Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 144 (2008) 1–26
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

An analysis of Hansen–Scheinkman moment estimators for discretely and randomly sampled diffusions

Yacine Aït-Sahalia[a,*], Per A. Mykland[b,1]

[a]Department of Economics, Princeton University and NBER, Princeton, NJ 08544-1021, USA
[b]Department of Statistics, The University of Chicago, Chicago, IL 60637-1514, USA

Available online 8 February 2008

Abstract

We derive closed-form expansions for the asymptotic distribution of Hansen and Scheinkman [1995. Back to the future: generating moment implications for continuous-time Markov processes. Econometrica 63, 767–804] moment estimators for discretely, and possibly randomly, sampled diffusions. This result makes it possible to select optimal moment conditions as well as to assess the efficiency of the resulting parameter estimators relative to likelihood-based estimators, or to an alternative type of moment conditions.

JEL classification: C22

Keywords: Diffusions; Discrete sampling; Random sampling; Moment conditions; Efficiency

1. Introduction

Hansen and Scheinkman (1995) (HS thereafter) derived moment conditions for estimating the parameters of continuous-time Markov processes using discrete time data. The HS moment conditions are correctly centered, so the resulting parameter estimators are consistent. One impediment, however, to the wide application of HS moment conditions in practice is the fact that the asymptotic variance of the resulting parameter estimators is not known explicitly, beyond its generic GMM expression (see Conley et al., 1997, p. 540). Indeed, the intervening matrices in the asymptotic variance take the form of expected values which cannot be calculated explicitly. Since the HS moment conditions involve the choice of a set of "test functions", the selection of optimal test functions would be greatly facilitated if one could, for instance, analyze their impact on the variance of the parameter estimators in closed form. So would the comparison with alternative estimation strategies, such as likelihood-based inference. These are the objectives of this paper.

Furthermore, in typical quote or transaction-level financial data, not only are the observations sampled discretely in time, but it is often the case that the time separating successive observations is itself random. In

*Corresponding author. Tel.: +1 609 258 4015.
E-mail addresses: yacine@princeton.edu (Y. Aït-Sahalia), mykland@galton.uchicago.edu (P.A. Mykland).
[1]Tel.: +1 773 702 8044.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.11.002

Aït-Sahalia and Mykland (2003), we developed methods to analyze the distribution of likelihood-based estimators for diffusions under these circumstances, compared the relative impact of discrete vs. random sampling, and in Aït-Sahalia and Mykland (2004) provided a general approach to deriving explicitly the asymptotic properties of estimators based on arbitrary moment conditions. In regular circumstances that are satisfied by the HS moment conditions, the estimator $\hat{\beta}$ of the parameter vector β_0 is consistent and $\sqrt{T}(\hat{\beta}-\beta_0)$ converges in law to $N(0,\Omega_\beta)$ as the time span T over which observations occur tends to infinity.

For any such estimator, the corresponding asymptotic variance Ω_β are generally unknown in closed form. The solution we proposed is to derive Taylor expansions for Ω_β starting with a leading term that corresponds to the limiting case where the sampling is continuous in time. The expansion is with respect to a parameter ε which indexes the sampling intervals separating successive observations, as in $\Delta \overset{L}{=} \varepsilon\Delta_0$, where Δ_0 is possibly random with a given fixed distribution. Sampling at a deterministic constant interval corresponds to the special case where $\mathrm{Var}[\Delta_0]=0$. Our Taylor expansions are of the form

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

While the limiting term as ε goes to zero corresponds to continuous sampling, by adding higher-order terms in ε, we progressively correct this leading term for the discreteness of the sampling. This method can then be used to analyze the relative merits of different estimation approaches, by comparing the order in ε at which various effects manifest themselves, and when they are equal the relative magnitudes of the corresponding coefficients in the expansion.

In this paper, we apply and extend these tools to the specific set of HS moment conditions for diffusions, analyze the properties of the estimators and compare them to the Cramér–Rao lower bounds. In particular, we give explicit expressions for the asymptotic variance matrix of the HS estimators in the Taylor series form (1) for arbitrary test functions. We then turn to the determination of optimal test functions and the relative efficiency of the resulting estimators compared to likelihood benchmarks. Both are made possible by the explicit computation of (1).

Let the parameter vector be written as $\beta' = (\theta,\gamma)$ where θ is the parameter entering the drift function and γ that entering the diffusion function. HS propose two sets of moment conditions, called C1 and C2, respectively, whose definition we will recall below. C1 is based on the stationary distribution of the process only, while C2 involves its transitions over the time interval corresponding to the frequency of observation.

A quick summary of our results is as follows:

- In the case of estimating θ, for known γ, our message is upbeat. The C1 and C2 estimators are fully efficient to first order in ε, confirming the result of Conley et al. (1997): based on the leading term in our expansion, namely the continuous record limit, the distinction between fixed and random sampling is irrelevant. Moreover, they are also very close to being efficient to second order in ε. There, the asymptotic variance is proportional to $\mathrm{Var}[\Delta_0]$. Therefore that result is inherently dependent on the randomness of the sampling. On the other hand, a perhaps disappointing result is that up to second order in ε, the C2 estimator is no more efficient than the C1 estimator.
- For estimating γ, however, the efficiency is substantially inferior to that of the likelihood estimate (by an order in ε). Assuming one is not going to use the likelihood, it would seem that a good way of using the C1 and C2 estimators is therefore to estimate γ by some other method, and then to estimate θ using C1 or C2. In view of the existing results on volatility inference for high frequency data, this is a feasible approach: a related estimation strategy is proposed by Phillips and Yu (2005).

The paper is organized as follows. Section 2 sets up the model and summarizes our approach to analyze the asymptotic variance of general estimators in the context of discretely and possibly randomly sampled estimators of diffusions. Section 3 then applies the method to derive closed-form expansions for the asymptotic variance of HS estimators. In Section 4, we use these expressions to study the choice of optimal test functions and the efficiency of HS estimators relative to likelihood-based estimators. An application of these results to a specific example of a diffusion process is contained in Section 5. Section 6 concludes, while proofs are in the Appendix.

2. The setup

This paper shares a common setup with our earlier work on the topic of estimating discretely and randomly sampled diffusions using either likelihood or generic moment conditions (Aït-Sahalia and Mykland, 2003, 2004). We start by briefly summarizing this theory in a special case (consistent estimators, exact identification, restrictive choice of moment functions, scalar parameters) that will be sufficient for its application to the HS moment conditions.

Suppose that we observe a stationary diffusion process

$$\mathrm{d}X_t = \mu(X_t;\theta)\,\mathrm{d}t + \sigma(X_t;\gamma)\,\mathrm{d}W_t \tag{2}$$

at discrete times in the interval $[0, T]$, and we wish to estimate the parameter vector $\beta' = (\theta, \gamma)$ which lies in an open and bounded set.

For simplicity of notation only, we assume that both θ and γ are scalars. Going from one to d parameters in each of the drift and diffusion functions presents no conceptual difficulties, although it complicates the expression of the various results. Subject to basic smoothness of the drift and diffusion functions, all the drift (resp. diffusion) behave identically as far as the rates of convergence—hence ultimately their efficiency—are concerned; the important difference is between drift and diffusion parameters, and is already made apparent in the one-parameter-in-drift, one-parameter-in-diffusion, case we are focusing on here.

We make the same primitive assumptions on (μ, σ) as in Aït-Sahalia and Mykland (2004, Assumption 1, pp. 2188–2189). Note that these assumptions imply the existence of a weak solution of the differential equation (2), see for example Karatzas and Shreve (1991, Chapter 5.3). We emphasize that all the probability measures are defined on the same sample space, which can be taken to be the set of continuous functions on $[0, T]$, or any extension thereof. Likelihood ratios (Radon–Nikodym derivatives) are therefore well defined subject to measure theoretic equivalence. In fact, such likelihood ratios, in conjunction with Girsanov's Theorem, are a key tool to show existence and uniqueness of the stochastic differential equation, as discussed in Karatzas and Shreve (1991).

The assumptions made also imply that the diffusion process is stationary; its stationary density is then

$$\pi(x;\beta) = \frac{\xi(\beta)\exp\{2\int^x (\mu(y;\theta)/\sigma^2(y;\gamma))\,\mathrm{d}y\}}{\sigma^2(x;\gamma)}, \tag{3}$$

where the lower bound of integration is an arbitrary point in the domain $\mathscr{S} = (\underline{x}, \bar{x})$ of the diffusion and $\xi(\beta)$ is a constant designed to make π integrate to 1.

The observation times on the X process are $\tau_0 = 0, \tau_1, \tau_2, \ldots, \tau_{N_T}$, where N_T is the smallest integer such that $\tau_{N_T+1} > T$. In other words, we observe $Y_0, \Delta_1, Y_1, \Delta_2, Y_2, \ldots, \Delta_{N_T}, Y_{N_T}$ where $Y_i = X_{\tau_i}$. We assume Assumption 2 in Aït-Sahalia and Mykland (2004, p. 2190): the sampling intervals $\Delta_n = \tau_n - \tau_{n-1}$ are independent and identically distributed, Δ_n is drawn from a common distribution which is independent of Y_{n-1} and of the parameter β, and $\mathrm{E}[\Delta_0^2] < +\infty$. This assumption is undoubtedly restrictive in light of the empirical fact that durations appear to be serially correlated, and causally related to the price process (see Renault and Werker, 2003 for an analysis of the impact of this on volatility measurement).

The analysis that follows is nevertheless a first step away from sampling at a fixed time interval. And as we shall see, even under our restrictive i.i.d. sampling assumption, random sampling can have non-trivial effects on the estimators.

Throughout the paper, we denote by Δ a generic random variable with the common distribution of the Δ_n's and write

$$\Delta = \varepsilon\Delta_0, \tag{4}$$

where Δ_0 has a given fixed (but unknown) distribution (independent of ε), and ε is deterministic. While we assume that the distribution of the sampling intervals is independent of β, it may well depend upon its own nuisance parameters (such as an unknown arrival rate, for instance). An important special case occurs when the sampling happens to take place at a fixed deterministic interval $\bar{\Delta}$, corresponding to the distribution of Δ_n being a Dirac mass at $\bar{\Delta}$ and $\mathrm{Var}[\Delta_0] = 0$.

We shall see below that in some cases the first-order term in the asymptotic variance is proportional to $\mathrm{Var}[\Delta_0]$, so as soon as one goes beyond the limit of continuous sampling, the randomness of the sampling is the next order effect.

We consider moment conditions $h(y_1, y_0, \delta, \beta, \varepsilon)$, which are continuously differentiable in β. As we discuss below, the HS method cannot fully identify all the parameters jointly, so we will estimate θ with γ known, or vice versa, but not both together. To avoid unnecessary notation, consider a single moment condition at a time; we extend this to the overidentified case in Section 4.4. In standard GMM fashion, we form the sample average

$$m_T(\beta) \equiv N_T^{-1} \sum_{n=1}^{N_T-1} h(Y_n, Y_{n-1}, \Delta_n, \beta, \varepsilon) \tag{5}$$

and obtain $\hat{\beta}$ by setting $m_T(\beta)$ to 0 in this exactly identified case. Consistency of $\hat{\beta}$ is achieved if

$$\mathrm{E}_{\Delta, Y_1, Y_0}[h(Y_1, Y_0, \Delta, \beta_0, \varepsilon)] = 0, \tag{6}$$

where we denote by $\mathrm{E}_{\Delta, Y_1, Y_0}$ expectations taken with respect to the joint law of (Δ, Y_1, Y_0) at the true parameter β_0, and write E_{Δ, Y_1}, etc., for expectations taken from the appropriate marginal laws of (Δ, Y_1), etc.

Under regularity assumptions on h discussed below, and satisfied in particular by the HS moment functions, we have that $\sqrt{T}(\hat{\beta} - \beta_0) \to \mathrm{N}(0, \Omega_\beta)$, with

$$\Omega_\beta^{-1} = (\mathrm{E}[\Delta])^{-1} D_\beta' S_\beta^{-1} D_\beta, \tag{7}$$

where

$$D_\beta \equiv \mathrm{E}_{\Delta, Y_1, Y_0}[\dot{h}(Y_1, Y_0, \Delta, \beta_0, \varepsilon)],$$
$$S_{\beta,j} \equiv \mathrm{E}_{\Delta, Y_1, Y_0}[h(Y_{1+j}, Y_j, \Delta, \beta_0, \varepsilon) h(Y_1, Y_0, \Delta, \bar{\beta}, \varepsilon)']$$

and $S_\beta \equiv \sum_{j=-\infty}^{+\infty} S_{\beta,j} \equiv S_{\beta,0} + T_\beta$.

To compute these expected values, we rely on the standard infinitesimal generator A_{β_0}. This is the operator which returns

$$A_{\beta_0} \cdot f = \frac{\partial f}{\partial \delta} + \mu(y_1, \theta_0) \frac{\partial f}{\partial y_1} + \frac{1}{2} \sigma^2(y_1; \gamma_0) \frac{\partial^2 f}{\partial y_1^2} \tag{8}$$

when applied to functions f that are continuously differentiable once in δ, twice in y_1 and such that $\partial f / \partial y_1$ and $A_{\beta_0} \cdot f$ are both in L^2 and satisfy

$$\lim_{y_1 \to \underline{x}} \frac{\partial f / \partial y_1}{s(y_1; \beta)} = \lim_{y_1 \to \bar{x}} \frac{\partial f / \partial y_1}{s(y_1; \beta)} = 0, \tag{9}$$

where $s(x; \beta) \equiv \exp\{-2 \int^x (\mu(; y\theta)/\sigma^2(y; \gamma)) \, \mathrm{d}y\}$ is the scale density of the process (see Hansen et al., 1998). We denote by L^2 the Hilbert space of measurable real-valued functions f on $\mathscr{S}$ such that $\|f\|^2 \equiv \mathrm{E}[f(X_0)^2] < \infty$ with the expectation computed at the true value β_0. We define $\mathscr{D}$ to be the set of functions f which have these properties and are additionally continuously differentiable in β, and ε.

To calculate Taylor expansions in ε of the asymptotic variances when the sampling intervals are random, we introduced in Aït-Sahalia and Mykland (2003) the *generalized infinitesimal operator* Γ_{β_0} for the process X in (2). Our operator Γ_{β_0} is defined by its action on $f \in \mathscr{D}$ as follows[1]:

$$\Gamma_{\beta_0} \cdot f \equiv \Delta_0 A_{\beta_0} \cdot f + \frac{\partial f}{\partial \varepsilon}. \tag{10}$$

Note that Γ_{β_0} is a random operator which takes a fixed (or random) function into a random one. Define $\mathscr{D}^J$ as the set of functions f which with $J+2$ continuous derivatives in δ, $2(J+2)$ in y_1, such that f and its first J iterates by repeated applications of A_{β_0} all remain in $\mathscr{D}$ and additionally have $J+2$ continuous derivatives in β and ε.

[1]That operator in general contains an additional term which is zero under (6), as will be the case throughout this paper.

The Taylor expansion of a function $f(Y_1, Y_0, \Delta, \bar{\beta}, \varepsilon) \in \mathscr{D}^J$ is

$$\mathrm{E}_{Y_1}[f(Y_1, Y_0, \Delta, \bar{\beta}, \varepsilon)|\, Y_0, \Delta] = \sum_{j=0}^{J} \frac{\varepsilon^j}{j!}(\Gamma^j_{\beta_0} \cdot f)(Y_0, Y_0, 0, \beta_0, 0) + O_{\mathrm{p}}(\varepsilon^{J+1}). \tag{11}$$

Note that $\Delta = \varepsilon\Delta_0$, and both ε and Δ_0 appear on the right-hand side of the equation, the latter as part of the random operator Γ_{β_0}. All the expectations are taken with respect to the law of the process at the true value β_0.

The usefulness of this approach lies in its ability to deliver closed-form expressions for the terms of the Taylor series in (11) for arbitrary choices of h, including the special case of HS moment functions. Relative to Assumption 3 in Aït-Sahalia and Mykland (2004, p. 2193), HS moment conditions are a special case that does not exhibit a singularity, so $H = 0$ there, and Assumption 3 reduces to $h \in \mathscr{D}^J$ for some $J \geqslant 3$, which we assume here.

Let us close this section with a remark on the asymptotics. Even though we create a Taylor expansion in ε, the asymptotics is of the standard 'large T' variety. The expansion in ε is just a convenient way of analyzing the asymptotic results. This is different from the form of asymptotics where $T \to \infty$ and $\varepsilon \to 0$ at the same time. A third form of asymptotics lets $\varepsilon \to 0$ for fixed T. The latter two are also useful means of analysis. One of the advantages of the approach we have adopted is that it allows for explicit calculations of the asymptotic variance expansion.

In any event, we view these different types of asymptotics as complements rather than substitutes. And given the current state of our knowledge, we cannot say with confidence that one type of asymptotic analysis is necessarily superior to another. Ultimately, they are all approximations to the real small-sample situation. Comparing them would be very useful, but undertaking this would go substantially beyond the scope of the present paper, as it would require the development of a different machinery to compute the distributional properties of the estimators under those alternative asymptotics.

3. Estimators based on HS estimating equations

We now make these results specific in the special case of the HS class of moment functions, and compare them to efficient estimators. The HS moment conditions are in the form of expectations of the infinitesimal generator, one unconditional and one conditional, that can be applied to test functions. HS give two ways of forming estimating functions in the case of sampling at a fixed deterministic Δ, which are referred to as the C1 and C2 moment conditions. In what follows, we apply our general theory to determine the asymptotic properties of estimators using these estimating equations; our results give these properties when the sampling intervals are fixed and deterministic, but also when they are random.

The simplifying feature of the method of moments approach, which is not specific to the context of discretely sampled diffusions, is that it requires only the specification of a set of moments rather than the full conditional density of the diffusion. The flip side of this simplification, however, is that it will not, in general, make efficient use of the entire information contained in the sample. We will characterize precisely this loss of information in our specific context of discrete sampling from a diffusion.

Also, unlike the typical use of the method of moments, one cannot in general select as moment conditions within this framework the "natural" conditional moments of the process since explicit expressions for the conditional mean, variance, skewness, or first-order Euler equations from an optimization problem, etc. are not available in closed form. Rather, the moment conditions, i.e., our h functions, are in the form of the infinitesimal generator of the process applied to arbitrary test functions. As a result, it is useful to be able to obtain explicit expressions for the asymptotic variance of the estimator based on given test functions, as our methodology will allow, with an eye towards selecting optimal test functions. We will address the efficiency question in Section 4.

Kessler and Sørensen (1999) proposed to use the eigenfunctions of the infinitesimal operator as test functions; unfortunately, these are not explicit either, except in special cases. Duffie and Glynn (2004) introduced a family of GMM estimators with Poisson sampling occurring at an arrival intensity that can depend on X and on β. However, the method is specific to the type of random sampling assumed: in particular, it does not allow for the important special case where sampling occurs at fixed time intervals.

One additional aspect of the HS method is that it does not permit full identification of all the parameters of the model since multiplying the drift and diffusion functions by the same constant results in identical moment conditions. So parameters are only identified up to scale. For instance, in the Ornstein–Uhlenbeck example of Section 5, only the stationary variance $\gamma/(2\theta)$ can be identified, but not θ and γ separately. Because of this limitation of the method, we will use the method to estimate θ with γ known, or vice versa, but not both together.

3.1. The C1 moment condition

Let us start by analyzing C1, as in the empirical implementation of the method in Conley et al. (1997). The C1 method takes a sufficiently differentiable function $\psi(y_0, \beta)$ in the domain of the operator B_β defined below and forms the estimating function which in our notation is given by

$$h_{C1}(y_1, y_0, \delta, \beta, \varepsilon) = h_{C1}(y_0, \beta) \equiv B_\beta \cdot \psi(y_0, \beta) \equiv \mu(y_0, \theta)\frac{\partial\psi}{\partial y_0} + \frac{1}{2}\sigma^2(y_0; \gamma)\frac{\partial^2\psi}{\partial y_0^2}. \tag{12}$$

This is a function of (y_0, β) only. Note that the operator B_β differentiates with respect to the backward state variable y_0 as opposed to the forward state variable y_1 (as in our definition of A_β).

The C1 estimating equation relies on the fact that we have the unbiasedness condition

$$\mathrm{E}_{Y_0}[B_{\beta_0} \cdot \psi(Y_0, \beta)] = 0. \tag{13}$$

Once B_β is evaluated at β_0, this is true for any value of β in ψ, including β_0. A consequence of this is that the estimator is consistent because h_{C1} evaluated at β_0 has unconditional mean zero: recall (6).

Eq. (13) follows from the fact that X is a stationary process, hence the unconditional expectation of any function of X_t, such as $\mathrm{E}_{X_t}[\psi(X_t, \beta)]$, does not depend upon the date t at which it is evaluated: thus $(\partial/\partial t)\mathrm{E}_{X_t}[\psi(X_t, \beta)] = 0$, from which the result follows.

In our setup, h_{C1} only depends on (y_0, β), and not on $(y_1, \delta, \varepsilon)$. It follows that

$$(\Gamma_{\beta_0} \cdot h_{C1})(y_1, y_0, \delta, \beta, \varepsilon) \equiv 0 \tag{14}$$

identically, and hence the expansions of h_{C1} (for the function q_1), $\dot{h}_{C1}$ (for D_β) and $h_{C1} \times h_{C1}$ (for $S_{\beta,0}$) will stop at their leading term and be exact.

Of course, we could equivalently have taken the moment function to be of the form

$$\tilde{h}_{C1}(y_1, \beta) \equiv A_\beta \cdot \psi(y_1, \beta) = \mu(y_1, \theta)\frac{\partial\psi}{\partial y_1} + \frac{1}{2}\sigma^2(y_1; \gamma)\frac{\partial^2\psi}{\partial y_1^2} \tag{15}$$

i.e., as a function of y_1 instead of y_0. We would get the same result since the unconditional expectation of any function $f(Y_1, \beta) \in \mathscr{D}^J$ is obtained by computing in our method

$$\mathrm{E}_{Y_1}[f(Y_1, \beta)|Y_0, \Delta] = \sum_{j=0}^{J} \frac{\varepsilon^j}{j!}(\Gamma^j_{\beta_0} \cdot f)(Y_0, \beta_0) + O_p(\varepsilon^{J+1}). \tag{16}$$

Next, $(\Gamma^j_{\beta_0} \cdot f)(Y_0, \beta_0) = \Delta_0^j(A^j_{\beta_0} \cdot f)(Y_0, \beta_0)$ since $\partial f/\partial\varepsilon = 0$ and $\partial\beta/\partial\varepsilon = 0$ given that the estimating equation is unbiased. When taking unconditional expectations, we have

$$\mathrm{E}_{Y_0}[(A^j_{\beta_0} \cdot f)(Y_0, \beta_0)] = 0 \tag{17}$$

for all $j \geqslant 1$ because the expected value of the generator applied to any function is zero—this is indeed (13). That is, the expansion for the unconditional expectation over the law of Y_0 will stop after the leading $(j = 0)$ term. Therefore, computing $\mathrm{E}_{Y_1}[f(Y_1, \beta)]$ as prescribed by our method, i.e., through the law of iterated expectations in the form $\mathrm{E}_{\Delta, Y_0}[\mathrm{E}_{Y_1}[f(Y_1, \beta)|Y_0, \Delta]]$, will produce the same result as writing down directly $\mathrm{E}_{Y_0}[f(Y_0, \beta)]$. In other words, using (12) or (15) as moment functions will yield the same results. The form (12) gives the result directly, and we will therefore use it.

Because of the form of h_{C1}, the only difference between estimating θ and estimating γ appears in D_β. Also, because $h_{C1}(y_0, y_0, 0, \beta_0, 0)$ is non-zero, we have $\alpha_{C1} = 0$. The specific expressions are

$$q_{C1}(Y_0, \beta_0, \varepsilon) = q_{C1}(Y_0, \beta_0, 0) = B_{\beta_0} \cdot \psi(Y_0, \beta_0) \tag{18}$$

and

$$D_\beta = D_\beta^{(0)} = \begin{cases} D_\theta = \mathrm{E}_{Y_0}\left[\dfrac{\partial \mu(Y_0, \theta_0)}{\partial \theta}\dfrac{\partial \psi(Y_0, \beta_0)}{\partial y}\right] & \text{when estimating } \theta, \\ D_\gamma = \dfrac{1}{2}\mathrm{E}_{Y_0}\left[\dfrac{\partial \sigma^2(Y_0, \gamma_0)}{\partial \gamma}\dfrac{\partial^2 \psi(Y_0, \beta_0)}{\partial y^2}\right] & \text{when estimating } \gamma, \end{cases} \tag{19}$$

$$S_{\beta,0} = S_{\beta,0}^{(0)} = -\frac{1}{2}\mathrm{E}_{Y_0}\left[\sigma^2(Y_0, \gamma_0)\frac{\partial \mu(Y_0, \theta_0)}{\partial y}\left(\frac{\partial \psi(Y_0, \beta_0)}{\partial y}\right)^2\right] + \frac{1}{4}\mathrm{E}_{Y_0}\left[\sigma^4(Y_0, \gamma_0)\left(\frac{\partial^2 \psi(Y_0, \beta_0)}{\partial y^2}\right)^2\right]. \tag{20}$$

The more difficult calculation involves the time series term $T_\beta = S_\beta - S_{\beta,0}$. As part of the proof of the following theorem, we show that

$$T_\beta = \varepsilon^{-1} T_\beta^{(-1)} + T_\beta^{(0)} + O(\varepsilon), \tag{21}$$

where

$$T_\beta^{(-1)} = \frac{2}{\mathrm{E}[\Delta_0]}\mathrm{E}_{Y_0}[(h_{C1} \times r_{C1})], \tag{22}$$

$$T_\beta^{(0)} = \frac{(\mathrm{E}[\Delta_0^2] - 2\mathrm{E}[\Delta_0]^2)}{4\mathrm{E}[\Delta_0]^2}\left\{\mathrm{E}_{Y_0}\left[\sigma^4\left(\frac{\partial^2 \psi}{\partial y^2}\right)^2\right] - 2\mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial \psi}{\partial y}\right)^2\frac{\partial \mu}{\partial y}\right]\right\}. \tag{23}$$

We then put together the expansions of D_β, $S_{\beta,0}$ and T_β to obtain the expansion for Ω_β. The terms of order ε^0 are given by

$$\Omega_\theta^{(0)} = T_\beta^{(-1)}/(D_\theta^{(0)})^2, \quad \Omega_\gamma^{(0)} = T_\beta^{(-1)}/(D_\gamma^{(0)})^2, \tag{24}$$

when estimating θ or γ, respectively, while the terms of order ε^1 are

$$\Omega_\theta^{(1)} = \mathrm{E}[\Delta_0](S_{\beta,0}^{(0)} + T_\beta^{(0)})/(D_\theta^{(0)})^2, \quad \Omega_\gamma^{(1)} = \mathrm{E}[\Delta_0](S_{\beta,0}^{(0)} + T_\beta^{(0)})/(D_\gamma^{(0)})^2. \tag{25}$$

The specific expressions, which characterize the asymptotic properties of the estimators based on the moment condition h_{C1}, are now given in Theorem 1. This and all subsequent theorems are subject to Assumptions 1–3 in Aït-Sahalia and Mykland (2004, pp. 2188–2190, 2193), with $H = 0$ (see the last paragraph in Section 2 in this paper).

Theorem 1 (*Properties of the estimators based on the C1 condition*). *The asymptotic variance has the form*

$$\Omega_\beta = \Omega_\beta^{(0)} + \varepsilon\Omega_\beta^{(1)} + O(\varepsilon^2), \tag{26}$$

where, when estimating θ,

$$\Omega_\theta^{(0)} = \frac{\mathrm{E}_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\left(\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right)^2\right]}{\mathrm{E}_{Y_0}\left[\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right]^2},$$

$$\Omega_\theta^{(1)} = \frac{\mathrm{Var}[\Delta_0]\left(\mathrm{E}_{Y_0}\left[\sigma^4(Y_0,\gamma_0)\left(\frac{\partial^2\psi(Y_0,\beta_0)}{\partial y^2}\right)^2\right] - 2\mathrm{E}_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\frac{\partial\mu(Y_0,\theta_0)}{\partial y}\left(\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right)^2\right]\right)}{4\mathrm{E}[\Delta_0]\mathrm{E}_{Y_0}\left[\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right]^2} \tag{27}$$

and, when estimating γ,

$$\Omega_\gamma^{(0)} = \frac{4\mathrm{E}_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\left(\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right)^2\right]}{\mathrm{E}_{Y_0}\left[\frac{\partial\sigma^2(Y_0,\gamma_0)}{\partial\gamma}\frac{\partial^2\psi(Y_0,\beta_0)}{\partial y^2}\right]^2},$$

$$\Omega_\gamma^{(1)} = \frac{\mathrm{Var}[\Delta_0]\left(\mathrm{E}_{Y_0}\left[\sigma^4(Y_0,\gamma_0)\left(\frac{\partial^2\psi(Y_0,\beta_0)}{\partial y^2}\right)^2\right] - 2\mathrm{E}_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\frac{\partial\mu(Y_0,\theta_0)}{\partial y}\left(\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right)^2\right]\right)}{\mathrm{E}[\Delta_0]\mathrm{E}_{Y_0}\left[\frac{\partial\sigma^2(Y_0,\gamma_0)}{\partial\gamma}\frac{\partial^2\psi(Y_0,\beta_0)}{\partial y^2}\right]^2}. \tag{28}$$

It is interesting to note that the first-order term $\Omega_\beta^{(1)}$ in the asymptotic variance is proportional to $\mathrm{Var}[\Delta_0]$, so the effect of the random sampling can be non-trivial, even under our restrictive sampling assumptions. That is, as soon as one goes beyond the limit of continuous sampling (order 0 in ε), the randomness of the sampling is the next order effect. The same effect will occur with the C2 moment condition, although with an additional term independent of $\mathrm{Var}[\Delta_0]$, as we shall now see.

3.2. The C2 moment condition

Consider now the C2 moment condition. The C2 method takes two functions ψ_0 and ψ_1, again satisfying smoothness and regularity conditions, and forms the "back to the future" estimating function

$$h_{C2}(y_1,y_0,\delta,\beta,\varepsilon) = h_{C2}(y_1,y_0,\beta) = \{A_\beta\cdot\psi_1(y_1,\beta)\}\times\psi_0(y_0,\beta) - \{B_\beta\cdot\psi_0(y_0,\beta)\}\times\psi_1(y_1,\beta). \tag{29}$$

In general, B_β should be replaced by the infinitesimal generator associated with the reverse time process, A_β^*. But under regularity conditions, univariate stationary diffusions are time reversible (see Kent, 1978) and so the infinitesimal generator of the process is self-adjoint and so we can define h_{C2} above using the operator B_β (itself defined in (12)).

The C2 estimating equation relies on the fact that, when the operators A_β and B_β are evaluated at the true parameter β_0, then

$$\mathrm{E}_{Y_0,Y_1}[\{A_{\beta_0}\cdot\psi_1(Y_1,\beta)\}\times\psi_0(Y_0,\beta) - \{B_{\beta_0}\cdot\psi_0(Y_0,\beta)\}\times\psi_1(Y_1,\beta)] = 0. \tag{30}$$

Once A_β is evaluated at β_0, this is true for any value of β in ψ, including β_0. As a result, estimators based on the C2 moment condition are consistent (recall (6)).

Eq. (30) is again a consequence of the stationarity of the process X. Namely, the expectation of any function of $(X_t, X_{t+\delta})$, such as $\mathrm{E}_{X_t,X_{t+\delta}}[\psi_0(X_t,\beta)\psi_1(X_{t+\delta},\beta)]$, does not depend upon the date t (it can of course depend upon the time lag δ between the two observations): hence

$$\frac{\partial}{\partial t}\mathrm{E}_{X_t,X_{t+\delta}}[\psi_0(X_t,\beta)\psi_1(X_{t+\delta},\beta)] = 0 \tag{31}$$

from which (30) follows. Incidentally, C2 can alternatively be obtained, as shown in Aït-Sahalia (1996), by combining the Kolmogorov forward and backward equations characterizing the transition function $p(y_1|y_0,\Delta,\beta)$ in a way that eliminates the (unobservable with discrete data) derivatives of p with respect to time.

When considering this case, it is worthwhile to be explicit about how the D_β, S_β and Ω_β matrices depend on the distributions of Δ_0 and Y_0.

Theorem 2. *If h_{C2} is used to estimate either $\beta=\theta$ or $=\gamma$, then*

$$D_\beta^{(0)}=\tilde{D}_\beta^{(0)}D_\beta^{(1)}=\mathrm{E}[\Delta_0]\tilde{D}_\beta^{(1)},$$

$$S_\beta^{(-1)}=\frac{1}{\mathrm{E}[\Delta_0]}\tilde{S}_\beta^{(-1)}S_\beta^{(0)}=\tilde{S}_\beta^{(0)}+\frac{\mathrm{Var}[\Delta_0]}{\mathrm{E}[\Delta_0]^2}S_{\beta,0}^{(0)}, \tag{32}$$

where $\tilde{D}_\beta^{(0)}$, $\tilde{D}_\beta^{(1)}$, $\tilde{S}_\beta^{(-1)}$, $\tilde{S}_\beta^{(0)}$, and $S_{\beta,0}^{(0)}$ depend only on ψ_0, ψ_1, μ, σ^2 and the distribution of Y_0 (and not on the distribution of Δ_0). Specifically,

$$D_\beta^{(0)}=\begin{cases} D_\theta^{(0)}=\mathrm{E}_{Y_0}\left[\dfrac{\partial\mu}{\partial\theta}\left(\dfrac{\partial\psi_1}{\partial y}\psi_0-\psi_1\dfrac{\partial\psi_0}{\partial y}\right)\right] & \textit{when estimating } \theta,\\ D_\gamma^{(0)}=\dfrac{1}{2}\mathrm{E}_{Y_0}\left[\dfrac{\partial\sigma^2}{\partial\gamma}\left(\dfrac{\partial^2\psi_1}{\partial y^2}\psi_0-\psi_1\dfrac{\partial^2\psi_0}{\partial y^2}\right)\right] & \textit{when estimating } \gamma,\end{cases} \tag{33}$$

$$\tilde{D}_\beta^{(1)}=\begin{cases} \tilde{D}_\theta^{(1)}=\dfrac{1}{2}\mathrm{E}_{Y_0}\left[\sigma^2\dfrac{\partial\mu}{\partial\theta}\left(\dfrac{\partial^2\psi_0}{\partial y^2}\dfrac{\partial\psi_1}{\partial y}-\dfrac{\partial\psi_0}{\partial y}\dfrac{\partial^2\psi_1}{\partial y^2}\right)\right] & \textit{when estimating } \theta,\\ \tilde{D}_\gamma^{(1)}=\dfrac{1}{4}\mathrm{E}_{Y_0}\left[\sigma^2\dfrac{\partial\sigma^2}{\partial\gamma}\left(\dfrac{\partial^3\psi_0}{\partial y^3}\dfrac{\partial\psi_1}{\partial y}-\dfrac{\partial\psi_0}{\partial y}\dfrac{\partial^3\psi_1}{\partial y^3}\right)\right. & \\ \qquad\left.+\sigma^2\dfrac{\partial^2\sigma^2}{\partial y\partial\gamma}\left(\dfrac{\partial^2\psi_0}{\partial y^2}\dfrac{\partial\psi_1}{\partial y}-\dfrac{\partial\psi_0}{\partial y}\dfrac{\partial^2\psi_1}{\partial y^2}\right)\right] & \textit{when estimating } \gamma,\end{cases} \tag{34}$$

$$S_{\beta,0}^{(0)}=\mathrm{E}_{Y_0}[(\{A_{\beta_0}\cdot\psi_1(Y_0,\beta_0)\}\times\psi_0(Y_0,\beta_0)-\{B_{\beta_0}\cdot\psi_0(Y_0,\beta_0)\}\times\psi_1(Y_0,\beta_0))^2], \tag{35}$$

$$\tilde{S}^{(-1)}=\mathrm{E}_{Y_0}\left[\sigma^2\left(\psi_1\frac{\partial\psi_0}{\partial y}-\psi_0\frac{\partial\psi_1}{\partial y}\right)^2\right], \tag{36}$$

$$\tilde{S}^{(0)}=2(\mathrm{E}_{Y_0}[A_{\beta_0}\cdot(h_{\mathrm{C2}}\times\check{r}_{\mathrm{C2}})]+S_{\beta,0}^{(0)}), \tag{37}$$

where the function $\check{r}$ is defined in Aït-Sahalia and Mykland (2004, equation (29), p. 2196).

We can now state the asymptotic properties of the estimators based on the C2 moment condition:

Theorem 3 (*Properties of the estimators based on the C2 condition*). *If h_{C2} is used to estimate either θ or γ, we have*

$$\Omega_\beta=\Omega_\beta^{(0)}+\varepsilon\Omega_\beta^{(1)}+O(\varepsilon^2), \tag{38}$$

where, when estimating θ,

$$\Omega_\theta^{(0)}=\frac{\mathrm{E}_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\left(\psi_1(Y_0,\beta_0)\dfrac{\partial\psi_0(Y_0,\beta_0)}{\partial y}-\psi_0(Y_0,\beta_0)\dfrac{\partial\psi_1(Y_0,\beta_0)}{\partial y}\right)^2\right]}{\mathrm{E}_{Y_0}\left[\dfrac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\left(\dfrac{\partial\psi_1(Y_0,\beta_0)}{\partial y}\psi_0(Y_0,\beta_0)-\psi_1(Y_0,\beta_0)\dfrac{\partial\psi_0(Y_0,\beta_0)}{\partial y}\right)\right]^2} \tag{39}$$

and, when estimating γ,

$$\Omega^{(0)}_{\sigma^2} = \frac{4\mathrm{E}_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\left(\psi_1(Y_0,\beta_0)\dfrac{\partial\psi_0(Y_0,\beta_0)}{\partial y} - \psi_0(Y_0,\beta_0)\dfrac{\partial\psi_1(Y_0,\beta_0)}{\partial y}\right)^2\right]}{\mathrm{E}_{Y_0}\left[\left(\dfrac{\partial^2\psi_1(Y_0,\beta_0)}{\partial y^2}\psi_0(Y_0,\beta_0) - \psi_1(Y_0,\beta_0)\dfrac{\partial^2\psi_0(Y_0,\beta_0)}{\partial y^2}\right)\right]^2}. \tag{40}$$

The expressions for $\Omega^{(1)}_\theta$ *and* $\Omega^{(1)}_{\sigma^2}$, *which are more involved, are summarized by*

$$\Omega^{(1)}_\beta = \mathrm{E}[\Delta_0]\tilde{\Omega}^{(1)}_\beta + \frac{\mathrm{Var}[\Delta_0]}{\mathrm{E}[\Delta_0]}\frac{S^{(0)}_{\beta,0}}{(D^{(0)}_{\beta,0})^2}, \tag{41}$$

where

$$\tilde{\Omega}^{(1)}_\beta = \frac{\tilde{S}^{(0)}_\beta \tilde{D}^{(0)}_\beta - \tilde{S}^{(-1)}_\beta \tilde{D}^{(1)}_\beta}{(\tilde{D}^{(0)}_\beta)^3}. \tag{42}$$

The form given in (41)–(42) will be useful when assessing the efficiency properties of these estimators, which is the question we shall now investigate.

4. Efficiency properties of the Hansen–Scheinkman estimators

4.1. Comparison with likelihood-based estimators

In Aït-Sahalia and Mykland (2003), we studied the effect that the randomness of the sampling intervals might have when estimating a continuous-time model with discrete data, as would be the case with transaction-level returns data. We disentangled the effect of the sampling *randomness* from the effect of the sampling *discreteness*, and compare their relative magnitudes. We also examined the effect of simply ignoring the sampling randomness. We achieved this by comparing the properties of different likelihood-based estimators, which make different use of the observations on the state process and the times at which these observations have been recorded.

One of the estimators of β we considered is the full information maximum likelihood (FIML) estimator, using the bivariate observations (Y_n, Δ_n); another 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). Not surprisingly, 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.

These 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. Because of the time homogeneity of the model, the transition function p depends only on δ and not on $(t, t+\delta)$ separately. FIML makes use of $p(Y_n|Y_{n-1},\Delta_n,\beta)$, while IOML uses the expectation $\tilde{p}(Y_n|Y_{n-1},\beta)$ of $p(Y_n|Y_{n-1},\Delta_n,\beta)$ over the law of $\Delta_n|Y_{n-1}$. In practice, even though most diffusion models do not admit closed-form transition densities, the 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. FIML and IOML are always consistent estimators.

We are here particularly interested in comparing the C1 and C2 estimators with the FIML and IOML estimators. From Aït-Sahalia and Mykland (2003), we have that

$$\Omega^{\mathrm{FIML}}_\theta = \Omega^{(\mathrm{FIML},0)}_\theta + O(\varepsilon^2), \tag{43}$$

$$\Omega^{\mathrm{IOML}}_\theta = \Omega^{(\mathrm{IOML},0)}_\theta + \varepsilon\Omega^{(\mathrm{IOML},1)}_\theta + O(\varepsilon^2), \tag{44}$$

where

$$\Omega_\theta^{(\text{FIML},0)} = \Omega_\theta^{(\text{IOML},0)} = (\mathrm{E}_{Y_0}[\sigma^{-2}(Y_0,\gamma_0)(\partial\mu(Y_0,\theta_0)/\partial\theta)^2])^{-1} \tag{45}$$

which is the leading term in Ω_θ corresponding to efficient estimation of θ with a continuous record of observations. (The expressions are given in Aït-Sahalia and Mykland, 2003 for $\sigma^2 = \gamma^2$, but the extension to general $\sigma^2(y;\gamma)$ follows from the developments in Aït-Sahalia and Mykland, 2004.)

And the price of ignoring the sampling times $\tau_0, \tau_1, \ldots$ when estimating θ is, to first order, represented by

$$\Omega_\theta^{(\text{IOML},1)} = \frac{\mathrm{E}[\mathrm{Var}[\Delta_0|\chi_1^2\Delta_0]]}{\mathrm{E}[\Delta_0]}V, \tag{46}$$

and "χ_1^2" is a χ_1^2 distributed random variable independent of Δ_0, and

$$V = \frac{\left(\mathrm{E}_{Y_0}\left[\sigma_0^4\left(\frac{\partial^2\mu(Y_0,\beta_0)}{\partial y\partial\theta}\right)^2\right] - 2\mathrm{E}_{Y_0}\left[\sigma_0^2\frac{\partial\mu(Y_0,\theta_0)}{\partial y}\left(\frac{\partial\mu(Y_0,\beta_0)}{\partial\theta}\right)^2\right]\right)}{4\mathrm{E}_{Y_0}\left[\left(\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\right)^2\right]^2}. \tag{47}$$

Note that $V \geqslant 0$ by the asymptotic efficiency of FIML.

And the leading term in Ω_γ corresponding to efficient estimation of γ is

$$\Omega_\gamma^{\text{FIML}} = \varepsilon\Omega_\gamma^{(\text{FIML},1)} + O(\varepsilon^2), \tag{48}$$

$$\Omega_\gamma^{\text{IOML}} = \varepsilon\Omega_\gamma^{(\text{IOML},1)} + O(\varepsilon^2), \tag{49}$$

where

$$\Omega_\gamma^{(\text{FIML},1)} = \Omega_\gamma^{(\text{IOML},1)} = \mathrm{E}[\Delta_0](2\mathrm{E}_{Y_0}[(\partial\sigma(Y_0,\gamma_0)/\partial\gamma)^2\sigma(Y_0,\gamma_0)^{-2}])^{-1}. \tag{50}$$

In the special case where σ^2 is constant ($\gamma = \sigma^2$), this becomes the standard AVAR of MLE from i.i.d. Gaussian observations, i.e., $\Omega_\gamma^{(1)} = 2\sigma_0^4\mathrm{E}[\Delta_0]$.

These leading terms are achieved in particular when h is the likelihood score for θ and γ, respectively, as analyzed in Aït-Sahalia and Mykland (2003), but also by other estimating functions that are able to mimic the behavior of the likelihood score at the leading order. So, we now turn to a comparison of the AVAR of these two estimators to the likelihood-based FIML and IOML to find out whether this is the case for these classes of moment conditions.

4.2. *Efficiency of the C1 estimator*

Using Theorem 1, we can study the first-order efficiency of the C1 estimator relative to the likelihood-based estimators. For the purpose of estimating either θ for γ known, or vice versa, or more generally for a scalar parameter β so that $\theta = \theta(\beta)$ and $\gamma = \gamma(\beta)$, Conley et al. (1997) propose to use ψ given by

$$\frac{\partial\psi(y,\beta)}{\partial y} = \frac{\partial}{\partial\beta}\left(\frac{2\mu(y,\theta) - \partial\sigma^2(y,\gamma)/\partial y}{\sigma^2(y,\gamma)}\right). \tag{51}$$

This choice of ψ yields a C1 estimator $A_\beta \cdot \psi$ which is "test function efficient": see Conley et al. (1997, Sections 3.2–3.3 and Appendix C) where they show that this choice is approximately optimal among the class of moment conditions they consider, in the sense of being optimal in the limit of continuous sampling, corresponding to $\varepsilon \to 0$ in our setting. In the case of estimating θ for γ known, this in the same as saying that (51) minimizes $\Omega_\theta^{(0)}$ (which yields the same variance as (45)). Similarly, in the case of estimating γ for θ known, (51) minimizes $\Omega_\gamma^{(0)}$.

This choice of test function corresponds to using as a test function ψ the score from a quasi-ML (QML) estimator that would assume that the data are i.i.d. with distribution given by the stationary density π

given in (3). Indeed (51) corresponds to making ψ proportional to $\partial \log \pi/\partial \beta$ since

$$\frac{\partial \log \pi(y,\beta)}{\partial y} = \frac{2\mu(y,\theta) - \partial\sigma^2(y,\gamma)/\partial y}{\sigma^2(y,\gamma)}. \tag{52}$$

This is not, in general, equivalent to the QML estimator itself, which would make the moment condition $A_\beta \cdot \psi$ proportional to the score $\partial \log \pi/\partial \beta$, as opposed to the test function ψ, so there is an efficiency gain from using (51) instead of the QML estimator. (The two would coincide when $\partial \log \pi/\partial \beta$ is an eigenfunction of A_β.) But of course, the data are not i.i.d., so this is not the FIML which will be more even efficient, as we shall see.

We consider further the estimation of θ for γ known. With Theorem 1, one can see that something stronger than test function efficiency holds. $\Omega_\theta^{(0)}$ for this ψ coincides with the corresponding term $\Omega_\theta^{(\mathrm{FIML},0)}$. In other words, to first order in ε, $A_\beta \cdot \psi$ is fully efficient. This fact is easily seen by substituting (51) into (27), and comparing to the corresponding expression for FIML given in (45).

In view of this efficiency property, it is obvious that this choice of ψ also minimizes the expression (27). This is shown, with different expressions, in Conley et al. (1997, Appendix C).

To consider the efficiency question more carefully, we shall for now fix σ^2 to be independent of y, and continue to use the first-order optimal choice (27). Note that the relevant comparison is not with FIML but rather with IOML. IOML comes from a likelihood which uses the observations $Y_0, Y_1, \ldots$, but not the spacings $\Delta_1, \Delta_2, \ldots$ between the observations. The reason that this is the relevant comparison is that the C1 estimators also do not use these spacings. In view of the Cramér–Rao lower bound, the asymptotic variance of the IOML is the best possible that can be obtained using the partial data $Y_0, Y_1, \ldots$. Hence the discrepancy between the two is then the cost of using the C1 estimator relative to a maximally efficient estimator.

To see what happens, recall V defined in (47). We then have from Theorem 1 that, on the one hand, for the C1 estimator

$$\Omega_\theta^{(\mathrm{C1},1)} = \frac{\mathrm{Var}[\Delta_0]}{\mathrm{E}[\Delta_0]} V. \tag{53}$$

Thus, the cost of using the C1 estimator rather than IOML is summarized by

$$\frac{\Omega_\theta^{(\mathrm{C1},1)}}{\Omega_\theta^{(\mathrm{IOML},1)}} = \frac{\mathrm{Var}[\Delta_0]}{\mathrm{E}[\mathrm{Var}[\Delta_0|\chi_1^2\Delta_0]]}. \tag{54}$$

As it should from the Cramér–Rao lower bound, or can alternatively be seen directly from properties of conditional variances, this quotient is always greater than 1. It also depends only on the distribution of the sampling intervals, i.e., the law of Δ_0. The size of the quantity is explored in Aït-Sahalia and Mykland (2003, Section 5.4, pp. 511–514), where we showed in particular that

$$\mathrm{E}[\mathrm{Var}[\Delta_0|\chi_1^2\Delta_0]] = \mathrm{E}[\Delta_0^2] - \mathrm{E}\left[\left(\chi_1^2\Delta_0 \frac{m_4(\chi_1^2\Delta_0)}{m_2(\chi_1^2\Delta_0)}\right)^2\right], \tag{55}$$

where χ_1^2 and Δ_0 are independent random variables and

$$m_q(b) = \mathrm{E}_Z\left[Z^{-q} d_0\left(\frac{b}{Z^2}\right)\right], \tag{56}$$

where Z is $\mathrm{N}(0,1)$ and d_0 is the density function of Δ_0.

With those results in hand, it is easily seen that, for example, when Δ_0 is exponentially distributed,

$$\frac{\Omega_\theta^{(\mathrm{C1},1)}}{\Omega_\theta^{(\mathrm{IOML},1)}} = \frac{8}{3}. \tag{57}$$

If one wishes to compare to the FIML estimator rather than IOML, recall from (43) that the ε-term in the expansion of the asymptotic variance is zero, so compared to this both the C1 and IOML estimators are inefficient.

For the case of estimating γ for known θ, however, the test function efficiency in C1 does not yield first-order efficiency. Indeed, $\Omega_\gamma^{(\mathrm{C1})}$ is of order $O(1)$ as $\varepsilon \to 0$, that is $\Omega_\gamma^{(\mathrm{C1},0)} > 0$, while the asymptotic variance of both the FIML and the IOML is of order $O(\varepsilon)$. This lack of efficiency is not surprising since volatility estimation is inherently about (squared) changes or increments of the process, and the C1 set of moment conditions uses no information about the increments. One could therefore expect that C2, which is able to utilize the increments, would probably be better suited to estimating the set of γ parameters. As we will see, however, this is not the case.

4.3. Efficiency of the C2 estimator

Surprisingly, to first order in ε, nothing is gained by using the C2 estimator rather than using C1. Specifically, what we mean by this, is that when $\Omega_\theta^{(\mathrm{C2},0)}$ is minimized over ψ_0 and ψ_1, one obtains the same result as when $\Omega_\theta^{(\mathrm{C1},0)}$ is minimized over ψ. And similarly for $\Omega_\gamma^{(\mathrm{C2},0)}$.

To see this, let $\beta = \theta$ or $= \gamma$, and write the first-order asymptotic variances as functionals $\Omega_\beta^{(\mathrm{C1},0)}[\psi]$ and $\Omega_\beta^{(\mathrm{C2},0)}[\psi_0, \psi_1]$. One then sees that $\Omega_\beta^{(\mathrm{C2},0)}[\psi_0, \psi_1] = \Omega_\beta^{(\mathrm{C1},0)}[\psi]$ for the choice

$$\frac{\partial\psi(Y_0,\beta_0)}{\partial y} = -\frac{\partial\psi_0(Y_0,\beta_0)}{\partial y}\psi_1(Y_0,\beta_0) + \frac{\partial\psi_1(Y_0,\beta_0)}{\partial y}\psi_0(Y_0,\beta_0). \tag{58}$$

In other words, for any choice of ψ_0 and ψ_1 in C2, there is an equally good choice of ψ for C1.

In the case of estimation of θ, it is also not to be expected that C2 could improve on C1 in the sense discussed above, as C1 is already comparable to likelihood to first order in ε. For estimating γ, however, the result is quite disappointing. Since C2 involves transition information where C1 does not, one could have hoped that it would give better efficiency.

Is there improvement to higher order, at least? For this, we use the results in Theorem 2. The discussion applies to both $\beta = \theta$ and γ. Before stating results, note (for comparison with the C1 case) that $D_\beta^{(0)}$ and $S_{\beta,0}^{(0)}$ are the same in the C1 and C2 cases when one makes the identification (58). Thus, noting the form of $T_{\beta,0}^{(\mathrm{C1},0)}$

$$\Omega_\beta^{(\mathrm{C1},1)} = \frac{\mathrm{Var}[\varDelta_0]}{\mathrm{E}[\varDelta_0]}\frac{S_{\beta,0}^{(0)}}{(D_{\beta,0}^{(0)})^2}. \tag{59}$$

Recall that above, we have shown that

$$\Omega_\beta^{(\mathrm{C2},1)} = \mathrm{E}[\varDelta_0]\tilde{\Omega}_\beta^{(1)} + \frac{\mathrm{Var}[\varDelta_0]}{\mathrm{E}[\varDelta_0]}\frac{S_{\beta,0}^{(0)}}{(D_{\beta,0}^{(0)})^2}. \tag{60}$$

This sets the stage for:

Theorem 4. *Let β denote either θ or γ. For optimal choice* (51) *of ψ, and if ψ_0 and ψ_1 satisfy* (58),

$$\tilde{\Omega}_\beta^{(1)} \geqslant 0. \tag{61}$$

Proof of Theorem 4. Using the notation from the previous subsection, since $\Omega_\theta^{(\mathrm{C2},0)} = \Omega_\theta^{(\mathrm{IOML},0)}$, and since $\Omega_\theta^{(\mathrm{C2})} \geqslant \Omega_\theta^{(\mathrm{IOML})}$, it follows that $\Omega_\theta^{(\mathrm{C2},1)} \geqslant \Omega_\theta^{(\mathrm{IOML},1)}$. Since, for the optimal choice of ψ, $V = S_{\beta,0}^{(0)}/(D_{\beta,0}^{(0)})^2$, the inequality becomes

$$\mathrm{E}[\varDelta_0]\tilde{\Omega}_\beta^{(1)} + \frac{\mathrm{Var}[\varDelta_0]}{\mathrm{E}[\varDelta_0]}V \geqslant \frac{\mathrm{E}[\mathrm{Var}[\varDelta_0|\chi_1^2\varDelta_0]]}{\mathrm{E}[\varDelta_0]}V. \tag{62}$$

This must hold for any distribution of $\varDelta_0$ so long as $\mathrm{E}[\varDelta_0]>0$, $\mathrm{E}[\varDelta_0^2]<+\infty$, and $\mathrm{Var}[\varDelta_0]>0$. Having said that, one can then take a limit of a sequence of distributions of $\varDelta_0$ so that $\mathrm{Var}[\varDelta_0]=0$, while the two other conditions remain. This proves the result. $\square$

This would seem to suggest that if ψ is chosen optimally, one cannot to this order improve on the C1 estimator by using a C2 estimator. There are a couple of caveats: the improvement may occur to higher order, and we have not investigated this. We have no result on whether C2 can improve on C1 for a non-optimal ψ, but with ψ_0 and ψ_1 satisfying (58). We also do not know whether $\tilde{\Omega}_\beta^{(1)}>0$ is a possibility.

Since C1 is a special case of C2 (choose $\psi_0 = 1$ and $\psi_1 = \psi$), one can obviously make $\tilde{\Omega}_{\beta}^{(1)} = 0$ with the correct choice of ψ_0 and ψ_1.

4.4. The effect of overidentification

In the discussion of efficiency, we have so far only discussed the estimation of a scalar parameter with a scalar moment condition. This raises the question of whether efficiency properties can be improved by using several moment conditions.

To discuss this question, recall that the general GMM setup in Section 2 is as follows. We consider estimators for a d-dimension β using a vector of r moment conditions $h(y_1, y_0, \delta, \beta, \varepsilon)$, $r \geqslant d$. We form the sample average as in Eq. (5), and obtain $\hat{\beta}$ by minimizing the quadratic form

$$Q_T(\beta) \equiv m_T(\beta)' W_T m_T(\beta),$$

where W_T is an $r \times r$ positive definite weight matrix assumed to converge in probability to a positive definite limit W_β. Let

$$D_\beta \equiv \mathrm{E}_{\Delta, Y_1, Y_0}[\dot{h}(Y_1, Y_0, \Delta, \bar{\beta}, \varepsilon)], \quad S_{\beta,j} \equiv \mathrm{E}_{\Delta, Y_1, Y_0}[h(Y_{1+j}, Y_j, \Delta, \bar{\beta}, \varepsilon) h(Y_1, Y_0, \Delta, \bar{\beta}, \varepsilon)']$$

and $S_\beta \equiv \sum_{j=-\infty}^{+\infty} S_{\beta,j}$. If the system is exactly identified, $r = d$, the choice of W_T is irrelevant and the problem amounts to setting $m_T(\beta)$ to 0. If not, the optimal choice is to set W_T to be any consistent estimator of S_β^{-1}. $\sqrt{T}(\hat{\beta} - \bar{\beta})$ converges in law to $\mathrm{N}(0, \Omega_\beta)$, with

$$\Omega_\beta^{-1} = (\mathrm{E}[\Delta])^{-1} D'_\beta S_\beta^{-1} D_\beta. \tag{63}$$

When $r > d$ and the system is overidentified, the estimator has the same asymptotic variance (63) as obtained if we replaced the $r \times 1$ vector h with the $d \times 1$ vector $H = D'_\beta W_\beta h$ and were back in the exactly identified case. Note that H is a vector of linear combinations of the original moment conditions h. Therefore, in the scalar case ($d = 1$), adding a second moment condition $\tilde{h}$ to an existing such condition h has the following effect. The asymptotic variance behaves as if inference were carried out using an optimal combination $ah + \tilde{a}\tilde{h}$.

This has the following consequences. In the (C1) case, if $h = B_{\beta_0}\psi(y_0, \beta)$ and $\tilde{h} = B_{\beta_0}\tilde{\psi}(y_0, \beta)$, an optimal combination is still of the form of a (C1) estimator. Hence, overidentification cannot improve the asymptotic variance if the test function is already optimal. Obviously, overidentification is helpful if one wishes the data to help find the optimal test function, but that does not alter our efficiency result. Obviously, the same conclusion holds if one wishes to consider more than two test functions.

In the (C2) case, we can only assert that to first order (in ε), overidentification does not improve efficiency. There may be improvement, however, to higher order. The situation is as follows.

For the purposes of computing asymptotic variance, overidentification increases the set of possible optimal estimators (by taking linear combinations of (C2) type estimators). It is therefore a priori quite possible that efficiency improves. To see that this is not the case to first order, however, let (ψ_1, ψ_2) and $(\tilde{\psi}_1, \tilde{\psi}_2)$ be two pairs of test functions, and form (C2) moment conditions h_{C2} and $\tilde{h}_{\mathrm{C2}}$ on the basis of these as in Eq. (29). In obvious extension of the notation in Theorem 2, it is easy to see from the proof of this theorem that

$$\tilde{S}^{(-1)}(ah + \tilde{a}\tilde{h}) = \mathrm{E}_{Y_0}[\sigma^2(a\psi + \tilde{a}\tilde{\psi})^2],$$

where ψ is given by (58) and similarly for $\tilde{\psi}$. Hence, the asymptotic variance for the overidentified estimator in the (C2) case is the same as for the overidentified estimator in the (C1) case, with test functions ψ and $\tilde{\psi}$. Since overidentification does not improve the (C1) estimator, our claim follows.

5. Example: The Ornstein–Uhlenbeck process

We now apply the inference strategies of the previous section to a specific example, the stationary ($\theta > 0$) Ornstein–Uhlenbeck process

$$\mathrm{d}X_t = -\theta X_t \, \mathrm{d}t + \sigma \, \mathrm{d}W_t, \tag{64}$$

where $\gamma = \sigma^2$ and specialize the expressions resulting from the general theorems that precede. We also compare how the different estimation methods fare relative to MLE. 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 $\mathrm{e}^{-\delta\theta}y_0$ and variance $(1-\mathrm{e}^{-2\delta\theta})\gamma/2\theta$. The stationary density $\pi(y_0, \beta)$ is also Gaussian with mean 0 and variance $\sigma^2/(2\theta)$.

For this model, we have from Table III in Aït-Sahalia and Mykland (2003)

$$\Omega_\theta^{(\mathrm{FIML})} = 2\theta_0 + \varepsilon^2\left(\frac{2\theta_0^3\mathrm{E}[\varDelta_0^3]}{3\mathrm{E}[\varDelta_0]}\right) + O(\varepsilon^3), \tag{65}$$

$$\Omega_\theta^{(\mathrm{IOML})} = 2\theta_0 + \varepsilon\left(\frac{2\theta_0^2\mathrm{E}[\mathrm{Var}[\varDelta_0|\chi^2\varDelta_0]]}{\mathrm{E}[\varDelta_0]}\right) + O(\varepsilon^2) \tag{66}$$

and

$$\Omega_\gamma^{(\mathrm{FIML})} = \varepsilon(2\sigma_0^4\mathrm{E}[\varDelta_0]), \tag{67}$$

$$\Omega_\gamma^{(\mathrm{IOML})} = \varepsilon\left(\frac{4\sigma_0^4\mathrm{E}[\varDelta_0]}{2-\mathrm{E}[\mathrm{Var}[\chi^2|\chi^2\varDelta_0]]}\right) + O(\varepsilon^2). \tag{68}$$

The C1 estimation method involves an element of choice, namely the selection of the test function ψ. We presently give the expressions that follow from applying Theorem 1 to the Ornstein–Uhlenbeck process with ψ chosen to be proportional to $\partial \log \pi/\partial\beta$, as advocated in (51). In this case, π from (3) is the normal density with mean 0 and variance $\kappa^2 = \sigma^2/2\theta$, so one gets

$$\psi(y, \beta) = \frac{(y^2-\kappa^2)}{2\kappa^4}\frac{\partial\kappa^2}{\partial\beta}, \tag{69}$$

where β is either θ or σ^2. One can estimate θ for given σ^2, or vice versa. Note that $A_\beta \cdot \psi(y,\beta) = -2\theta\psi(y,\beta)$.

For the estimation of θ, the quantities from Theorem 1 are as follows:

$$\begin{aligned}
D_\theta &= \varepsilon\left(\frac{1}{\theta_0}\right),\\
S_{\theta,0} &= \varepsilon\left(\frac{2}{\mathrm{E}[\varDelta_0]\theta_0}\right) - 2\varepsilon^2\left(\frac{\mathrm{E}[\varDelta_0^2]}{\mathrm{E}[\varDelta_0]^2}+1\right) + O(\varepsilon^3),\\
T_\theta &= 4\varepsilon^2\left(\frac{\mathrm{E}[\varDelta_0^2]}{\mathrm{E}[\varDelta_0]^2}\right) + O(\varepsilon^3),\\
\Omega_\theta &= 2\theta_0 + \varepsilon\left(\frac{2\theta_0^2\mathrm{Var}[\varDelta_0]}{\mathrm{E}[\varDelta_0]}\right) + O(\varepsilon^2),
\end{aligned} \tag{70}$$

while for the estimation of σ^2, one obtains

$$\begin{aligned}
D_{\sigma^2} &= \varepsilon\frac{\theta}{\sigma_0^4},\\
S_{\sigma^2,0} &= \varepsilon\frac{\theta_0}{\sigma_0^4}\left(\frac{2}{\mathrm{E}[\varDelta_0]}\right) - 2\varepsilon^2\frac{\theta_0^2}{\sigma_0^4}\left(\frac{\mathrm{E}[\varDelta_0^2]}{\mathrm{E}[\varDelta_0]^2}+1\right) + O(\varepsilon^3),\\
T_{\sigma^2} &= 4\varepsilon^2\left(\frac{\theta_0^2\,\mathrm{E}[\varDelta_0^2]}{\sigma_0^4\mathrm{E}[\varDelta_0]^2}\right) + O(\varepsilon^3),\\
\Omega_{\sigma^2} &= 2\frac{\sigma_0^4}{\theta_0} + \varepsilon\left(\frac{2\sigma_0^4\mathrm{Var}[\varDelta_0]}{\mathrm{E}[\varDelta_0]}\right) + O(\varepsilon^2).
\end{aligned} \tag{71}$$

It is noteworthy that this is the only case where Ω_{σ^2} is of order $O(1)$ in ε as opposed to order $O(\varepsilon)$. While this follows from applying the general Theorem 1, a simple direct demonstration of this in the Ornstein–Uhlenbeck case is as follows. Note that $A_\beta \cdot \psi(y,\beta)$ for estimating θ is $f(y,\kappa^2)/\kappa^2$, where $f(y,\kappa^2) = y^2-\kappa^2$, while $A_\beta \cdot \psi(y,\beta)$ for estimating σ^2 is $-f(y,\kappa^2)/2\kappa^4$. Hence, if one sets $\hat{\kappa}^2 = N_T^{-1}\sum_i Y_i$, and if one lets $\hat{\theta}$ denote the estimator of θ for σ^2 known, and similarly define $\hat{\sigma}^2$, one gets $\hat{\theta}=\sigma^2/2\hat{\kappa}^2$ and $\hat{\sigma}^2=2\theta\hat{\kappa}^2$. It follows that

$$\sqrt{T}(\hat{\sigma}^2-\sigma^2)=2\theta\sqrt{T}(\hat{\kappa}^2-\kappa^2)=-\frac{\sigma^2}{\theta}\sqrt{T}(\hat{\theta}-\theta)+\mathrm{o_p}(1), \tag{72}$$

whence $\Omega_{\sigma^2}=(\sigma^4/\theta^2)\Omega_\theta$. Since Ω_θ is $O(1)$ in ε, then so is Ω_{σ^2}. This first-order efficiency loss is a natural consequence of the absence of conditioning information in the C1 method.

In the case of the C2 estimator, suppose that one can write $\psi_0(y,\beta_0)$ and $\psi_1(y,\beta_0)$ as a series

$$\psi_0(y,\beta_0)=\sum_{i\geqslant 0}a_i y^i,\quad \psi_1(y,\beta_0)=\sum_{i\geqslant 0}b_i y^i. \tag{73}$$

Under the optimality constraint (58), with

$$\psi(y,\beta)=\frac{y^2-\kappa^2}{2\kappa^4}\frac{\partial\kappa^2}{\partial\beta}, \tag{74}$$

we obtain that

$$\begin{aligned}\frac{y}{\kappa^4}\frac{\partial\kappa^2}{\partial\beta}&=\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\\&=-\frac{\partial\psi_0(Y_0,\beta_0)}{\partial y}\psi_1(Y_0,\beta_0)+\frac{\partial\psi_1(Y_0,\beta_0)}{\partial y}\psi_0(Y_0,\beta_0)\\&=-\sum_{i,j\geqslant 0}(i+1)a_{i+1}b_j y^{i+j}+\sum_{i,j\geqslant 0}(j+1)a_i b_{j+1}y^{i+j}\\&=\sum_{n\geqslant 0}y^n\sum_{i+j=n}[-(i+1)a_{i+1}b_j+(j+1)a_i b_{j+1}].\end{aligned} \tag{75}$$

It follows that for ψ_0 and ψ_1 to be first-order optimal, one needs

$$\sum_{i+j=n}[-(i+1)a_{i+1}b_j+(j+1)a_i b_{j+1}]=0 \tag{76}$$

for all $n\neq 0$. (The restriction for $n=1$ is irrelevant, since inference is unaltered by multiplying ψ by a constant.)

6. Conclusions and extensions

One can extend the theory to cover more general continuous-time Markov processes, such as jump-diffusions. In that case, the standard infinitesimal generator of the process applied to a smooth f takes the form

$$J_{\beta_0}\cdot f=A_{\beta_0}\cdot f+\int\{f(y_1+z,y_0,\delta,\beta,\varepsilon)-f(y_1,y_0,\delta,\beta,\varepsilon)\}v(\mathrm{d}z,y_0),$$

where A_{β_0}, defined in (8), is the contribution coming from the diffusive part of the stochastic differential equation and $v(\mathrm{d}z,y_0)$ is the Lévy jump measure specifying the number of jumps of size in $(z,z+\mathrm{d}z)$ per unit of time (see e.g., Protter, 1992). In that case, our generalized infinitesimal generator becomes

$$\Gamma_{\beta_0}\cdot f\equiv\Delta_0 J_{\beta_0}\cdot f+\frac{\partial f}{\partial\varepsilon}+\frac{\partial f}{\partial\beta}\frac{\partial\beta}{\partial\varepsilon}$$

that is, the same expression as (10) except that A_{β_0} is replaced by J_{β_0}. Of course, the asymptotic variance expressions we derived above, hence the efficiency comparisons, are dependent upon the nature of the generator of the process.

Another extension concerns the generation of the sampling intervals. For example, if the Δ_i's are random and i.i.d., then $\mathrm{E}[\Delta]$ has the usual meaning, but even if this is not the case, by $\mathrm{E}[\Delta]$ we mean the limit (in probability, or just the limit if the Δ_i's are non-random) of $\sum_{i=1}^n\Delta_i/n$ as n tends to infinity. This permits the inclusion of the random non-i.i.d. and the non-random (but possibly irregularly spaced) cases for the Δ_i's. At the cost of further complications, the theory can be extended to allow for dependence in the sampling intervals, whereby Δ_n is drawn conditionally on (Y_{n-1}, Δ_{n-1}).

Acknowledgments

Financial support from the NSF under Grants SES-0350772 and DMS-0532370 (Aït-Sahalia) and SES-0631605 and DMS-0604758 (Mykland) is gratefully acknowledged. We are also very grateful to the Editor, an Associate Editor and two anonymous referees for comments and suggestions.

Appendix A

This appendix contains the proofs of Theorems 1 and 3. We start with a summary of the form of D_β, $S_{\beta,0}$ and T_β for a generic h.

A.1. Summary of results for generic moment functions h

In the following, we specialize the general expansions for D_β, $S_{\beta,0}$ and $T_\beta = S_\beta - S_{\beta,0}$ from Aït-Sahalia and Mykland (2004) in the minimal case that is needed for our application here to HS, namely a single moment function h at a time, with no singularity ($H = 0$). Our expansions have the generic form

$$D_\beta = D_\beta^{(0)} + \varepsilon D_\beta^{(1)} + \varepsilon^2 D_\beta^{(2)} + O(\varepsilon^3), \tag{A.1}$$

$$S_{\beta,0} = S_{\beta,0}^{(0)} + \varepsilon S_{\beta,0}^{(1)} + \varepsilon^2 S_{\beta,0}^{(2)} + O(\varepsilon^3), \tag{A.2}$$

$$T_\beta = \varepsilon^{-1} T_\beta^{(-1)} + T_\beta^{(0)} + \varepsilon T_\beta^{(1)} + \varepsilon^2 T_\beta^{(2)} + O(\varepsilon^3). \tag{A.3}$$

To identify the terms in the above, we apply Lemma 1 in Aït-Sahalia and Mykland (2004) to a single moment function h and get for D_β and $S_{\beta,0}$:

$$D_\beta = \mathrm{E}_{Y_0}[\dot{h}] + \varepsilon \mathrm{E}_{\Delta,Y_0}[\Gamma_{\beta_0} \cdot \dot{h}] + \frac{\varepsilon^2}{2} \mathrm{E}_{\Delta,Y_0}[\Gamma_{\beta_0}^2 \cdot \dot{h}] + O(\varepsilon^3), \tag{A.4}$$

$$S_{\beta,0} = \mathrm{E}_{Y_0}[h^2] + \varepsilon \mathrm{E}_{\Delta,Y_0}[\Gamma_{\beta_0} \cdot (h^2)] + \frac{\varepsilon^2}{2} \mathrm{E}_{\Delta,Y_0}[\Gamma_{\beta_0}^2 \cdot (h^2)] + O(\varepsilon^3). \tag{A.5}$$

As for T_β, the simplest situation arises when h is a martingale,

$$\mathrm{E}_{\Delta,Y_1}[h(Y_1, Y_0, \Delta, \beta_0, \varepsilon) | Y_0] = 0. \tag{A.6}$$

When (A.6) is satisfied, $S_{\beta,j} = 0$ for all $j \neq 0$, and so $T_\beta = 0$. But this is not the case in general for the HS moment functions, for which

$$\begin{aligned} \mathrm{E}_{\Delta,Y_1}[h(Y_1, Y_0, \Delta, \beta_0, \varepsilon) | Y_0] &\equiv q(Y_0, \beta_0, \varepsilon) \\ &= q(Y_0, \beta_0, 0) + O(\varepsilon) \end{aligned} \tag{A.7}$$

is non-zero and that conditional expectation is of order $\alpha = 0$ in ε in the more general setup of Aït-Sahalia and Mykland (2004, equation (27), p. 2195). Recall that $\mathrm{E}_{\Delta,Y_1,Y_0}[h(Y_1, Y_0, \Delta, \beta_0, \varepsilon)] = 0$ hence by the law of iterated expectations we have that

$$\mathrm{E}_{Y_0}[q(Y_0, \beta_0, \varepsilon)] = 0 \tag{A.8}$$

and in particular $\mathrm{E}_{Y_0}[q(Y_0, \beta_0, 0)] = 0$ with $q(Y_0, \beta_0, 0) = h(Y_0, Y_0, 0, \beta_0, 0)$.

While the index α ($= 0$ here) and the function q play a crucial role in determining the order in ε of the matrix T_β, the function r will play an important role in the determination of its coefficients. We define r as

$$r(y_0, \beta_0, \varepsilon) = -\int_0^\infty U_t \cdot A_{\beta_0} \cdot q(y_0, \beta_0, \varepsilon) \mathrm{E}[\tau_{N(t)+1}] \, \mathrm{d}t, \tag{A.9}$$

where $U_\delta \cdot f(y_0, \delta, \beta, \varepsilon) \equiv \mathrm{E}_{Y_1}[f(Y_1, Y_0, \Delta, \beta, \varepsilon) | Y_0 = y_0, \Delta = \delta]$ is the conditional expectations operator. Lemma 3 of Aït-Sahalia and Mykland (2004) showed that

$$T_\beta = \frac{2}{\mathrm{E}[\Delta_0]}(\varepsilon^{-1}\mathrm{E}_{Y_0}[h \times r] + \mathrm{E}_{\Delta_0, Y_0}[\Gamma_{\beta_0} \cdot (h \times r)] + O(\varepsilon)). \tag{A.10}$$

We therefore need to say more about the function r in order to be able to implement the computations required in (A.10). r satisfies the differential equation

$$\frac{\partial}{\partial y}\left[\frac{\partial r(y, \beta_0, 0)}{\partial y} \frac{1}{s(y; \beta_0)}\right] = -\frac{2q(y, \beta_0, 0)}{\sigma^2(y; \gamma_0)s(y; \beta_0)} \tag{A.11}$$

which can be used to evaluate the terms of the expansion

$$\begin{aligned} r(Y_1, \beta_0, \varepsilon) = r(Y_0, \beta_0, 0) + (Y_1 - Y_0)\frac{\partial r(Y_0, \beta_0, 0)}{\partial y} \\ + \frac{1}{2}(Y_1 - Y_0)^2 \frac{\partial^2 r(Y_0, \beta_0, 0)}{\partial y^2} + \varepsilon \frac{\partial r(Y_0, \beta_0, 0)}{\partial \varepsilon} + \mathrm{o_p}(\varepsilon) \end{aligned} \tag{A.12}$$

for a given h: this is what we will need to do for the HS moment functions. In what follows, we use subscripts "C1" or "C2" when denoting the h and related functions (such as q and r) corresponding to each of the two situations.

Appendix B. Proof of Theorem 1

To calculate T_β for the h_{C1} moment function, we start with a lemma:

Lemma 1. *For any function $\phi(Y_0, \beta_0)$ suitably differentiable in y, such that the expected values below exist, we have*

$$\mathrm{E}_{Y_0}[\phi q_{\mathrm{C1}}] = -\frac{1}{2}\mathrm{E}_{Y_0}\left[\sigma^2 \frac{\partial \psi}{\partial y}\frac{\partial \phi}{\partial y}\right], \tag{B.13}$$

$$\mathrm{E}_{Y_0}\left[\sigma^2 \phi \frac{\partial r_{\mathrm{C1}}}{\partial y}\right] = -\mathrm{E}_{Y_0}\left[\sigma^2 \frac{\partial \psi}{\partial y}\phi\right], \tag{B.14}$$

$$\mathrm{E}_{Y_0}\left[\sigma^4 \phi \frac{\partial^2 r_{\mathrm{C1}}}{\partial y^2}\right] = -\mathrm{E}_{Y_0}\left[\sigma^4 \frac{\partial^2 \psi}{\partial y^2}\phi\right], \tag{B.15}$$

where all the functions are evaluated at $\varepsilon = 0$ and $\beta = \beta_0$.

Proof. Based on the form of q_{C1} given in (18), we have

$$\begin{aligned} \mathrm{E}_{Y_0}[\phi q_{\mathrm{C1}}] &= \mathrm{E}_{Y_0}[\phi \times (B_{\beta_0} \cdot \psi)] \\ &= \mathrm{E}_{Y_0}\left[\left(\mu \frac{\partial \psi}{\partial y} + \frac{\sigma^2}{2}\frac{\partial^2 \psi}{\partial y^2}\right)\phi\right] \\ &= \mathrm{E}_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial}{\partial y}\left(\frac{\partial \psi}{\partial y}\phi\right) + \frac{\sigma^2}{2}\frac{\partial^2 \psi}{\partial y^2}\phi\right] \\ &= -\frac{1}{2}\mathrm{E}_{Y_0}\left[\sigma^2 \frac{\partial \psi}{\partial y}\frac{\partial \phi}{\partial y}\right] \end{aligned}$$

which proves (B.13).

Next, recall the fact that $\pi(y;\beta_0)=c/(s(y,\gamma_0)\sigma^2(y,\gamma_0))$ where c is the integration constant needed to ensure that $\int\pi(y;\beta_0)\,\mathrm{d}y=1$. Using integration by parts, we have

$$\begin{aligned}
\mathrm{E}_{Y_0}\left[\sigma^2\phi\frac{\partial r_{C1}}{\partial y}\right] &= \int\frac{\partial r_{C1}}{\partial y}\sigma^2\phi\pi\,\mathrm{d}y\\
&= c\int\frac{\partial r_{C1}}{\partial y}\frac{1}{s}\phi\,\mathrm{d}y\\
&= -c\int\frac{\partial}{\partial y}\left[\frac{\partial r_{C1}}{\partial y}\frac{1}{s}\right]\left(\int^{y}\phi\,\mathrm{d}z_0\right)\mathrm{d}y\\
&= c\int\frac{2q_{C1}}{\sigma^2 s}\left(\int^{y}\phi\,\mathrm{d}z_0\right)\mathrm{d}y\\
&= 2\int q_{C1}\left(\int^{y}\phi\,\mathrm{d}z_0\right)\pi\,\mathrm{d}y\\
&= 2\mathrm{E}_{Y_0}\left[\left(\int^{Y_0}\phi\,\mathrm{d}z_0\right)q_{C1}\right]\\
&= -\mathrm{E}_{Y_0}\left[\sigma^2\frac{\partial\psi}{\partial y}\phi\right], \qquad \text{(B.16)}
\end{aligned}$$

where in the second-to-last equality, the integration constant in $\int^{Y_0}\phi\,\mathrm{d}z_0$ is irrelevant because $\mathrm{E}_{Y_0}[q_{C1}(Y_0,\beta_0,0)]=0$. The last equality follows from (B.13).

Using again (A.11), that is

$$\frac{\partial}{\partial y}\left[\frac{\partial r_{C1}}{\partial y}\frac{1}{s}\right]=-\frac{2q_{C1}}{\sigma^2 s}$$

we have

$$\frac{\partial^2 r_{C1}}{\partial y^2}\frac{1}{s}=\frac{\partial r_{C1}}{\partial y}\frac{\partial s}{\partial y}\frac{1}{s^2}-\frac{2q_{C1}}{\sigma^2 s}=-\frac{\partial r_{C1}}{\partial y}\frac{1}{s}\frac{2\mu}{\sigma^2}-\frac{2q_{C1}}{\sigma^2 s}$$

since

$$\frac{\partial s}{\partial y}=-\frac{2\mu}{\sigma^2}s$$

hence

$$\frac{\partial^2 r_{C1}}{\partial y^2}=-\frac{\partial r_{C1}}{\partial y}\frac{2\mu}{\sigma^2}-\frac{2q_{C1}}{\sigma^2}.$$

With again $\pi=c/(s\sigma^2)$, it follows that

$$\begin{aligned}
\mathrm{E}_{Y_0}\left[\sigma^4\phi\frac{\partial^2 r_{C1}}{\partial y^2}\right] &= -2\mathrm{E}_{Y_0}\left[\sigma^2\phi\mu\frac{\partial r_{C1}}{\partial y}\right]-2\mathrm{E}_{Y_0}[\sigma^2\phi q_{C1}]\\
&= 2\mathrm{E}_{Y_0}\left[\sigma^2\frac{\partial\psi}{\partial y}\phi\mu\right]+\mathrm{E}_{Y_0}\left[\sigma^2\frac{\partial\psi}{\partial y}\frac{\partial(\sigma^2\phi)}{\partial y}\right]\\
&= 2\mathrm{E}_{Y_0}\left[\sigma^2\frac{\partial\psi}{\partial y}\phi\mu\right]+\mathrm{E}_{Y_0}\left[\left(\sigma^4\frac{\partial\phi}{\partial y}+\sigma^2\frac{\partial\sigma^2}{\partial y}\phi\right)\frac{\partial\psi}{\partial y}\right] \qquad \text{(B.17)}
\end{aligned}$$

by applying (B.13) and (B.14).

But recall now that

$$\mathrm{E}_{Y_0}[\phi\mu]=-\frac{1}{2}\mathrm{E}_{Y_0}\left[\sigma^2\frac{\partial\phi}{\partial y}\right]$$

so that

$$\begin{aligned}
\mathrm{E}_{Y_0}\left[\sigma^4\phi\frac{\partial^2 r_{C1}}{\partial y^2}\right] &= -\mathrm{E}_{Y_0}\left[\sigma^2\frac{\partial}{\partial y}\left(\sigma^2\frac{\partial\psi}{\partial y}\phi\right)\right] + \mathrm{E}_{Y_0}\left[\left(\sigma^4 + \sigma^2\frac{\partial\sigma^2}{\partial y}\right)\frac{\partial\psi\partial\phi}{\partial y\,\partial y}\right] \\
&= -\mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\sigma^2\partial\psi}{\partial y\,\partial y}\phi + \sigma^2\frac{\partial^2\psi}{\partial y^2}\phi + \sigma^2\frac{\partial\psi\partial\phi}{\partial y\,\partial y}\right)\right] + \mathrm{E}_{Y_0}\left[\left(\sigma^4\frac{\partial\phi}{\partial y} + \sigma^2\frac{\partial\sigma^2}{\partial y}\phi\right)\frac{\partial\psi}{\partial y}\right] \\
&= -\mathrm{E}_{Y_0}\left[\sigma^4\frac{\partial^2\psi}{\partial y^2}\phi\right]
\end{aligned}$$

and the lemma is proved. □

Returning to T_β, we have from (A.10)

$$\begin{aligned}
T_\beta &= \frac{2}{\mathrm{E}[\Delta_0]}(\varepsilon^{-1}\mathrm{E}_{Y_0}[(h_{C1}\times r_{C1})] + \mathrm{E}_{\Delta_0,Y_0}[(\Gamma_{\beta_0}\cdot(h_{C1}\times r_{C1}))]) + O(\varepsilon) \\
&\equiv \varepsilon^{-1}T_\beta^{(-1)} + T_\beta^{(0)} + O(\varepsilon).
\end{aligned} \tag{B.18}$$

To compute $T_\beta^{(-1)}$, we therefore need to calculate

$$\begin{aligned}
\mathrm{E}_{Y_0}[(h_{C1}\times r_{C1})] &= \mathrm{E}_{Y_0}[h_{C1}(Y_0, Y_0, 0, \beta_0, 0)\times r_{C1}(Y_0,\beta_0,0)] \\
&= \mathrm{E}_{Y_0}[(B_{\beta_0}\cdot\psi(Y_0,\beta_0))\times r_{C1}(Y_0,\beta_0,0)] \\
&= -\frac{1}{2}\mathrm{E}_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\frac{\partial\psi(Y_0,\beta_0)\partial r_{C1}(Y_0,\beta_0,0)}{\partial y\qquad\qquad\partial y}\right]
\end{aligned}$$

because of (B.13).

Next, we apply (B.14) to get

$$\mathrm{E}_{Y_0}\left[\sigma^2\frac{\partial\psi\partial r_{C1}}{\partial y\ \partial y}\right] = -\mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi}{\partial y}\right)^2\right]$$

and thus

$$\begin{aligned}
T_\beta^{(-1)} &= \frac{2}{\mathrm{E}[\Delta_0]}\mathrm{E}_{Y_0}[(h_{C1}\times r_{C1})] \\
&= \frac{2}{\mathrm{E}[\Delta_0]}\left(-\frac{1}{2}\right)\left(-\mathrm{E}_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\left(\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right)^2\right]\right) \\
&= \frac{1}{\mathrm{E}[\Delta_0]}\mathrm{E}_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\left(\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right)^2\right].
\end{aligned} \tag{B.19}$$

Regarding the next order term in T_β, we have

$$T_\beta^{(0)} = \frac{2}{\mathrm{E}[\Delta_0]}\mathrm{E}_{\Delta_0,Y_0}[(\Gamma_{\beta_0}\cdot(h_{C1}\times r_{C1}))] = \frac{2}{\mathrm{E}[\Delta_0]}\mathrm{E}_{\Delta_0,Y_0}[h_{C1}\times(\Gamma_{\beta_0}\cdot r_{C1})]$$

since by (14) and the independence of h_{C1} on y_1, we have $\Gamma_{\beta_0}\cdot h_{C1} = 0$ and $\partial h_{C1}/\partial y_1 = 0$ so that

$$\begin{aligned}
\Gamma_{\beta_0}\cdot(h_{C1}\times r_{C1}) &= (\Gamma_{\beta_0}\cdot h_{C1})\times r_{C1} + h_{C1}\times(\Gamma_{\beta_0}\cdot r_{C1}) + \Delta_0\sigma_0^2\frac{\partial r_{C1}\partial h_{C1}}{\partial y_1\ \partial y_1} \\
&= h_{C1}\times(\Gamma_{\beta_0}\cdot r_{C1}).
\end{aligned}$$

Using the definition of the operator Γ_{β_0}, we have $\Gamma_{\beta_0}\cdot r_{C1} = \Delta_0(A_{\beta_0}\cdot r_{C1}) + \partial r_{C1}/\partial\varepsilon$ with

$$\frac{\partial r_{C1}}{\partial\varepsilon}(y,\beta_0,0) = \frac{\partial\check{r}_{C1}}{\partial\varepsilon}(y,\beta_0,0) + \frac{1\mathrm{E}[\Delta_0^2]}{2\mathrm{E}[\Delta_0]}q_{C1}(y,\beta_0,0)$$

from equation (29) in Aït-Sahalia and Mykland (2004, p. 2196). But $q_{C1}(y,\beta_0,\varepsilon)\equiv q_{C1}(y,\beta_0,0)$ identically, i.e., q_{C1} does not depend on ε and therefore $\check{r}_{C1}$ does not depend on ε either. Hence $(\partial\check{r}_{C1}/\partial\varepsilon)(y,\beta_0,0)=0$.

Therefore

$$\begin{aligned}\mathrm{E}_{\Delta_0,Y_0}[h_{C1}\times(\Gamma_{\beta_0}\cdot r_{C1})] &= \mathrm{E}_{\Delta_0,Y_0}\left[(B_{\beta_0}\cdot\psi)\times\left(\Delta_0 A_{\beta_0}\cdot r_{C1}+\frac{1}{2}\frac{\mathrm{E}[\Delta_0^2]}{\mathrm{E}[\Delta_0]}q_{C1}\right)\right]\\ &= \mathrm{E}[\Delta_0]\underbrace{\mathrm{E}_{Y_0}[(B_{\beta_0}\cdot\psi)(A_{\beta_0}\cdot r_{C1})]}_{\equiv K_1}+\frac{1}{2}\frac{\mathrm{E}[\Delta_0^2]}{\mathrm{E}[\Delta_0]}\underbrace{\mathrm{E}_{Y_0}[(B_{\beta_0}\cdot\psi)q_{C1}]}_{\equiv K_2}.\end{aligned}\tag{B.20}$$

Consider first the term K_2 in (B.20). Applying (B.13), we have

$$\begin{aligned}K_2 &= \mathrm{E}_{Y_0}[(B_{\beta_0}\cdot\psi)q_{C1}] = \mathrm{E}_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\frac{\partial(B_{\beta_0}\cdot\psi)}{\partial y}\right]\\ &= \mathrm{E}_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\frac{\partial}{\partial y}\left(\mu\frac{\partial\psi}{\partial y}+\frac{\sigma^2}{2}\frac{\partial^2\psi}{\partial y^2}\right)\right]\\ &= \mathrm{E}_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\left(\mu\frac{\partial^2\psi}{\partial y^2}+\frac{\partial\mu}{\partial y}\frac{\partial\psi}{\partial y}+\frac{1}{2}\frac{\partial\sigma^2}{\partial y}\frac{\partial^2\psi}{\partial y^2}+\frac{\sigma^2}{2}\frac{\partial^3\psi}{\partial y^3}\right)\right]\\ &= \mathrm{E}_{Y_0}\left[\frac{\sigma^2}{2}\frac{\partial}{\partial y}\left(\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial y^2}\right)-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\left(\frac{\partial\mu}{\partial y}\frac{\partial\psi}{\partial y}+\frac{1}{2}\frac{\partial\sigma^2}{\partial y}\frac{\partial^2\psi}{\partial y^2}+\frac{\sigma^2}{2}\frac{\partial^3\psi}{\partial y^3}\right)\right]\\ &= \mathrm{E}_{Y_0}\left[\frac{\sigma^2}{4}\frac{\partial\sigma^2}{\partial y}\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial y^2}+\frac{\sigma^4}{4}\left(\left(\frac{\partial^2\psi}{\partial y^2}\right)^2+\frac{\partial\psi}{\partial y}\frac{\partial^3\psi}{\partial y^3}\right)-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\left(\frac{\partial\mu}{\partial y}\frac{\partial\psi}{\partial y}+\frac{1}{2}\frac{\partial\sigma^2}{\partial y}\frac{\partial^2\psi}{\partial y^2}\right)-\frac{\sigma^4}{4}\frac{\partial\psi}{\partial y}\frac{\partial^3\psi}{\partial y^3}\right]\\ &= \mathrm{E}_{Y_0}\left[\frac{\sigma^4}{4}\left(\frac{\partial^2\psi}{\partial y^2}\right)^2-\frac{\sigma^2}{2}\frac{\partial\mu}{\partial y}\left(\frac{\partial\psi}{\partial y}\right)^2\right].\end{aligned}\tag{B.21}$$

Regarding the term K_1 in (B.20), we apply similarly (B.13) to obtain

$$\begin{aligned}K_1 &= \mathrm{E}_{Y_0}\left[(B_{\beta_0}\cdot\psi)(A_{\beta_0}\cdot r_{C1})\right] = \mathrm{E}_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\frac{\partial(A_{\beta_0}\cdot r_{C1})}{\partial y}\right]\\ &= \mathrm{E}_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\frac{\partial}{\partial y}\left(\mu\frac{\partial r_{C1}}{\partial y}+\frac{\sigma^2}{2}\frac{\partial^2 r_{C1}}{\partial y^2}\right)\right]\\ &= \mathrm{E}_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\left(\mu\frac{\partial^2 r_{C1}}{\partial y^2}+\frac{\partial\mu}{\partial y}\frac{\partial r_{C1}}{\partial y}+\frac{1}{2}\frac{\partial\sigma^2}{\partial y}\frac{\partial^2 r_{C1}}{\partial y^2}+\frac{\sigma^2}{2}\frac{\partial^3 r_{C1}}{\partial y^3}\right)\right]\\ &= \mathrm{E}_{Y_0}\left[\frac{\sigma^2}{2}\frac{\partial}{\partial y}\left(\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\frac{\partial^2 r_{C1}}{\partial y^2}\right)-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\left(\frac{\partial\mu}{\partial y}\frac{\partial r_{C1}}{\partial y}+\frac{1}{2}\frac{\partial\sigma^2}{\partial y}\frac{\partial^2 r_{C1}}{\partial y^2}+\frac{\sigma^2}{2}\frac{\partial^3 r_{C1}}{\partial y^3}\right)\right]\\ &= \mathrm{E}_{Y_0}\left[\frac{\sigma^2}{4}\frac{\partial\sigma^2}{\partial y}\frac{\partial\psi}{\partial y}\frac{\partial^2 r_{C1}}{\partial y^2}+\frac{\sigma^4}{4}\left(\frac{\partial^2\psi}{\partial y^2}\frac{\partial^2 r_{C1}}{\partial y^2}+\frac{\partial\psi}{\partial y}\frac{\partial^3 r_{C1}}{\partial y^3}\right)-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\left(\frac{\partial\mu}{\partial y}\frac{\partial r_{C1}}{\partial y}+\frac{1}{2}\frac{\partial\sigma^2}{\partial y}\frac{\partial^2 r_{C1}}{\partial y^2}\right)-\frac{\sigma^4}{4}\frac{\partial\psi}{\partial y}\frac{\partial^3 r_{C1}}{\partial y^3}\right]\\ &= \mathrm{E}_{Y_0}\left[\frac{\sigma^4}{4}\frac{\partial^2\psi}{\partial y^2}\frac{\partial^2 r_{C1}}{\partial y^2}-\frac{\sigma^2}{2}\frac{\partial\mu}{\partial y}\frac{\partial\psi}{\partial y}\frac{\partial r_{C1}}{\partial y}\right].\end{aligned}$$

Next, we apply (B.14) to get

$$\mathrm{E}_{Y_0}\left[\sigma^2\frac{\partial\mu}{\partial y}\frac{\partial\psi}{\partial y}\frac{\partial r_{C1}}{\partial y}\right] = -\mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi}{\partial y}\right)^2\frac{\partial\mu}{\partial y}\right].$$

Then we apply (B.15) to obtain

$$\mathrm{E}_{Y_0}\left[\sigma^4\frac{\partial^2\psi}{\partial y^2}\frac{\partial^2 r_{C1}}{\partial y^2}\right] = -\mathrm{E}_{Y_0}\left[\sigma^4\frac{\partial^2\psi}{\partial y^2}\frac{\partial^2\psi}{\partial y^2}\right].$$

Therefore

$$
\begin{aligned}
K_1 &= \mathrm{E}_{Y_0}[(B_{\beta_0}\cdot\psi)(A_{\beta_0}\cdot r_{\mathrm{C}1})] \\
&= \mathrm{E}_{Y_0}\left[\frac{\sigma^4\partial^2\psi\partial^2 r_{\mathrm{C}1}}{4\,\partial y^2\ \partial y^2} - \frac{\sigma^2\partial\mu\partial\psi\partial r_{\mathrm{C}1}}{2\,\partial y\partial y\ \partial y}\right] \\
&= \frac{1}{2}\mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi}{\partial y}\right)^2\frac{\partial\mu}{\partial y}\right] - \frac{1}{4}\mathrm{E}_{Y_0}\left[\sigma^4\frac{\partial^2\psi\partial^2\psi}{\partial y^2\,\partial y^2}\right].
\end{aligned}
\tag{B.22}
$$

Replacing (B.21) and (B.22) into (B.20), we therefore have

$$
\begin{aligned}
T_\beta^{(0)} &= \frac{2}{\mathrm{E}[\Delta_0]}\{\mathrm{E}_{\Delta_0,Y_0}[h_{\mathrm{C}1}\times(\Gamma_{\beta_0}\cdot r_{\mathrm{C}1})]\} \\
&= 2K_1 + \frac{\mathrm{E}[\Delta_0^2]}{\mathrm{E}[\Delta_0]^2}K_2 \\
&= \mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi}{\partial y}\right)^2\frac{\partial\mu}{\partial y}\right] - \mathrm{E}_{Y_0}\left[\frac{\sigma^4\partial^2\psi\partial^2\psi}{2\,\partial y^2\,\partial y^2}\right] + \frac{\mathrm{E}[\Delta_0^2]}{\mathrm{E}[\Delta_0]^2}\mathrm{E}_{Y_0}\left[\frac{\sigma^4}{4}\left(\frac{\partial^2\psi}{\partial y^2}\right)^2 - \frac{\sigma^2\partial\mu}{2\,\partial y}\left(\frac{\partial\psi}{\partial y}\right)^2\right] \\
&= \frac{\left(\mathrm{E}[\Delta_0^2]-2\mathrm{E}[\Delta_0]^2\right)}{4\mathrm{E}[\Delta_0]^2}\left\{\mathrm{E}_{Y_0}\left[\sigma^4\left(\frac{\partial^2\psi}{\partial y^2}\right)^2\right] - 2\mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi}{\partial y}\right)^2\frac{\partial\mu}{\partial y}\right]\right\}.
\end{aligned}
\tag{B.23}
$$

We then put everything together: (B.19) and (B.23) give the expansion of T_β in (B.18); (19)–(20) for D_β and $S_{\beta,0}$; and from there follows the expansion for Ω_β given in (26).

Appendix C. Proof of Theorem 2

The moment function $h_{\mathrm{C}2}$ depends on (y_1, y_0, β), but not on (δ, ε). As in the $h_{\mathrm{C}1}$ case, the only difference between estimating θ and estimating σ^2 appears in D_β. Recall that $q_{\mathrm{C}2}$ is defined by

$$
\begin{aligned}
\mathrm{E}_{\Delta,Y_1}[h_{\mathrm{C}2}(Y_1,Y_0,\Delta,\beta_0,\varepsilon)|Y_0] &\equiv q_{\mathrm{C}2}(Y_0,\beta_0,\varepsilon) \\
&= q_2(Y_0,\beta_0,0) + \varepsilon\frac{\partial q_2(Y_0,\beta_0,0)}{\partial\varepsilon} + O(\varepsilon^2)
\end{aligned}
$$

and note that, even though $h_{\mathrm{C}2}$ does not depend on ε, $q_{\mathrm{C}2}(y_0,\beta_0,\varepsilon)$ does depend on ε (unlike $q_{\mathrm{C}1}$): the dependence of $h_{\mathrm{C}2}$ on y_1 implies that $q_{\mathrm{C}2}(y_0,\beta_0,\varepsilon)$ does not reduce to $q_{\mathrm{C}2}(y_0,\beta_0,0)$. The specific expressions for $q_{\mathrm{C}2}(y_0,\beta_0,0)$ and $\partial q_{\mathrm{C}2}(Y_0,\beta_0,0)/\partial\varepsilon$ are

$$
\begin{aligned}
q_{\mathrm{C}2}(y_0,\beta_0,0) &= h_{\mathrm{C}2}(y_0,y_0,0,\beta_0,0) \\
&= \{A_{\beta_0}\cdot\psi_1(y_0,\beta_0)\}\times\psi_0(y_0,\beta_0) - \{B_{\beta_0}\cdot\psi_0(y_0,\beta_0)\}\times\psi_1(y_0,\beta_0)
\end{aligned}
\tag{C.24}
$$

and

$$
\begin{aligned}
\frac{\partial q_{\mathrm{C}2}(y_0,\beta_0,0)}{\partial\varepsilon} &= \mathrm{E}[\Delta_0](A_{\beta_0}\cdot h_{\mathrm{C}2})(y_0,y_0,0,\beta_0,0) \\
&= \mathrm{E}[\Delta_0]\{\{A_{\beta_0}^2\cdot\psi_1(y_0,\beta_0)\}\times\psi_0(y_0,\beta_0) - \{B_{\beta_0}\cdot\psi_0(y_0,\beta_0)\}\times\{A_{\beta_0}\cdot\psi_1(y_0,\beta_0)\}\}.
\end{aligned}
\tag{C.25}
$$

Next, we have $D_\beta = D_\beta^{(0)} + D_\beta^{(1)}\varepsilon + O(\varepsilon^2)$ and $S_{\beta,0} = S_{\beta,0}^{(0)} + O(\varepsilon)$ with

$$
D_\beta^{(0)} = \begin{cases}
D_\theta^{(0)} = \mathrm{E}_{Y_0}\left[\dfrac{\partial\mu}{\partial\theta}\left(\dfrac{\partial\psi_1}{\partial y}\psi_0 - \psi_1\dfrac{\partial\psi_0}{\partial y}\right)\right] & \text{when estimating } \theta, \\
D_\gamma^{(0)} = \dfrac{1}{2}\mathrm{E}_{Y_0}\left[\dfrac{\partial\sigma^2}{\partial\gamma}\left(\dfrac{\partial^2\psi_1}{\partial y^2}\psi_0 - \psi_1\dfrac{\partial^2\psi_0}{\partial y^2}\right)\right] & \text{when estimating } \gamma,
\end{cases}
\tag{C.26}
$$

$$D_{\beta}^{(1)} = \begin{cases} D_{\theta}^{(1)} = \dfrac{1}{2}\mathrm{E}[\varDelta_0]\mathrm{E}_{Y_0}\left[\sigma^2\dfrac{\partial\mu}{\partial\theta}\left(\dfrac{\partial^2\psi_0}{\partial y^2}\dfrac{\partial\psi_1}{\partial y} - \dfrac{\partial\psi_0}{\partial y}\dfrac{\partial^2\psi_1}{\partial y^2}\right)\right] & \text{when estimating } \theta, \\ D_{\gamma}^{(1)} = \dfrac{1}{4}\mathrm{E}[\varDelta_0]\mathrm{E}_{Y_0}\left[\sigma^2\dfrac{\partial\sigma^2}{\partial\gamma}\left(\dfrac{\partial^3\psi_0}{\partial y^3}\dfrac{\partial\psi_1}{\partial y} - \dfrac{\partial\psi_0}{\partial y}\dfrac{\partial^3\psi_1}{\partial y^3}\right)\right. & \\ \qquad\left. +\sigma^2\dfrac{\partial^2\sigma^2}{\partial y\partial\gamma}\left(\dfrac{\partial^2\psi_0}{\partial y^2}\dfrac{\partial\psi_1}{\partial y} - \dfrac{\partial\psi_0}{\partial y}\dfrac{\partial^2\psi_1}{\partial y^2}\right)\right] & \text{when estimating } \gamma, \end{cases} \tag{C.27}$$

$$S_{\beta,0}^{(0)} = \mathrm{E}_{Y_0}[(\{A_{\beta_0}\cdot\psi_1(Y_0,\beta_0)\}\times\psi_0(Y_0,\beta_0) - \{B_{\beta_0}\cdot\psi_0(Y_0,\beta_0)\}\times\psi_1(Y_0,\beta_0))^2]. \tag{C.28}$$

As for T_β, we have from (A.10)

$$\begin{aligned} T_\beta &= \frac{2}{\mathrm{E}[\varDelta_0]}(\varepsilon^{-1}\mathrm{E}_{Y_0}[(h_{\mathrm{C2}}\times r_{\mathrm{C2}})] + \mathrm{E}_{\varDelta_0,Y_0}[(\varGamma_{\beta_0}\cdot(h_{\mathrm{C2}}\times r_{\mathrm{C2}}))]) + O(\varepsilon) \\ &\equiv \varepsilon^{-1}T_\beta^{(-1)} + T_\beta^{(0)} + O(\varepsilon). \end{aligned} \tag{C.29}$$

The first term is

$$\begin{aligned} T_\beta^{(-1)} &= \frac{2}{\mathrm{E}[\varDelta_0]}\mathrm{E}_{Y_0}[h_{\mathrm{C2}}(Y_0,Y_0,0,\beta_0,0)\times r_{\mathrm{C2}}(Y_0,\beta_0,0)] \\ &= \frac{2}{\mathrm{E}[\varDelta_0]}\mathrm{E}_{Y_0}[q_{\mathrm{C2}}(Y_0,\beta_0,0)\times r_{\mathrm{C2}}(Y_0,\beta_0,0)] \\ &= \frac{1}{\mathrm{E}[\varDelta_0]}\mathrm{E}_{Y_0}\left[\sigma^2\left(\psi_1\frac{\partial\psi_0}{\partial y} - \psi_0\frac{\partial\psi_1}{\partial y}\right)\frac{\partial r_{\mathrm{C2}}}{\partial y}\right] \\ &= \frac{1}{\mathrm{E}[\varDelta_0]}\mathrm{E}_{Y_0}\left[\sigma^2\left(\psi_1\frac{\partial\psi_0}{\partial y} - \psi_0\frac{\partial\psi_1}{\partial y}\right)^2\right] \end{aligned} \tag{C.30}$$

with the second equality following from (C.24), the third from (C.35) and the last from (C.36).

Regarding the next order term in T_β, we have

$$T_\beta^{(0)} = \frac{2}{\mathrm{E}[\varDelta_0]}\mathrm{E}_{\varDelta_0,Y_0}[\varGamma_{\beta_0}\cdot(h_{\mathrm{C2}}\times r_{\mathrm{C2}})]. \tag{C.31}$$

Since h_{C2} does not depend on ε, and in view of the function $\check{r}$ defined in Aït-Sahalia and Mykland (2004, equation (29), p. 2196), we have

$$\begin{aligned} T_\beta^{(0)} &= \frac{2}{\mathrm{E}[\varDelta_0]}\mathrm{E}_{\varDelta_0,Y_0}[\varGamma_{\beta_0}\cdot(h_{\mathrm{C2}}\times\check{r}_{\mathrm{C2}})] + \frac{\mathrm{E}[\varDelta_0^2]}{\mathrm{E}[\varDelta_0]^2}\mathrm{E}_{\varDelta_0,Y_0}[h_{\mathrm{C2}}q] \\ &= 2\mathrm{E}_{Y_0}[A_{\beta_0}\cdot(h_{\mathrm{C2}}\times\check{r}_{\mathrm{C2}})] + \frac{\mathrm{E}[\varDelta_0^2]}{\mathrm{E}[\varDelta_0]^2}S_{\beta,0}^{(0)} \end{aligned} \tag{C.32}$$

since, by iterated conditional expectations,

$$\begin{aligned} \mathrm{E}_{\varDelta_0,Y_0}[h_{\mathrm{C2}}q] &= \mathrm{E}_{\varDelta_0,Y_0}[h_{\mathrm{C2}}^2] \\ &= S_{\beta,0}^{(0)}. \end{aligned} \tag{C.33}$$

Thus $S_\beta^{(-1)} = T_\beta^{(-1)}$, while

$$\begin{aligned} S_\beta^{(0)} &= S_{\beta,0}^{(0)} + T_{\beta,0}^{(0)} \\ &= 2(\mathrm{E}_{Y_0}[A_{\beta_0}\cdot(h_{\mathrm{C2}}\times\check{r}_{\mathrm{C2}})] + S_{\beta,0}^{(0)}) + \frac{\mathrm{Var}[\varDelta_0]}{\mathrm{E}[\varDelta_0]^2}S_{\beta,0}^{(0)}. \end{aligned} \tag{C.34}$$

We then put together the expansions of D_β, $S_{\beta,0}$ and T_β to obtain the expansion for Ω_β given in (38). The terms of order ε^0 are given in the statement of the theorem, while the terms of order ε^1 are

$$\Omega_\theta^{(1)} = \frac{\mathrm{E}[\Delta_0](D_\theta^{(0)}(S_{\beta,0}^{(0)} + T_\beta^{(0)}) - 2D_\theta^{(1)}T_\beta^{(-1)})}{(D_\theta^{(0)})^3},$$
$$\Omega_{\sigma^2}^{(1)} = \frac{\mathrm{E}[\Delta_0](D_{\sigma^2}^{(0)}(S_{\beta,0}^{(0)} + T_\beta^{(0)}) - 2D_{\sigma^2}^{(1)}T_\beta^{(-1)})}{(D_{\sigma^2}^{(0)})^3}$$

when estimating θ or γ, respectively.

We have used the following.

Lemma 2. *For any function $\phi(y_1, y_0, \beta_0)$ suitably differentiable in y_0, such that the expected values below exist, we have*

$$\mathrm{E}_{Y_0}[\phi q_{C2}] = \frac{1}{2}\mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\left(\frac{\partial\phi}{\partial y_1} + \frac{\partial\phi}{\partial y_0}\right)\right], \tag{C.35}$$

$$\mathrm{E}_{Y_0}\left[\sigma^2\phi\frac{\partial r_{C2}}{\partial y}\right] = \mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\right], \tag{C.36}$$

$$\mathrm{E}_{Y_0}\left[\sigma^4\phi\frac{\partial^2 r_{C2}}{\partial y^2}\right] = \mathrm{E}_{Y_0}\left[\sigma^4\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\right], \tag{C.37}$$

where all the functions are evaluated at $y_1 = y_0$, $\varepsilon = 0$ and $\beta = \beta_0$.

Proof. From the form of q_{C2} given in (C.24), we have

$$\begin{aligned}
\mathrm{E}_{Y_0}[\phi q_{C2}] &= \mathrm{E}_{Y_0}[\phi \times (\{A_{\beta_0}\cdot\psi_1(y_0,\beta_0)\}\times\psi_0(y_0,\beta_0) - \{B_{\beta_0}\cdot\psi_0(y_0,\beta_0)\}\times\psi_1(y_0,\beta_0))] \\
&= \mathrm{E}_{Y_0}\left[\phi\times\left(\left\{\mu\frac{\partial\psi_1}{\partial y} + \frac{\sigma^2}{2}\frac{\partial^2\psi_1}{\partial y^2}\right\}\psi_0 - \left\{\mu\frac{\partial\psi_0}{\partial y} + \frac{\sigma^2}{2}\frac{\partial^2\psi_0}{\partial y^2}\right\}\times\psi_1\right)\right] \\
&= \mathrm{E}_{Y_0}\left[\mu\phi\left\{\frac{\partial\psi_1}{\partial y}\psi_0 - \frac{\partial\psi_0}{\partial y}\psi_1\right\}\psi_0 + \frac{\sigma^2}{2}\phi\left\{\frac{\partial^2\psi_1}{\partial y^2}\psi_0 - \frac{\partial^2\psi_0}{\partial y^2}\psi_1\right\}\right] \\
&= -\mathrm{E}_{Y_0}\left[\frac{\sigma^2}{2}\frac{\mathrm{d}}{\mathrm{d}y_0}\left(\phi\left\{\frac{\partial\psi_1}{\partial y}\psi_0 - \frac{\partial\psi_0}{\partial y}\psi_1\right\}\right)\right] + \mathrm{E}_{Y_0}\left[\frac{\sigma^2}{2}\phi\left\{\frac{\partial^2\psi_1}{\partial y^2}\psi_0 - \frac{\partial^2\psi_0}{\partial y^2}\psi_1\right\}\right] \\
&= -\mathrm{E}_{Y_0}\left[\frac{\sigma^2}{2}\left\{\frac{\partial\psi_1}{\partial y}\psi_0 - \frac{\partial\psi_0}{\partial y}\psi_1\right\}\frac{\mathrm{d}\phi}{\mathrm{d}y_0}\right] - \mathrm{E}_{Y_0}\left[\frac{\sigma^2}{2}\left\{\frac{\partial^2\psi_1}{\partial y^2}\psi_0 - \frac{\partial^2\psi_0}{\partial y^2}\psi_1\right\}\phi\right] \\
&\quad + \mathrm{E}_{Y_0}\left[\frac{\sigma^2}{2}\phi\left\{\frac{\partial^2\psi_1}{\partial y^2}\psi_0 - \frac{\partial^2\psi_0}{\partial y^2}\psi_1\right\}\right] \\
&= \frac{1}{2}\mathrm{E}_{Y_0}\left[\sigma^2\left\{\frac{\partial\psi_1}{\partial y}\psi_0 - \frac{\partial\psi_0}{\partial y}\psi_1\right\}\frac{\mathrm{d}\phi}{\mathrm{d}y_0}\right]
\end{aligned}$$

with the fourth equality following from

$$\mathrm{E}_{Y_0}[\mu(Y_0,\theta_0)f(Y_0)] = -\mathrm{E}_{Y_0}\left[\frac{\sigma^2(Y_0,\gamma_0)}{2}\frac{\mathrm{d}f(Y_0)}{\mathrm{d}y_0}\right]. \tag{C.38}$$

Then we have

$$\frac{\mathrm{d}\phi}{\mathrm{d}y_0} = \frac{\mathrm{d}\phi(y_0,y_0,\beta_0)}{\mathrm{d}y_0} = \frac{\partial\phi}{\partial y_1} + \frac{\partial\phi}{\partial y_0}.$$

Next, as in the development (B.16) in Lemma 1, we have

$$
\begin{aligned}
\mathrm{E}_{Y_0}\left[\sigma^2\phi\frac{\partial r_{C2}}{\partial y}\right] &= \int \frac{\partial r_{C2}}{\partial y}\sigma^2\phi\pi\,\mathrm{d}y \\
&= 2\mathrm{E}_{Y_0}\left[\left(\int^{Y_0}\phi\,\mathrm{d}z_0\right)q_{C2}\right]
\end{aligned}
$$

from which it follows that

$$
\mathrm{E}_{Y_0}\left[\sigma^2\phi\frac{\partial r_{C2}}{\partial y}\right] = \mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\right]
$$

by applying (C.35).

Next, the same development as (B.17) in Lemma 1 now gives

$$
\mathrm{E}_{Y_0}\left[\sigma^4\phi\frac{\partial^2 r_{C2}}{\partial y^2}\right] = -2\mathrm{E}_{Y_0}\left[\sigma^2\phi\mu\frac{\partial r_{C2}}{\partial y}\right] - 2\mathrm{E}_{Y_0}[\sigma^2\phi q_{C2}]
$$

and we apply (C.35) and (C.36) to obtain

$$
\mathrm{E}_{Y_0}\left[\sigma^4\phi\frac{\partial^2 r_{C2}}{\partial y^2}\right] = -2\mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\mu\right] - \mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\frac{\mathrm{d}(\sigma^2\phi)}{\mathrm{d}y_0}\right].
$$

But from (C.38) it follows that

$$
\begin{aligned}
-2\mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\mu\right] &= \mathrm{E}_{Y_0}\left[\sigma^2\frac{\mathrm{d}}{\mathrm{d}y_0}\left(\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\right)\right] \\
&= \mathrm{E}_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\frac{\mathrm{d}(\sigma^2\phi)}{\mathrm{d}y_0}\right] \\
&\quad + \mathrm{E}_{Y_0}\left[\sigma^4\phi\frac{\partial}{\partial y}\left(\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\right)\right]
\end{aligned}
$$

and therefore

$$
\begin{aligned}
\mathrm{E}_{Y_0}\left[\sigma^4\phi\frac{\partial^2 r_{C2}}{\partial y^2}\right] &= \mathrm{E}_{Y_0}\left[\sigma^4\phi\frac{\partial}{\partial y}\left(\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\right)\right] \\
&= \mathrm{E}_{Y_0}\left[\sigma^4\phi\left(\frac{\partial^2\psi_0}{\partial y^2}\psi_1 - \psi_0\frac{\partial^2\psi_1}{\partial y^2}\right)\right]
\end{aligned}
$$

which completes the proof of the lemma. □

References

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9, 385–426.

Aït-Sahalia, Y., 2002. Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. Econometrica 70, 223–262.

Aït-Sahalia, Y., Mykland, P.A., 2003. The effects of random and discrete sampling when estimating continuous-time diffusions. Econometrica 71, 483–549.

Aït-Sahalia, Y., Mykland, P.A., 2004. Estimators of diffusions with randomly spaced discrete observations: a general theory. The Annals of Statistics 32, 2186–2222.

Conley, T.G., Hansen, L.P., Luttmer, E.G.J., Scheinkman, J.A., 1997. Short-term interest rates as subordinated diffusions. Review of Financial Studies 10, 525–577.

Duffie, D., Glynn, P., 2004. Estimation of continuous-time Markov processes sampled at random times. Econometrica 72, 1773–1808.

Hansen, L.P., Scheinkman, J.A., 1995. Back to the future: generating moment implications for continuous-time Markov processes. Econometrica 63, 767–804.

Hansen, L.P., Scheinkman, J.A., Touzi, N., 1998. Identification of scalar diffusions using eigenvectors. Journal of Econometrics 86, 1–32.

Karatzas, I., Shreve, S.E., 1991. Brownian Motion and Stochastic Calculus. Springer, New York.

Kent, J., 1978. Time reversible diffusions. Advanced Applied Probability 10, 819–835.

Kessler, M., Sørensen, M., 1999. Estimating equations based on eigenfunctions for a discretely observed diffusion. Bernoulli 5, 299–314.

Phillips, P.C., Yu, J., 2005. A two-stage realized volatility approach to the estimation of diffusion processes from discrete observations. Technical Report, Singapore Management University.

Protter, P., 1992. Stochastic Integration and Differential Equations: A New Approach. Springer, New York.

Renault, E., Werker, B.J., 2003. Causality effects in return volatility measures with random times. Technical Report, University of North Carolina and Tilburg University.