*(Y_0, \beta_0, 0)$ is finite. This follows from Theorem 1 applied to Taylor-expand in ε the conditional expectation of the particular function $f = h_1$:

$$E_{\Delta,Y_1}[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon)|Y_0] = \varepsilon^2 q_1^*(Y_0, \beta_0, 0) + o_p(\varepsilon^2).$$

Specifically, we obtain that

$$\begin{aligned} q_1^*(y, \beta_0, 0) &= -\sigma_0^4 \mathrm{var}[\Delta_0]\left(\frac{1}{2\sigma_0^2}\frac{\partial^2 \lambda(y, \beta_0)}{\partial y \partial \theta}\mu(y, \theta_0) + \frac{1}{4}\frac{\partial^3 \lambda(y, \beta_0)}{\partial y^2 \partial \theta}\right) \\ &= -\frac{1}{2}\sigma_0^2 \mathrm{var}[\Delta_0](A_{\beta_0} \cdot \dot{\lambda}_\theta)(y, \beta_0), \end{aligned} \tag{I.5}$$

which reduces to (6.3) when expressed as a function of the drift function μ.

Let $\Im_k$ denote the σ-field generated by $\{Y_0, \ldots, Y_k\}$, i.e., by the first k observations. We have

$$\begin{aligned} T_\theta = S_\theta - S_{\theta,0} &= 2\sum_{k=1}^{\infty} S_{\beta,k} = 2\sum_{j=1}^{\infty} E[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon)h_1(Y_{k+1}, Y_k, \bar{\beta}, \varepsilon)] \\ &= 2\sum_{k=1}^{\infty} E[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon)E[h_1(Y_{k+1}, Y_k, \bar{\beta}, \varepsilon)|\Im_k]] \\ &= 2\sum_{k=1}^{\infty} E[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon)\varepsilon^2 q_1^*(Y_k, \beta_0, \varepsilon)] \\ &= 2\varepsilon^2 \sum_{k=1}^{\infty} E[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon)E[q_1^*(Y_k, \beta_0, \varepsilon)|Y_1]] \\ &= 2\varepsilon \frac{1}{E[\Delta_0]} E_{\Delta,Y_1,Y_0}[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon)r_1(Y_1, \beta_0, \varepsilon)] \end{aligned} \tag{I.6}$$

where

$$r_1(y, \beta_0, \varepsilon) = \varepsilon E[\Delta_0]\sum_{k=1}^{\infty} E_{Y_k}[q_1^*(Y_k, \beta_0, \varepsilon)|Y_1 = y] \tag{I.7}$$

$$\to r_1(y, \beta_0, 0) = \int_0^\infty E_{X_t}[q_1^*(X_t, \beta_0, 0)|X_0 = y]\,dt \quad \text{as} \quad \varepsilon \to 0. \tag{I.8}$$

By a standard martingale argument, at $\varepsilon = 0$, $r_1(y, \beta_0, 0)$ is the solution of the ordinary differential equation

$$\mu(y, \theta)\frac{\partial r_1(y, \beta_0, 0)}{\partial y} + \frac{1}{2}\sigma^2 \frac{\partial^2 r_1(y, \beta_0, 0)}{\partial y^2} = -q_1^*(y, \beta_0, 0) \tag{I.9}$$

that satisfies $E_{Y_0}[r_1(Y_0, \beta_0, 0)] = 0$. Still using Theorem 1, the rest of the analysis is based on expanding $r(y, \beta, \varepsilon)$ in powers of ε. We get the order of T_β through the Taylor expansion

$$\begin{aligned} r_1(Y_1, \beta_0, \varepsilon) = r_1(Y_0, \beta_0, 0) &+ (Y_1 - Y_0)\frac{\partial r_1(Y_0, \beta_0, 0)}{\partial y} + \frac{1}{2}(Y_1 - Y_0)^2 \frac{\partial^2 r_1(Y_0, \beta_0, 0)}{\partial y^2} \\ &+ \varepsilon \frac{\partial r_1(Y_0, \beta_0, 0)}{\partial \varepsilon} + o_p(\varepsilon), \end{aligned} \tag{I.10}$$

in (I.6) to obtain

$$
\begin{aligned}
T_\theta = {} & 2\varepsilon \frac{1}{E[\Delta_0]} E_{\Delta, Y_1, Y_0}[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon) r_1(Y_0, \beta_0, 0)] \\
& + 2\varepsilon \frac{1}{E[\Delta_0]} E_{\Delta, Y_1, Y_0}\left[(Y_1 - Y_0) h_1(Y_1, Y_0, \bar{\beta}, \varepsilon) \frac{\partial r_1(Y_0, \beta_0, 0)}{\partial y} \right] \\
& + \varepsilon \frac{1}{E[\Delta_0]} E_{\Delta, Y_1, Y_0}\left[(Y_1 - Y_0)^2 h(Y_1, Y_0, \bar{\beta}, \varepsilon) \frac{\partial^2 r_1(Y_0, \beta_0, 0)}{\partial y^2} \right] \\
& + 2\varepsilon \frac{1}{E[\Delta_0]} E_{\Delta, Y_1, Y_0}\left[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon) \left(\varepsilon \frac{\partial r_1(Y_0, \beta_0, 0)}{\partial \varepsilon} + o_p(\varepsilon) \right) \right].
\end{aligned} \tag{I.11}
$$

From this it follows that $T_\theta = T_\theta^{(2)} \varepsilon^2 + O(\varepsilon^3)$ with

$$
T_\theta^{(2)} = 4 E_{Y_0}[q_1^*(Y_0, \beta_0, 0) G_1(Y_0, \beta_0)] \tag{I.12}
$$

where

$$
g_1(y_0, \beta_0) = \partial h_1(y_1, y_0, \bar{\beta}, \varepsilon)/\partial y_1|_{y_1 = y_0, \bar{\beta} = \beta_0, \varepsilon = 0} \equiv (\partial h_1/\partial y_1)(y_0, y_0, \beta_0, 0) \tag{I.13}
$$

and G_1 is its primitive (because q_1^* has mean 0, the constant of integration in G_1 is irrelevant for (I.12)). Given the expression of h_1, g_1 simplifies to $g_1(y_0, \beta_0) = \dot{\mu}(y_0, \theta_0)/\sigma_0^2$ and thus

$$
G_1(Y_0, \beta_0) = \sigma_0^{-2} \int^{Y_0} \dot{\mu}(z, \theta_0)\, dz.
$$

Simplifying further,

$$
T_\theta^{(2)} = 4 E_{Y_0}[q_1^*(Y_0, \beta_0, 0) G_1(Y_0, \beta_0)] = -2 E_{Y_0}[\dot{q}_1^*(Y_0, \beta_0, 0)].
$$

Since

$$
\dot{q}_1^*(Y_0, \beta_0, 0) = -\frac{1}{2} \sigma_0^2 \mathrm{var}[\Delta_0] \left\{ (A_{\beta_0} \cdot \ddot{\lambda}_\theta)(y, \beta_0) + \dot{\mu}_\theta(y, \theta_0) \frac{\partial \dot{\lambda}_\theta(y, \beta_0)}{\partial y} \right\}, \tag{I.14}
$$

we obtain

$$
T_\theta^{(2)} = \sigma_0^{(2)} \mathrm{var}[\Delta_0] E_{Y_0}\left[\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0) \frac{\partial^2 \lambda}{\partial y \partial \theta}(Y_0, \beta_0) \right].
$$

From the expressions for D_θ, $S_{\theta, 0}$, and T_θ, (6.2) follows.

J. *Proof of Theorem 5*

Consider now h_2, the second component of the vector h in (4.23). For reasons of mathematical convenience, let us replace h_2 by εh_2 in what follows. Premultiplying a moment condition by the nonrandom quantity ε does not affect any of the results regarding the properties of the resulting estimator. The technique we use is similar to that applied to the analysis of h_1. We now have

$$
\begin{aligned}
& E_{\Delta Y_1}[h_2(Y_1, Y_0, \beta, \varepsilon) | Y_0] \\
& \quad = E_{Y_1}[\varepsilon \dot{l}_{\sigma^2}(Y_1 | Y_0, E[\Delta], \theta, \sigma^2) | Y_0] \\
& \quad = -\varepsilon \frac{1}{2\sigma^4}(\sigma^2 - \sigma_0^2) - \varepsilon^2 \left\{ \frac{\sigma_0^4}{\sigma^4} \frac{E[\Delta_0^2]}{E[\Delta_0]} \lambda(Y_0, \beta_0) + E[\Delta_0] \frac{1}{\sigma^4} \left(\mu(Y_0, \theta) \mu(Y_0, \theta_0) \right.\right. \\
& \qquad \left.\left. - \frac{1}{2} \mu(Y_0, \theta)^2 + \frac{1}{2} \sigma_0^2 \frac{\partial \mu(Y_0, \theta_0)}{\partial y} \right) \right\} + O_p(\varepsilon^3).
\end{aligned}
$$

To compute the bias of the diffusion estimator, we have in this case $\bar{\beta} = (\theta_0, \bar{\sigma}^2)'$ and write $\bar{\sigma}^2 = \sigma_0^2 + \varepsilon b_1 + O(\varepsilon^2)$. Now it is the equation

$$E_{\Delta, Y_1, Y_0}[h_2(Y_1, Y_0, \bar{\beta}, \varepsilon)] \equiv 0 \tag{J.1}$$

that defines $\bar{\beta}$. We then apply (3.13) to the function $f = h_2$ and determine b_1 by setting the first term of the Taylor expansion of h_2 to zero, yielding $b_1 = (i^{(0)}_{\sigma^2\sigma^2})^{-1}\breve{b}$ where

$$\breve{b} = -\frac{\text{var}[\Delta_0]}{E[\Delta_0]} E_{Y_0}[\lambda(Y_0, \beta_0)] = \frac{\text{var}[\Delta_0]}{4\sigma_0^2 E[\Delta_0]} E_{Y_0}[\mu(Y_0, \theta_0)^2]. \tag{J.2}$$

Replacing $i^{(0)}_{\sigma^2\sigma^2}$ by its value, b_1 reduces to (6.4). Note that since $\breve{b} > 0$ (with $\breve{b} = 0$ only if Δ_0 is nonrandom), the bias in the estimator of σ^2 is always of positive sign.

Then by applying Theorem 1, we obtain that

$$\begin{aligned}
D_{\sigma^2} &= -\varepsilon \frac{1}{2\sigma_0^4} + \varepsilon^2 \frac{\text{var}[\Delta_0]}{2\sigma_0^4 E[\Delta_0]} E_{Y_0}\left[\frac{\partial \mu}{\partial y}(Y_0, \theta_0)\right] + O(\varepsilon^3), \\
S_{\sigma^2, 0} &= \varepsilon^2 \left(\frac{3\text{var}[\Delta_0] + 2E[\Delta_0]^2}{4\sigma_0^4 E[\Delta_0]^2}\right) + \varepsilon^3 \frac{1}{4\sigma_0^4 E[\Delta_0]^3} E_{Y_0}\left[\frac{\partial \mu}{\partial y}(Y_0, \theta_0)\right] \\
&\quad \times (-6\text{var}[\Delta_0]^2 + (3(E[\Delta_0^3] - E[\Delta_0]^3) - 5\text{var}[\Delta_0]E[\Delta_0])E[\Delta_0]) + O(\varepsilon^4),
\end{aligned}$$

which we write as $D_{\sigma^2} = \varepsilon D^{(1)}_{\sigma^2} + \varepsilon^2 D^{(2)}_{\sigma^2} + O(\varepsilon^3)$ and $S_{\sigma^2, 0} = \varepsilon^2 S^{(2)}_{\sigma^2, 0} + \varepsilon^3 S^{(3)}_{\sigma^2, 0} + O(\varepsilon^4)$.

To evaluate $T_{\sigma^2} = S_{\sigma^2} + S_{\sigma^2, 0}$, we find that in this case (I.4) becomes

$$E_{\Delta, Y_1}[h_2(Y_1, Y_0, \bar{\beta}, \varepsilon) | Y_0] = \varepsilon^2 q_2(Y_0, \beta_0, 0) + O_p(\varepsilon^3)$$

with

$$q_2(y, \beta_0, 0) = -\frac{\text{var}[\Delta_0]}{E[\Delta_0]}(\lambda(y, \beta_0) - E_{Y_0}[\lambda(Y_0, \beta_0)]),$$

replacing q_1^*. This expression for q_2 reduces to (6.6) when expressed as a function of μ. Note that in this case (I.13) reduces to $G_2(Y_0, \beta_0) = \int^{y_0} g_2(z_0 \beta_0)\, dz_0 = 0$.

Define the function r_2 analogously to r_1 in (I.7) except that q_2 is used in place of q_1^*. By the same reasoning, we have that

$$\begin{aligned}
T_{\sigma^2} &= 2\varepsilon \frac{1}{E[\Delta_0]} E_{\Delta, Y_1, Y_0}[h_2(Y_1, Y_0, \bar{\beta}, \varepsilon) r_2(Y_0, \beta_0, 0)] \\
&\quad + 2\varepsilon \frac{1}{E[\Delta_0]} E_{\Delta, Y_1, Y_0}\left[(Y_1 - Y_0) h_1(Y_1, Y_0, \bar{\beta}, \varepsilon) \frac{\partial r_1(Y_0, \beta_0, 0)}{\partial y}\right] \\
&\quad + \varepsilon \frac{1}{E[\Delta_0]} E_{\Delta, Y_1, Y_0}\left[(Y_1 - Y_0)^2 h(Y_1, Y_0, \bar{\beta}, \varepsilon) \frac{\partial^2 r_1(Y_0, \beta_0, 0)}{\partial y^2}\right] \\
&\quad + 2\varepsilon \frac{1}{E[\Delta_0]} E_{\Delta, Y_1, Y_0}\left[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon)\left(\varepsilon \frac{\partial r_1(Y_0, \beta_0, 0)}{\partial \varepsilon} + o_p(\varepsilon)\right)\right],
\end{aligned} \tag{J.3}$$

and we need to Taylor-expand the function h_2. For that purpose, we need to compute the partial derivatives of h_2, evaluated at $(y_0, y_0, \beta_0, 0)$:

$$\begin{aligned}
&\frac{\partial h_2(Y_0, Y_0, \beta_0, 0)}{\partial y_1} = g_2(Y_0, \beta_0) = 0, \qquad \frac{\partial^2 h_2(Y_0, Y_0, \beta_0, 0)}{\partial y_1^2} = \frac{1}{\sigma_0^4 E[\Delta_0]}, \\
&\frac{\partial^3 h_2(Y_0, Y_0, \beta_0, 0)}{\partial y_1^3} = 0,
\end{aligned}$$

$$\frac{\partial h_2(Y_0, Y_0, \beta_0, 0)}{\partial \varepsilon} = -\frac{1}{2\sigma_0^2}, \qquad \frac{\partial^2 h_2(Y_0, Y_0, \beta_0, 0)}{\partial y_1 \partial \varepsilon} = -\frac{\mu(Y_0, \beta_0)}{\sigma_0^4},$$
$$\frac{\partial h_2(Y_0, Y_0, \beta_0, 0)}{\partial \sigma^2} = 0, \qquad \frac{\partial^2 h_2(Y_0, Y_0, \beta_0, 0)}{\partial y_1 \partial \sigma^2} = 0.$$

We identify separately the order ε^3 contribution to T_{σ^2} from the four terms in (J.3). They are in the form of a coefficient of ε^2, to be premultiplied by ε. The coefficient of ε^2 contributed by the first term is

$$A_1 \equiv E_{Y_0}[q_2(Y_0, \beta_0, 0) r_2(Y_0, \beta_0, 0)].$$

Similarly, from the second term comes

$$\begin{aligned} A_2 \equiv{}& \frac{3}{2}\sigma_0^2 E[\Delta_0^2] E_{Y_0}\left[\mu(Y_0, \theta_0) \frac{\partial^2 h_2(Y_0, Y_0, 0, \beta_0, 0)}{\partial y_1^2} \frac{\partial r_2(Y_0, \beta_0, 0)}{\partial y}\right] \\ &+ E[\Delta_0] E_{Y_0}\left[\mu(Y_0, \theta_0) \frac{\partial h_2(Y_0, Y_0, 0, \beta_0, 0)}{\partial \varepsilon} \frac{\partial r_2(Y_0, \beta_0, 0)}{\partial y}\right] \\ &+ \sigma_0^2 E[\Delta_0] E_{Y_0}\left[\frac{\partial^2 h_2(Y_0, Y_0, 0, \beta_0, 0)}{\partial y_1 \partial \varepsilon} \frac{\partial r_2(Y_0, \beta_0, 0)}{\partial y}\right] \\ ={}& \frac{3\mathrm{var}[\Delta_0^2]}{2\sigma_0^2 E[\Delta_0]} E_{Y_0}\left[\mu(Y_0, \theta_0) \frac{\partial r_2(Y_0, \beta_0, 0)}{\partial y}\right], \end{aligned}$$

then from the third term

$$\begin{aligned} A_3 \equiv{}& \frac{3}{2}\sigma_0^4 E[\Delta_0^2] E_{Y_0}\left[\frac{\partial^2 h_2(Y_0, Y_0, 0, \beta_0, 0)}{\partial y_1^2} \frac{\partial^2 r_2(Y_0, \beta_0, 0)}{\partial y^2}\right] \\ &+ \sigma_0^2 E[\Delta_0] E_{Y_0}\left[\frac{\partial h_2(Y_0, Y_0, 0, \beta_0, 0)}{\partial \varepsilon} \frac{\partial^2 r_2(Y_0, \beta_0, 0)}{\partial y^2}\right] \\ ={}& \frac{1}{2E[\Delta_0]}\left(3E[\Delta_0^2] - E[\Delta_0]^2\right) E_{Y_0}\left[\frac{\partial^2 r_2(Y_0, \beta_0, 0)}{\partial y^2}\right], \end{aligned}$$

and from the fourth and last term

$$A_4 \equiv \frac{1}{2}\sigma_0^2 E[\Delta_0] E_{Y_0}\left[\frac{\partial^2 h_2(Y_0, Y_0, 0, \beta_0, 0)}{\partial y_1^2} \frac{\partial r_2(Y_0, \beta_0, 0)}{\partial \varepsilon}\right] = \frac{1}{2\sigma_0^2} E_{Y_0}\left[\frac{\partial r_2(Y_0, \beta_0, 0)}{\partial \varepsilon}\right] = 0,$$

since $E_{Y_0}[r_2(Y_0, \beta_0, \varepsilon)] = 0$ from (I.7) and $E_{Y_0}[q_2(Y_0, \beta_0, \varepsilon)] = 0$.

Hence

$$E_{Y_0}\left[\frac{\partial r_2(Y_0, \beta_0, 0)}{\partial \varepsilon}\right] = \frac{\partial}{\partial \varepsilon} E_{Y_0}[r_2(Y_0, \beta_0, \varepsilon)] = 0,$$

and in particular for $\varepsilon \to 0$. So finally we obtain

$$T_{\sigma^2} = \frac{2}{E[\Delta_0]}\varepsilon^3 \left\{A_1 + A_2 + \frac{1}{2}A_3\right\} + O(\varepsilon^4). \tag{J.4}$$

Collecting the results for D_{σ^2}, $S_{\sigma^2,0}$, and T_{σ^2} gives (6.5).

K. *Proof of Theorem 6*

We are now using both moment functions in the vector $h = (h_1, h_2)'$ of equation (4.23), with h_2 again replaced by εh_2, to estimate jointly $\beta = (\theta, \sigma^2)'$. Regarding first the bias, we have that

$$\bar{\theta} - \theta_0 = -\frac{i^{(1)}_{\theta\sigma^2}}{i^{(1)}_{\theta\theta}}(\bar{\sigma}^2 - \sigma_0^2) + O(\varepsilon^2) \tag{K.1}$$

which becomes

$$-\varepsilon \frac{i^{(1)}_{\theta\sigma^2}}{i^{(1)}_{\theta\theta} i^{(0)}_{\sigma^2\sigma^2}} \breve{b} + O(\varepsilon^2) = \varepsilon b_1 + O(\varepsilon^2)$$

if σ^2 is unknown (and $O(\varepsilon^2)$ otherwise, as in the first situation considered in this Theorem where $b_1 = 0$ when $\sigma^2 = \sigma_0^2$ is known). When σ^2 is estimated, b_1 is given by the expression in (6.8). In any event, the bias of the estimator of σ^2 is identical whether θ is known or has to be estimated.

Define q_1 by $E_{\Delta, Y_1}[h_1(Y_1, Y_0, \tilde{\beta}, \varepsilon)|Y_0] = \varepsilon^2 q_1^*(Y_0, \beta_0, 0) + o_p(\varepsilon^2)$. Compared to q_1^*, the extra term in q_1 due to the estimation of σ^2 is given by

$$q_1(y, \beta_0, 0) - q_1^*(y, \beta_0, 0) = -2\sigma_0^2 \mathrm{var}[\Delta_0] E_{Y_0}[\kappa(Y_0, \theta_0, \sigma_0^2)] \times \left(\frac{i^{(1)}_{\theta\sigma^2}}{i^{(1)}_{\theta\theta}} \left(\frac{\partial \mu}{\partial \theta}(y, \theta_0) \right)^2 - \frac{1}{2} \frac{\partial^2 \mu}{\partial y \partial \theta}(y, \theta_0) \right)$$

(where q_1^* is defined in (I.5)). The result is given by (6.10) as a function of μ.

By applying Theorem 1 further, the structure of the matrices D_β, S_β, and T_β is as follow:

$$D_\beta = \begin{pmatrix} d_{\theta\theta} & d_{\theta\sigma^2} \\ d_{\sigma^2\theta} & d_{\sigma^2\sigma^2} \end{pmatrix}, \qquad S_{\beta,0} = \begin{pmatrix} s_{\theta\theta,0} & s_{\theta\sigma^2,0} \\ s_{\sigma^2\theta,0} & s_{\sigma^2\sigma^2,0} \end{pmatrix},$$

$$T_\beta = S_\beta - S_{\beta,0} = \begin{pmatrix} t_{\theta\theta} & t_{\theta\sigma^2} \\ t_{\sigma^2\theta} & t_{\sigma^2\sigma^2} \end{pmatrix},$$

$$d_{\theta\theta} = \varepsilon D_\theta^{(1)} + \varepsilon^2 \left(\frac{\mathrm{var}[\Delta_0] E_{Y_0}\left[\frac{\partial \mu}{\partial y}(Y_0, \theta_0)\right] \left(2E_{Y_0}\left[\left(\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\right)^2\right]\right)^2 + 3\sigma_0^2 E_{Y_0}\left[\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0) \frac{\partial^2 \mu}{\partial \theta^2}(Y_0, \theta_0)\right] E_{Y_0}\left[\frac{\partial^2 \mu}{\partial y \partial \theta}(Y_0, \theta_0)\right]}{4\sigma_0^2 E_{Y_0}\left[\left(\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\right)^2\right]} - \frac{\mathrm{var}[\Delta_0] E_{Y_0}\left[\frac{\partial \mu}{\partial y}(Y_0, \theta_0)\right] E_{Y_0}\left[\frac{\partial^3 \mu}{\partial y \partial \theta^2}(Y_0, \theta_0)\right]}{4} \right) + O(\varepsilon^3),$$

$$d_{\sigma^2\sigma^2} = \varepsilon D_{\sigma^2}^{(1)} + \varepsilon^2 D_{\sigma^2}^{(2)} + O(\varepsilon^3),$$

$$d_{\theta\sigma^2} = -\varepsilon \frac{E[\Delta_0] E_{Y_0}\left[\frac{\partial^2 \mu}{\partial y \partial \theta}(Y_0, \theta_0)\right]}{2\sigma_0^2} + \varepsilon^2 \left(\frac{\mathrm{var}[\Delta_0] E_{Y_0}\left[\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\right] E_{Y_0}\left[\frac{\partial^2 \mu}{\partial y \partial \theta}(Y_0, \theta_0)\right] \left(2E_{Y_0}\left[\left(\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\right)^2\right] + \sigma_0^2 E_{Y_0}\left[\frac{\partial^2 \mu}{\partial y \partial \theta}(Y_0, \theta_0)\right]\right)}{8\sigma_0^2 E_{Y_0}\left[\left(\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\right)^2\right]} - \frac{E[\Delta_0]^2 \left(E_{Y_0}\left[\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0) \frac{\partial^2 \mu}{\partial y^2}(Y_0, \theta_0)\right] + 2E_{Y_0}\left[\frac{\partial^2 \mu}{\partial y \partial \theta}(Y_0, \theta_0) \frac{\partial \mu}{\partial y}(Y_0, \theta_0)\right]\right)}{12\sigma_0^2} \right) + O(\varepsilon^3) \equiv \varepsilon d_{\theta\sigma^2}^{(1)} + \varepsilon^2 d_{\theta\sigma^2}^{(2)} + O(\varepsilon^3),$$

$$d_{\sigma^2\theta} = -\varepsilon^2 \frac{E[\Delta_0]E_{Y_0}\left[\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\right]}{2\sigma_0^2} + O(\varepsilon^3) \equiv \varepsilon^2 d^{(2)}_{\sigma^2\theta} + O(\varepsilon^3),$$

$$s_{\theta\theta,0} = \varepsilon S^{(1)}_{\sigma^2,0} + \varepsilon^2 \left(S^{(2)}_{\sigma^2,0} - \frac{\begin{array}{c}\operatorname{var}[\Delta_0]E_{Y_0}\left[\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\right]\left(2E_{Y_0}\left[\left(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\right)^2\right]\right)^2 \\ +\sigma_0^2 E_{Y_0}\left[\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\frac{\partial^2\mu}{\partial\theta^2}(Y_0,\theta_0)\right]E_{Y_0}\left[\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\right]\end{array}}{2\sigma_0^2 E_{Y_0}\left[\left(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\right)^2\right]} \right) + O(\varepsilon^3),$$

$$s_{\theta\sigma^2,0} = s_{\sigma^2\theta,0} = \varepsilon^2 \frac{E[\Delta_0]E_{Y_0}\left[\frac{\partial^2\mu}{\partial y\partial\theta}(Y_0,\theta_0)\right]}{2\sigma_0^2} + O(\varepsilon^3) \equiv \varepsilon^2 s^{(2)}_{\sigma^2\theta,0} + O(\varepsilon^3),$$

$$s_{\sigma^2\sigma^2,0} = \varepsilon^2 S^{(2)}_{\sigma^2,0} + \varepsilon^3 S^{(3)}_{\sigma^2,0} O(\varepsilon^4).$$

As for the time series terms,

$$t_{\theta\theta} = \varepsilon^2\left(T^{(2)}_\theta - 2E_{Y_0}[(\dot q_1(Y_0,\beta_0,0) - \dot q_1^*(Y_0,\beta_0,0))]\right) + O(\varepsilon^3) \equiv \varepsilon^2 t^{(2)}_{\theta\theta} + O(\varepsilon^3),$$

$$t_{\sigma^2\sigma^2} = \varepsilon^3 T^{(3)}_{\sigma^2} + O(\varepsilon^3),$$

$$t_{\theta\sigma^2} = t_{\sigma^2\theta} = \varepsilon^2\left(2E_{Y_0}[G_1(Y_0,\beta_0)q_2(Y_0,\beta_0,0)]\right) + O(\varepsilon^3) \equiv \varepsilon^2 t^{(2)}_{\theta\sigma^2} + O(\varepsilon^3).$$

The leading terms of $t_{\theta\theta}$ and $t_{\sigma^2\sigma^2}$ are identical to their one-parameter counterparts, except that q_1 is different from the one-parameter case. That is,

$$\begin{aligned} t^{(2)}_{\theta\theta} &= 4E_{Y_0}[G_1(Y_0,\beta_0)q_1(Y_0,\beta_0,0)] \\ &= -2E_{Y_0}[\dot q_1(Y_0,\beta_0,0)] \\ &= -2E_{Y_0}[\dot q_1^*(Y_0,\beta_0,0)] - 2E_{Y_0}[(\dot q_1(Y_0,\beta_0,0) - \dot q_1^*(Y_0,\beta_0,0))] \\ &= T^{(2)}_\theta - 2E_{Y_0}[(\dot q_1(Y_0,\beta_0,0) - \dot q_1^*(Y_0,\beta_0,0))] \end{aligned}$$

whereas $t^{(3)}_{\sigma^2\sigma^2} = T^{(3)}_{\sigma^2}$. The expression for $t^{(1)}_{\theta\sigma^2}$ follows by gathering the terms:

$$t^{(2)}_{\theta\sigma^2} = 2E_{Y_0}[G_1(Y_0,\beta_0)q_2(Y_0,\beta_0,0)].$$

L. *Proof of Proposition 6*

(i) We proceed as in the proof of Theorem 4, using the h_1 moment alone. We now have for $\theta - \theta_0 = b_1(\omega_0)\varepsilon + O(\varepsilon^2)$:

$$\begin{aligned} E_{\Delta,Y_1,Y_0}[h_1(Y_1,Y_0,\beta,\varepsilon)] &= E_{\Delta,Y_1,Y_0}[|\dot l_\theta(Y_1|Y_0,\varepsilon\omega_0,\theta,\sigma^2)|] \\ &= \varepsilon^2\left(\frac{\omega_0 b_1(\omega_0)}{\sigma_0^2}E_{Y_0}\left[\left(\frac{\partial\mu}{\partial\theta}(Y_0,\theta_0)\right)^2\right] + \frac{\omega_0(E[\Delta_0]-\omega_0)}{12}\right. \\ &\quad \times\left(2E_{Y_0}\left[\frac{\partial\mu}{\partial y}(Y_0,\theta_0)\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y\partial\theta}\right]\right. \\ &\quad \left.\left. + E_{Y_0}\left[\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\frac{\partial^2\mu(Y_0,\theta_0)}{\partial y^2}\right]\right)\right) + O(\varepsilon^3). \end{aligned} \tag{L.1}$$

The bias term $b_1(\omega_0)$ is such that the leading term in (L.l) is zero. This gives the result (6.12).

Next, by applying Theorem 1 to Taylor-expand the quantities D_β and $S_{\beta,0}$ in (4.25), we obtain that

$$D_\theta = -\varepsilon \frac{\omega_0}{\sigma_0^2} E_{Y_0}\left[\left(\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\right)^2\right] + O(\varepsilon^2),$$

$$S_{\theta,0} = \varepsilon \frac{E[\Delta_0]}{\sigma_0^2} E_{Y_0}\left[\left(\frac{\partial \mu}{\partial \theta}(Y_0, \theta_0)\right)^2\right] + O(\varepsilon^2),$$

which we write as $D_\theta = \varepsilon D_\theta^{(1)} + O(\varepsilon^2)$ and $S_{\theta,0} = \varepsilon S_{\theta,0}^{(1)} + O(\varepsilon^2)$. Note that $D_\theta^{(1)}$ now contains ω_0 whereas $S_{\theta,0}^{(1)}$ still contains $E[\Delta_0]$.

This leaves us with the evaluation of the time series term $T_\theta = S_\theta - S_{\theta,0}$. For this purpose, we show that

$$E_{\Delta, Y_1}[h_1(Y_1, Y_0, \bar{\beta}, \varepsilon)|Y_0] = \varepsilon^{\alpha_1} q_1^*(Y_0, \beta_0, 0) + o_p(\varepsilon^2), \tag{L.2}$$

where $\alpha_1 = 1$ (instead of $\alpha_1 = 2$ when $\omega_0 = E[\Delta_0]$) and $q_1^*(Y_0, \beta_0, 0)$ is finite:

$$q_1^*(y, \beta_0, 0) = (E[\Delta_0] - \omega_0)\frac{1}{\sigma_0^2}\left(\mu(y, \theta_0)\frac{\partial \mu(y, \theta_0)}{\partial \theta} + \frac{\sigma_0^2}{2}\frac{\partial^2 \mu(y, \theta_0)}{\partial y \partial y \partial \theta}\right). \tag{L.3}$$

From this it follows that $T_\theta = T_\theta^{(1)}\varepsilon^1 + O(\varepsilon^2)$ with

$$T_\theta^{(1)} = 4E_{Y_0}[q_1^*(Y_0, \beta_0, 0)G_1(Y_0, \beta_0)] + \frac{2}{E[\Delta_0]}E_{Y_0}[q_1^*(Y_0, \beta_0, 0)r_1(Y_0, \beta_0, 0)] \tag{L.4}$$

where

$$g_1(y_0, \beta_0) = \partial h_1(y_1, y_0, \bar{\beta}, \varepsilon)/\partial y_1|_{y_1 = y_0, \bar{\beta} = \beta_0, \varepsilon = 0} \equiv (\partial h_1/\partial y_1)(y_0, y_0, \beta_0, 0), \tag{L.5}$$

G_1 is its primitive, and $r_1(y, \beta_0, 0)$ is the solution of the ordinary differential equation

$$\mu(y, \theta)\frac{\partial r_1(y, \beta_0, 0)}{\partial y} + \frac{1}{2}\sigma^2 \frac{\partial^2 r_1(y, \beta_0, 0)}{\partial y^2} = -q_1^*(y, \beta_0, 0) \tag{L.6}$$

that satisfies $E_{Y_0}[r_1(Y_0, \beta_0, 0)] = 0$. From the expressions for D_θ, $S_{\theta,0}$, and T_θ, the asymptotic variance of the estimator is

$$\check{V}_\theta = \check{V}_\theta^{(0)} + \check{V}_\theta^{(1)}\varepsilon + O(\varepsilon^2)$$

with

$$\check{V}_\theta^{(0)} = \left(S_{\theta,0}^{(1)} + T_\theta^{(1)}\right)\Big/\left(D_\theta^{(1)}\right)^2.$$

(ii) We now proceed as in the proof of Theorem 5, using the h_2 moment alone. We have

$$\begin{aligned} E_{\Delta, Y_1}[h_2(Y_1, Y_0, \bar{\beta}, \varepsilon)|Y_0] &= E_{Y_1}[\varepsilon \dot{l}_{\sigma^2}(Y_1|Y_0, \varepsilon\omega_0, \theta_0, \bar{\sigma}^2(\omega_0))|Y_0] \\ &= \frac{1}{2s^{(0)}}\left(\frac{\sigma_0^2 E[\Delta_0]}{\omega_0 s^{(0)}} - 1\right)\varepsilon + O_p(\varepsilon^2), \end{aligned}$$

where $\bar{\beta} = (\bar{\sigma}^2(\omega_0), \theta_0)$ and

$$\bar{\sigma}^2(\omega_0) = s^{(0)} + s^{(1)}\varepsilon + O(\varepsilon^2). \tag{L.7}$$

Hence

$$E_{\Delta, Y_1, Y_0}[h_2(Y_1, Y_0, \beta, \varepsilon)|Y_0] = \frac{1}{2s^{(0)}}\left(\frac{\sigma_0^2 E[\Delta_0]}{\omega_0 s^{(0)}} - 1\right)\varepsilon + O_p(\varepsilon^2). \tag{L.8}$$

By definition of $\bar{\beta}$, $s^{(0)}$ is the value that makes the leading term of (L.8) zero, that is:

$$s^{(0)} = \frac{E[\Delta_0]}{\omega_0}\sigma_0^2 \tag{L.9}$$

and so we have

$$\bar{\sigma}^2(\omega_0) = \frac{E[\Delta_0]}{\omega_0}\sigma_0^2 + O(\varepsilon). \tag{L.10}$$

REFERENCES

AÏT-SAHALIA, Y. (1996): "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527–560.

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

BARNDORFF-NIELSEN, O. E., AND D. R. COX (1989): *Asymptotic Techniques for Use in Statistics*. London, U.K.: Chapman and Hall.

BARTLETT, M. S. (1953): "Approximate Confidence Intervals," *Biometrika*, 40, 12–19.

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

COX, D. R. (1975): "Partial Likelihood," *Biometrika*, 62, 269–276.

DACUNHA-CASTELLE, D., AND D. FLORENS-ZMIROU (1986): "Estimation of the Coefficients of a Diffusion from Discrete Observations," *Stochastics*, 19, 263–284.

DUFFIE, D., AND P. GLYNN (1997): "Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals," Tech. Rep., Stanford University.

ENGLE, R. F., AND J. R. RUSSELL (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

FLORENS, J.-P., AND D. FOUGERE (1996): "Noncausality in Continuous Time," *Econometrica*, 64, 1195–1212.

GOURIÉROUX, C., AND A. MONFORT (1995): *Statistics and Econometric Models*, Vol. 1. Cambridge: Cambridge U. Press.

GOURIÉROUX, C., A. MONFORT, E. RENAULT, AND A. TROGNON (1987): "Generalized Residuals," *Journal of Econometrics*, 34, 5–32.

GRANGER, C. W. J. (1969): "Investigating Causal Relations by Econometric Models and Cross Spectral Methods," *Econometrica*, 37, 424–430.

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

HANSEN, L. P., AND J. A. SCHEINKMAN (1995): "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767–804.

HANSEN, L. P., J. A. SCHEINKMAN, AND N. TOUZI (1998): "Identification of Scalar Diffusions Using Eigenvectors," *Journal of Econometrics*, 86, 1–32.

HASBROUCK, J. (1991): "Measuring the Information Content of Stock Trades," *Journal of Finance*, 46, 179–207.

——— (1996): "Modeling Market Microstructure Time Series," in *Handbook of Statistics, Volume 14*, ed. by C. R. Rao and G. S. Maddala. Amsterdam: North-Holland, 647–692.

KARATZAS, I., AND S. E. SHREVE (1991): *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.

KARLIN, S., AND H. M. TAYLOR (1981): *A Second Course in Stochastic Processes*. New York: Academic Press.

MASRY, E. (1978): "Alias-free Sampling: An Alternative Conceptualization and its Applications," *IEEE Transactions on Information Theory*, 24, 317–324.

——— (1983): "Spectral and Probability Density Estimation from Irregularly Observed Data," in *Time Series Analysis of Irregularly Observed Data, Lecture Notes in Statistics*, ed. by E. Parzen. New York: Springer-Verlag, 224–250.

PROTTER, P. (1992): *Stochastic Integration and Differential Equations: A New Approach*. New York: Springer-Verlag.

ROBINSON, P. M. (1977): "Estimation of a Time Series Model from Unequally Spaced Data," *Stochastic Processes and Applications*, 6, 9–24.

RYDBERG, T. H., AND N. SHEPHARD (1999): "A Modelling Framework for the Prices and Times of Trades Made on the NYSE," Tech. Rep., Nuffield College, Oxford University.

SIMS, C. (1972): "Money, Income and Causality," *American Economic Review*, 62, 540–552.

WONG, W. H. (1986): "Theory of Partial Likelihood," *Annals of Statistics*, 14, 88–123.