The University of Chicago
Department of Statistics
TECHNICAL REPORT SERIES

An Analysis of Hansen-Scheinkman Moment Estimators for Discretely and Randomly Sampled Diffusions*

Yacine Ait-Sahalia
Department of Economics
Princeton University and NBER

Per A. Mykland
Department of Statistics
The University of Chicago

TECHNICAL REPORT NO. 557

September 23, 2005

5734 S. University Avenue
Chicago, IL 60637

* Financial support from the NSF under grants SES-0350772 (Ait-Sahalia) and DMS-0204639 (Mykland) is grattefully acknowledged.
. † Princeton, NJ 08544-1021. Phone: (609) 258-4015. E-mail: yacine@princeton.edu.
‡ Chicago, IL 60637-1514. Phone: (773) 702-8044. E-mail: mykland@galton.uchicago.edu.

An Analysis of Hansen-Scheinkman Moment Estimators for Discretely and Randomly Sampled Diffusions*

Yacine Aït-Sahalia
Department of Economics
Princeton University and NBER†

Per A. Mykland
Department of Statistics
The University of Chicago‡

September 23, 2005

Abstract

We derive closed form expansions for the asymptotic distribution of Hansen and Scheinkman (1995) 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.

KEYWORDS: Diffusions; discrete sampling; random sampling; moment conditions; efficiency.
JEL CODES: C22.

*Financial support from the NSF under grants SES-0350772 (Aït-Sahalia) and DMS-0204639 (Mykland) is gratefully acknowledged.

†Princeton, NJ 08544-1021. Phone: (609) 258-4015. E-mail: yacine@princeton.edu.

‡Chicago, IL 60637–1514. Phone: (773) 702-8044. E-mail: mykland@galton.uchicago.edu.

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 page 540 in Conley et al. (1997)). 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 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 the asymptotic properties of estimators based on arbitrary moment conditions. Our method provides explicit expressions for all relevant characteristic of the asymptotic distribution. In regular circumstances, the estimator $\hat{\beta}$ of the parameter vector β_0 converges in probability to some $\bar{\beta}$ and $\sqrt{T}(\hat{\beta} - \bar{\beta})$ 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 and, when applicable the bias $\bar{\beta} - \beta_0$, are generally unknown in closed form. The solution we proposed is to derive Taylor expansions for Ω_β and $\bar{\beta} - \beta_0$ 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 $\Delta \overset{L}{=} \epsilon \Delta_0$, where Δ_0 is possibly random with a given finite distribution. Discrete (but not random) sampling corresponds to the special case where $\text{Var}[\Delta_0] = 0$. Our Taylor expansions are of the form

$$\Omega_\beta = \Omega_\beta^{(0)} + \epsilon\,\Omega_\beta^{(1)} + \epsilon^2\,\Omega_\beta^{(2)} + O\left(\epsilon^3\right). \tag{1}$$

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 Cramer-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 vector of parameters entering the drift function and γ those entering the diffusion function. Recall that 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. Not only are the C1 and C2 estimators fully efficient to first order in ϵ (confirming the result of Conley et al. (1997) in the special case of constant Δ), but they are also very close to being efficient to second order in ϵ. On the other hand, a 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 substantial theory on volatility inference for high frequency data, this is a feasible approach.

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) and Aït-Sahalia and Mykland (2004)). Namely, suppose that we observe the process

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\gamma)dW_t \tag{2}$$

at discrete times in the interval $[0,T]$, and we wish to estimate the d-dimensional parameter vector $\beta' = (\theta',\gamma')$ which lies in an open and bounded set. We let $\mathcal{S} = (\underline{x},\bar{x})$ denote the domain of the diffusion X_t. Define the scale and speed densities of the process

$$s\left(x;\beta\right) \equiv \exp\left\{-2\int^x \left(\mu\left(;y\theta\right)/\sigma^2(y;\gamma)\right)dy\right\} \tag{3}$$

$$m\left(x;\beta\right) \equiv 1/\left(\sigma^2(x;\gamma)s\left(x;\beta\right)\right), \tag{4}$$

the scale and speed measures $S\left(x;\beta\right) \equiv \int^x s\left(w;\beta\right)dw$ and $M\left(x;\beta\right) \equiv \int^x m\left(w;\beta\right)dw$, and the transformation

$$g\left(x;\gamma\right) \equiv \int^x \frac{du}{\sigma(x;\gamma)}. \tag{5}$$

We assume in Assumption 1 below conditions that make this transformation well-defined. By Itô's Lemma, $\tilde{X}_t \equiv g\left(X_t;\gamma\right)$ defined on $\tilde{\mathcal{S}} = \left(g\left(\underline{x};\gamma\right),g\left(\bar{x};\gamma\right)\right)$ satisfies $d\tilde{X}_t = \tilde{\mu}\left(\tilde{X}_t;\beta\right)dt + dW_t$ with

$$\tilde{\mu}\left(x;\beta\right) \equiv \frac{\mu\left(g^{\text{inv}}\left(x;\eta\right);\theta\right)}{\sigma(g^{\text{inv}}\left(x;\eta\right);\eta)} - \frac{1}{2}\frac{\partial\sigma(g^{\text{inv}}\left(x;\eta\right);\eta)}{\partial x}$$

where g^{inv} denotes the reciprocal transformation. We also define the scale and speed densities of $\tilde{X}$, $\tilde{s}$ and $\tilde{m}$, and $\tilde{\lambda}\left(x;\beta\right) \equiv -\left(\tilde{\mu}\left(x;\beta\right)^2 + \partial\tilde{\mu}\left(x;\beta\right)/\partial x\right)/2$.

We will denote by L^2 the Hilbert space of measurable real-valued functions f on $\mathcal{S}$ such that $\|f\|^2 \equiv E\left[f(X_0)^2\right] < \infty$ for all values of β. When f is a function of other variables, in addition to the state variable y_1, we say that $f \in L^2$ if it satisfies the integrability condition for every given value of the other variables.

We call the observations times on the X process $\tau_0 = 0$, τ_1, τ_2, ..., τ_{N_T}, where N_T is the smallest integer such that $\tau_{N_T+1} > T$. We suppose that the sampling intervals $\Delta_n = \tau_n - \tau_{n-1}$ are i.i.d., and independent of the X process and of the parameter β (see Assumption 2 below). In other words, we observe

$$Y_0, \Delta_1, Y_1, \Delta_2, Y_2, ..., \Delta_{N_T-1}, Y_{N_T-1}$$

where $Y_i = X_{\tau_i}$.

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

$$\Delta = \epsilon\Delta_0 \tag{6}$$

where Δ_0 has a given finite distribution and ϵ is deterministic. While we assume that the distribution of the sampling intervals is independent of β, it may well depend upon it own nuisance parameters (such as an unknown arrival rate). 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 $\bar{\Delta}$.

In Aït-Sahalia and Mykland (2004), we considered $r-$dimensional moment conditions $h(y_1, y_0, \delta, \beta, \epsilon)$, which are continuously differentiable in the $d-$dimensional parameter vector β, $r \geq d$. In standard GMM fashion, we form the sample average

$$m_T(\beta) \equiv N_T^{-1} \sum_{n=1}^{N_T-1} h(Y_n, Y_{n-1}, \Delta_n, \beta, \epsilon) \tag{7}$$

and obtain $\hat{\beta}$ by minimizing the quadratic form

$$Q_T(\beta) \equiv m_T(\beta)' \, W_T \, m_T(\beta) \tag{8}$$

where W_T is an $r \times r$ positive definite weight matrix assumed to converge in probability to a positive definite limit W_β. If the system is exactly identified, $r = d$, the choice of W_T is irrelevant and minimizing (8) amounts to setting $m_T(\beta)$ to 0. The function h is known in different strands of the literature either as a "moment function" [see e.g., Hansen (1982)] or an "estimating equation" [see e.g., Godambe (1960) and Heyde (1997).]

Consistency of $\hat{\beta}$ follows from

$$E_{\Delta,Y_1,Y_0}\left[h(Y_1,Y_0,\Delta,\beta_0,\epsilon)\right]=0. \tag{9}$$

where we denote by E_{Δ,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. Some otherwise fairly natural estimating strategies lead to inconsistent estimators. To allow for this, we do not assume that (9) is necessarily satisfied. Rather, we simply assume that the equation $E_{\Delta,Y_1,Y_0}\left[h(Y_1,Y_0,\Delta,\beta,\epsilon)\right]=0$ admits a unique root in β, which we define as $\bar{\beta}=\bar{\beta}(\beta_0,\epsilon)$.

For the estimator to be consistent, it must be that $\bar{\beta}\equiv\beta_0$ but, again, this will not be the case for every estimation method (although it will be for the HS moment conditions). However, in all the cases we consider, and one may argue for *any* reasonable estimation method, the bias will disappear in the limit where $\epsilon\to 0$, i.e., $\bar{\beta}(\beta_0,0)=\beta_0$ (so that there is no bias in the limiting case of continuous sampling) and we have an expansion of the form

$$\bar{\beta}=\bar{\beta}(\beta_0,\epsilon)=\beta_0+b^{(1)}\epsilon+b^{(2)}\epsilon^2+O\left(\epsilon^3\right). \tag{10}$$

For the HS moment conditions, we have simply $b^{(i)}=0$ for all $i\geq 1$ as the estimators are correctly centered.

With N_T/T converging in probability to $(E\left[\Delta\right])^{-1}$, it follows from standard arguments that $\sqrt{T}(\hat{\beta}-\bar{\beta})$ converges in law to $N(0,\Omega_\beta)$, with

$$\Omega_\beta^{-1}=(E\left[\Delta\right])^{-1}D_\beta' S_\beta^{-1}D_\beta. \tag{11}$$

where

$$D_\beta\equiv E_{\Delta,Y_1,Y_0}\left[\dot{h}(Y_1,Y_0,\Delta,\bar{\beta},\epsilon)\right],\ S_{\beta,j}\equiv E_{\Delta,Y_1,Y_0}\left[h(Y_{1+j},Y_j,\Delta,\bar{\beta},\epsilon)h(Y_1,Y_0,\Delta,\bar{\beta},\epsilon)'\right]$$

and $S_\beta\equiv\sum_{j=-\infty}^{+\infty}S_{\beta,j}$. If $r>d$, the weight matrix W_T is assumed to be any consistent estimator of S_β^{-1}; otherwise its choice is irrelevant. A consistent first-step estimator of $\bar{\beta}$, needed to compute the optimal weight matrix, can be obtained by minimizing (8) with $W_T=Id$.

The standard infinitesimal generator A_{β_0} is the operator which returns

$$A_{\beta_0} \cdot f = \frac{\partial f}{\partial \delta} + \mu(y_1, \theta_0)\frac{\partial f}{\partial y_1} + \frac{1}{2}\sigma^2(y_1; \gamma_0)\frac{\partial^2 f}{\partial y_1^2} \tag{12}$$

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

$$\lim_{y_1 \to \underline{x}} \frac{\partial f/\partial y_1}{s(y_1; \beta)} = \lim_{y_1 \to \bar{x}} \frac{\partial f/\partial y_1}{s(y_1; \beta)} = 0 \tag{13}$$

(see Hansen et al. (1998)). We define $\mathcal{D}$ 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 $f \in \mathcal{D}$ as follows:

$$\Gamma_{\beta_0} \cdot f \equiv \Delta_0 \, A_{\beta_0} \cdot f + \frac{\partial f}{\partial \epsilon} + \frac{\partial f}{\partial \beta}\frac{\partial \beta}{\partial \epsilon} \tag{14}$$

Define $\mathcal{D}^J$ 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 $\mathcal{D}$ and additionally have $J+2$ continuous derivatives in β and ϵ.

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

$$E_{Y_1}\left[f(Y_1, Y_0, \Delta, \bar{\beta}, \epsilon)|Y_0, \Delta = \delta\right] = \sum_{j=0}^{J} \frac{\epsilon^j}{j!}\left(\Gamma_{\beta_0}^j \cdot f\right)(Y_0, Y_0, 0, \beta_0, 0) + O_p\left(\epsilon^{J+1}\right). \tag{15}$$

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 (15) for arbitrary choices of the moment functions h.

As described in Assumption 3 below, the moment functions h selected to conduct inference are of the generic form

$$h(y_1, y_0, \delta, \beta, \epsilon) = \tilde{h}(y_1, y_0, \delta, \beta, \epsilon) + \frac{H(y_1, y_0, \delta, \beta, \epsilon)}{\delta},$$

where $\tilde{h} \in \mathcal{D}^J$ and $H \in \mathcal{D}^{J+1}$. The term H captures the singularity (i.e., powers of $1/\delta$) which can occur in some estimation strategies (this will not be the case for the HS moment conditions, but is necessary for likelihood inference, for instance). In Appendix B, we recall the general results in

Aït-Sahalia and Mykland (2004) regarding the matrices D_β, $S_{\beta,0}$ and $T_\beta = S_\beta - S_{\beta,0}$ as functions of the specification of the vector of moment conditions h.

We will now make these results specific in the special case of the HS class of moment functions, and compare them to other situations.

3. Estimators Based on Hansen-Scheinkman Estimating Equations

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. 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. Unlike the typical use of the method of moments, however, one cannot in general select as moment conditions the "natural" conditional moments of the process since explicit expressions for the conditional mean, variance, skewness, 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.

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. 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.

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.

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

$$h_{C1}(y_1, y_0, \delta, \beta, \epsilon) = h_{C1}(y_0, \beta) \equiv B_\beta \cdot \psi(y_0, \beta) \equiv \mu(y_0, \theta)\frac{\partial \psi}{\partial y_0} + \frac{1}{2}\sigma^2(y_0; \gamma)\frac{\partial^2 \psi}{\partial y_0^2}. \tag{16}$$

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

$$E_{Y_0}\left[B_{\beta_0} \cdot \psi(Y_0, \beta)\right] = 0. \tag{17}$$

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 unbiased (recall (9)): we have $\bar{\beta}(\beta_0, \epsilon) \equiv \beta_0$ identically because h_{C1} evaluated at β_0 has unconditional mean zero.

Equation (17) follows from the fact that X is a stationary process, hence the unconditional expectation of any function of X_t, such as $E_{X_t}[\psi(X_t, \beta)]$, does not depend upon the date t at which it is evaluated: thus $(\partial/\partial t)E_{X_t}[\psi(X_t, \beta)] = 0$, from which the result follows.

In our setup, h_{C1} only depends on (y_0, β), and not on (y_1, δ, ϵ). It follows that

$$\left(\Gamma_{\beta_0} \cdot h_{C1}\right)(y_1, y_0, \delta, \beta, \epsilon) \equiv 0 \tag{18}$$

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

$$\tilde{h}_{C1}(y_1, \beta) \equiv A_\beta \cdot \psi(y_1, \beta) = \mu(y_1, \theta)\frac{\partial \psi}{\partial y_1} + \frac{1}{2}\sigma^2(y_1; \gamma)\frac{\partial^2 \psi}{\partial y_1^2} \tag{19}$$

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 $f(Y_1, \beta) \in \mathcal{D}^J$ is obtained by computing in our method

$$E_{Y_1}\left[f(Y_1,\beta)|\, Y_0, \Delta\right] = \sum_{j=0}^{J} \frac{\epsilon^j}{j!}\left(\Gamma^j_{\beta_0} \cdot f\right)(Y_0, \beta_0) + O_p\left(\epsilon^{J+1}\right).$$

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

$$E_{Y_0}\left[\left(A^j_{\beta_0} \cdot f\right)(Y_0, \beta_0)\right] = 0$$

for all $j \geq 1$ because the expected value of the generator applied to any function is zero – this is indeed (17). That is, the expansion for the unconditional expectation over the law of Y_0 will stop after the leading ($j = 0$) term. Therefore, computing $E_{Y_1}[f(Y_1, \beta)]$ as prescribed by our method, i.e., through the law of iterated expectations in the form $E_{\Delta, Y_0}\left[E_{Y_1}\left[f(Y_1, \beta)|\, Y_0, \Delta\right]\right]$, will produce the same result as writing down directly $E_{Y_0}[f(Y_0, \beta)]$. In other words, using (16) or (19) as moment functions will yield the same results. The form (16) 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

$$q_{C1}(Y_0, \beta_0, \epsilon) = q_{C1}(Y_0, \beta_0, 0) = B_{\beta_0} \cdot \psi(Y_0, \beta_0), \tag{20}$$

and

$$D_\beta = D_\beta^{(0)} = \begin{cases} D_\theta = E_{Y_0}\left[\frac{\partial \mu(Y_0,\theta_0)}{\partial \theta} \frac{\partial \psi(Y_0,\beta_0)}{\partial y}\right] & \text{when estimating } \theta \\ D_\gamma = \frac{1}{2} E_{Y_0}\left[\frac{\partial \sigma^2(Y_0,\gamma_0)}{\partial \gamma} \frac{\partial^2 \psi(Y_0,\beta_0)}{\partial y^2}\right] & \text{when estimating } \gamma \end{cases} \tag{21}$$

$$S_{\beta,0} = S_{\beta,0}^{(0)} = -\frac{1}{2} E_{Y_0}\left[\sigma^2(Y_0, \gamma_0) \frac{\partial \mu(Y_0, \theta_0)}{\partial y} \left(\frac{\partial \psi(Y_0, \beta_0)}{\partial y}\right)^2\right] + \frac{1}{4} E_{Y_0}\left[\sigma^4(Y_0, \gamma_0) \left(\frac{\partial^2 \psi(Y_0, \beta_0)}{\partial y^2}\right)^2\right] \tag{22}$$

The more difficult calculation involves the time series term T_β. As part of the proof of the following theorem, we show that

$$T_\beta = \epsilon^{-1} T_\beta^{(-1)} + T_\beta^{(0)} + O(\epsilon) \tag{23}$$

where

$$\begin{aligned} T_\beta^{(-1)} &= \frac{2}{E[\Delta_0]} E_{Y_0}\left[(h_{C1} \times r_{C1})\right] \\ T_\beta^{(0)} &= \frac{\left(E[\Delta_0^2] - 2E[\Delta_0]^2\right)}{4E[\Delta_0]^2} \left\{ E_{Y_0}\left[\sigma^4 \left(\frac{\partial^2\psi}{\partial y^2}\right)^2\right] - 2E_{Y_0}\left[\sigma^2\left(\frac{\partial\psi}{\partial y}\right)^2 \frac{\partial\mu}{\partial y}\right]\right\}. \end{aligned}$$

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

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

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

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

The specific expressions, which characterize the asymptotic properties of the estimators based on the moment condition h_{C1}, are given by:

Theorem 1. *(Properties of the Estimators based on the C1 Condition) If h_{C1} is used to estimate either θ or γ, the estimator is unbiased and we have*

$$\Omega_\beta = \Omega_\beta^{(0)} + \epsilon\Omega_\beta^{(1)} + O(\epsilon^2) \tag{24}$$

where, when estimating θ,

$$\begin{aligned} \Omega_\theta^{(0)} &= \frac{E_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\left(\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right)^2\right]}{E_{Y_0}\left[\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right]^2} \\ \Omega_\theta^{(1)} &= \frac{Var[\Delta_0]\left(E_{Y_0}\left[\sigma^4(Y_0,\gamma_0)\left(\frac{\partial^2\psi(Y_0,\beta_0)}{\partial y^2}\right)^2\right] - 2E_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\frac{\partial\mu(Y_0,\theta_0)}{\partial y}\left(\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right)^2\right]\right)}{4\,E[\Delta_0]\;E_{Y_0}\left[\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right]^2} \end{aligned} \tag{25}$$

and, when estimating γ,

$$\begin{aligned} \Omega_\gamma^{(0)} &= \frac{4E_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\left(\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right)^2\right]}{E_{Y_0}\left[\frac{\partial\sigma^2(Y_0,\gamma_0)}{\partial\gamma}\frac{\partial^2\psi(Y_0,\beta_0)}{\partial y^2}\right]^2} \\ \Omega_\gamma^{(1)} &= \frac{Var[\Delta_0]\left(E_{Y_0}\left[\sigma^4(Y_0,\gamma_0)\left(\frac{\partial^2\psi(Y_0,\beta_0)}{\partial y^2}\right)^2\right] - 2E_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\frac{\partial\mu(Y_0,\theta_0)}{\partial y}\left(\frac{\partial\psi(Y_0,\beta_0)}{\partial y}\right)^2\right]\right)}{E[\Delta_0]\;E_{Y_0}\left[\frac{\partial\sigma^2(Y_0,\gamma_0)}{\partial\gamma}\frac{\partial^2\psi(Y_0,\beta_0)}{\partial y^2}\right]^2}. \end{aligned} \tag{26}$$

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

$$h_{C2}(y_1, y_0, \delta, \beta, \epsilon) = h_{C2}(y_1, y_0, \beta) = \{A_\beta \cdot \psi_1(y_1, \beta)\} \times \psi_0(y_0, \beta) - \{B_\beta \cdot \psi_0(y_0, \beta)\} \times \psi_1(y_1, \beta). \quad (27)$$

In general, B_β should be replaced by the infinitesimal generator associated with the reverse time process, A^*_β (see). 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 16).

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

$$E_{Y_0,Y_1}\left[\left\{A_{\beta_0} \cdot \psi_1(Y_1, \beta)\right\} \times \psi_0(Y_0, \beta) - \left\{B_{\beta_0} \cdot \psi_0(Y_0, \beta)\right\} \times \psi_1(Y_1, \beta)\right] = 0. \quad (28)$$

Once A_β is evaluated at β_0, this is true for any value of β in ψ, including β_0.

Equation (28) 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 $E_{X_t, X_{t+\delta}}\left[\psi_0(X_t, \beta)\psi_1(X_{t+\delta}, \beta)\right]$, does not depend upon the date t (it can of course depend upon the time lag δ between the two observations): hence

$$\frac{\partial}{\partial t} E_{X_t, X_{t+\delta}}\left[\psi_0(X_t, \beta)\psi_1(X_{t+\delta}, \beta)\right] = 0$$

from which (28) follows.

C2 can alternatively be obtained, as shown in Aït-Sahalia (1996), by combining the Kolmogorov forward and backward equations characterizing the transition function $p\left(y_1 | y_0, \Delta . \beta\right)$ in a way that eliminates the time derivatives of p, which are unobservable with discrete data. Combining the two equations yields the "transition discrepancy":

$$\begin{aligned} \nu\left(y_1 | y_0, \Delta, \beta\right) \equiv & \left\{\frac{1}{2}\frac{\partial^2}{\partial y_1^2}\left(\sigma^2\left(y_1, \gamma\right) p\left(y_1 | y_0, \Delta, \beta\right)\right) - \frac{\partial}{\partial y_1}\left(\mu\left(y_1, \theta\right) p\left(y_1 | y_0, \Delta, \beta\right)\right)\right\} \\ & - \left\{\mu\left(y_0, \theta\right)\frac{\partial}{\partial y_0}\left(p\left(y_1 | y_0, \Delta, \beta\right)\right) + \frac{1}{2}\sigma^2\left(y_0, \gamma\right)\frac{\partial^2}{\partial y_0^2}\left(p\left(y_1 | y_0, \Delta, \beta\right)\right)\right\} \end{aligned}$$

which, for every (x, y) in $\mathcal{S}^2$ and $\Delta > 0$, must satisfy

$$\nu\left(y_1|y_0, \Delta, \beta\right) = 0.$$

It follows that

$$E_{Y_0,Y_1}\left[\nu\left(Y_1|Y_0, \Delta, \beta\right)\psi_1(Y_1, \beta)\psi_0(Y_0, \beta)\right] = 0. \tag{29}$$

for all suitably differentiable (ψ_0, ψ_1). Now, writing the expected value in integral form, then integration by parts, and the fact that

$$\pi\left(y, \beta\right)\mu\left(y, \theta\right) = \frac{1}{2}\frac{\partial}{\partial y}\left(\sigma^2\left(y, \gamma\right)\pi\left(y, \beta\right)\right),$$

it follows that the equality (29) implies the equality (28).

When considering this case, it is worth while to be careful 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 $= \sigma^2$,*

$$\begin{array}{ll} D_\beta^{(0)} = \tilde{D}_\beta^{(0)} & D_\beta^{(1)} = E[\Delta_0]\tilde{D}_\beta^{(1)} \\ S_\beta^{(-1)} = \dfrac{1}{E[\Delta_0]}\tilde{S}_\beta^{(-1)} & S_\beta^{(0)} = \tilde{S}_\beta^{(0)} + \dfrac{Var\left[\Delta_0\right]}{E\left[\Delta_0\right]^2}S_{\beta,0}^{(0)} \end{array} \tag{30}$$

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,

$$D_\beta^{(0)} = \begin{cases} D_\theta^{(0)} = E_{Y_0}\left[\frac{\partial\mu}{\partial\theta}\left(\frac{\partial\psi_1}{\partial y}\psi_0 - \psi_1\frac{\partial\psi_0}{\partial y}\right)\right] & \textit{when estimating } \theta \\ D_\gamma^{(0)} = \frac{1}{2}E_{Y_0}\left[\frac{\partial\sigma^2}{\partial\gamma}\left(\frac{\partial^2\psi_1}{\partial y^2}\psi_0 - \psi_1\frac{\partial^2\psi_0}{\partial y^2}\right)\right] & \textit{when estimating } \sigma^2 \end{cases} \tag{31}$$

$$\tilde{D}_\beta^{(1)} = \begin{cases} \tilde{D}_\theta^{(1)} = \frac{1}{2}E_{Y_0}\left[\sigma^2\frac{\partial\mu}{\partial\theta}\left(\frac{\partial^2\psi_0}{\partial y^2}\frac{\partial\psi_1}{\partial y} - \frac{\partial\psi_0}{\partial y}\frac{\partial^2\psi_1}{\partial y^2}\right)\right] & \textit{when estimating } \theta \\ \tilde{D}_\gamma^{(1)} = \frac{1}{4}E_{Y_0}\left[\sigma^2\frac{\partial\sigma^2}{\partial\gamma}\left(\frac{\partial^3\psi_0}{\partial y^3}\frac{\partial\psi_1}{\partial y} - \frac{\partial\psi_0}{\partial y}\frac{\partial^3\psi_1}{\partial y^3}\right)\right. & \\ \qquad\left. + \sigma^2\frac{\partial^2\sigma^2}{\partial y\partial\gamma}\left(\frac{\partial^2\psi_0}{\partial y^2}\frac{\partial\psi_1}{\partial y} - \frac{\partial\psi_0}{\partial y}\frac{\partial^2\psi_1}{\partial y^2}\right)\right] & \textit{when estimating } \sigma^2 \end{cases} \tag{32}$$

$$S_{\beta,0}^{(0)} = E_{Y_0}\left[\left(\left\{A_{\beta_0}\cdot\psi_1(Y_0, \beta_0)\right\}\times\psi_0(Y_0, \beta_0) - \left\{B_{\beta_0}\cdot\psi_0(Y_0, \beta_0)\right\}\times\psi_1(Y_0, \beta_0)\right)^2\right] \tag{33}$$

$$\tilde{S}^{(-1)} = E_{Y_0}\left[\sigma^2\left(\psi_1\frac{\partial\psi_0}{\partial y} - \psi_0\frac{\partial\psi_1}{\partial y}\right)^2\right] \tag{34}$$

$$\tilde{S}^{(0)} = 2\left(E_{Y_0}\left[A_{\beta_0}\cdot(h_{C2}\times\breve{r}_{C2})\right] + S_{\beta,0}^{(0)}\right) \tag{35}$$

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 σ^2, the estimator is unbiased and we have*

$$\Omega_\beta = \Omega_\beta^{(0)} + \epsilon\Omega_\beta^{(1)} + O(\epsilon^2) \tag{36}$$

where, when estimating θ,

$$\Omega_\theta^{(0)} = \frac{E_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\left(\psi_1(Y_0,\beta_0)\frac{\partial\psi_0(Y_0,\beta_0)}{\partial y} - \psi_0(Y_0,\beta_0)\frac{\partial\psi_1(Y_0,\beta_0)}{\partial y}\right)^2\right]}{E_{Y_0}\left[\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\left(\frac{\partial\psi_1(Y_0,\beta_0)}{\partial y}\psi_0(Y_0,\beta_0) - \psi_1(Y_0,\beta_0)\frac{\partial\psi_0(Y_0,\beta_0)}{\partial y}\right)\right]^2} \tag{37}$$

and, when estimating σ^2,

$$\Omega_{\sigma^2}^{(0)} = \frac{4\ E_{Y_0}\left[\sigma^2(Y_0,\gamma_0)\left(\psi_1(Y_0,\beta_0)\frac{\partial\psi_0(Y_0,\beta_0)}{\partial y} - \psi_0(Y_0,\beta_0)\frac{\partial\psi_1(Y_0,\beta_0)}{\partial y}\right)^2\right]}{E_{Y_0}\left[\left(\frac{\partial^2\psi_1(Y_0,\beta_0)}{\partial y^2}\psi_0(Y_0,\beta_0) - \psi_1(Y_0,\beta_0)\frac{\partial^2\psi_0(Y_0,\beta_0)}{\partial y^2}\right)\right]^2}. \tag{38}$$

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

$$\Omega_\beta^{(1)} = E\left[\Delta_0\right]\tilde{\Omega}_\beta^{(1)} + \frac{Var\left[\Delta_0\right]}{E\left[\Delta_0\right]}\frac{S_{\beta,0}^{(0)}}{(D_{\beta,0}^{(0)})^2} \tag{39}$$

where

$$\tilde{\Omega}_\beta^{(1)} = \frac{\tilde{S}_\beta^{(0)}\tilde{D}_\beta^{(0)} - \tilde{S}_\beta^{(-1)}\tilde{D}_\beta^{(1)}}{(\tilde{D}_\beta^{(0)})^3}. \tag{40}$$

The form given in (39)-(40) will be useful in the next section.

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 three likelihood-based estimators, which make different use of the observations on the state process and the times at which these observations have been recorded. We designed these estimators in such a way that each one of them is subject to a specific subset of the effects we wish to measure. As a result, the differences in their properties allowed us to zero in and isolate these different effects.

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

Both FIML and IOML confront the randomness issue head-on. FIML uses the recorded sampling times, IOML does not, but still recognizes their relevance by integrating them out in the absence of observations on them. Because the data are always discretely sampled, each estimator is subject to the "cost of discreteness," which we define to be the additional variance relative to the variance of an asymptotically efficient estimator based on the full continuous-time sample path. It also represents the error that one would make if one were to use continuous-time asymptotics when the data are in fact discretely sampled. However, FIML is only subject to the cost of discreteness, while IOML is penalized by both the fact that the data are discrete (the continuous-time sample path is not observed) and randomly spaced in time (the sampling intervals are not observed). The additional variance of IOML over that of FIML will therefore be identified as the "cost of randomness," or the cost of not observing the randomly-spaced sampling intervals. But if in fact one had recorded the observations times but chosen not to use them in the empirical estimation phase, then what we call the cost of randomness can be interpreted as the cost of throwing away, or not using, these data.

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

All three estimators rely on maximizing a version of the likelihood function of the observations. Let $p(y_1|y_0, \delta, \beta)$ denote the transition function of the process X. Because of the time homogeneity of the model, the transition function p depends only on δ and not on $(t, t+\delta)$ separately. All three estimators make use of some functional of the density p: namely, $p(Y_n|Y_{n-1}, \Delta_n, \beta)$ for FIML; the expectation $\tilde{p}(Y_n|Y_{n-1}, \beta)$ of $p(Y_n|Y_{n-1}, \Delta_n, \beta)$ over the law of $\Delta_n|Y_{n-1}$ for IOML; and $p(Y_n|Y_{n-1}, \bar{\Delta}, \beta)$ for PFML (i.e., like FIML except that $\bar{\Delta}$ is used in place of the actual Δ_n). In practice, even though most diffusion models do not admit closed-form transition densities, all three estimators can be calculated for any diffusion X using arbitrarily accurate closed-form approximations of the transition function p (see Aït-Sahalia (2002)). We also show that $\tilde{p}$ can be obtained in closed form. While FIML and IOML are always consistent estimators, i.e., $\bar{\beta} = \beta_0$, this is not the case for PFML when the sampling intervals are random. 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

$$\Omega_\theta^{\text{FIML}} = \Omega_\theta^{(\text{FIML},0)} + O\left(\epsilon^2\right) \tag{41}$$

$$\Omega_\theta^{\text{IOML}} = \Omega_\theta^{(\text{IOML},0)} + \epsilon\, \Omega_\theta^{(\text{IOML},1)} + O\left(\epsilon^2\right), \tag{42}$$

where

$$\Omega_\theta^{(\text{FIML},0)} = \Omega_\theta^{(\text{IOML},0)} = \left(E_{Y_0}\left[\sigma^{-2}(Y_0, \gamma_0)\left(\partial\mu(Y_0, \theta_0)/\partial\theta\right)^2\right]\right)^{-1} \tag{43}$$

which is the the leading term in Ω_θ corresponding to efficient estimation of θ with a continuous record of observations.

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

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

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

$$V = \frac{\left(E_{Y_0}\left[\sigma_0^4\left(\frac{\partial^2\mu(Y_0,\beta_0)}{\partial y\partial\theta}\right)^2\right] - 2E_{Y_0}\left[\sigma_0^2\frac{\partial\mu(Y_0,\theta_0)}{\partial y}\left(\frac{\partial\mu(Y_0,\beta_0)}{\partial\theta}\right)^2\right]\right)}{4\ E_{Y_0}\left[\left(\frac{\partial\mu(Y_0,\theta_0)}{\partial\theta}\right)^2\right]^2}. \tag{44}$$

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

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

$$\Omega_\gamma^{\text{FIML}} = \epsilon\ \Omega_\gamma^{(\text{FIML},1)} + O\left(\epsilon^2\right)$$
$$\Omega_\gamma^{\text{IOML}} = \epsilon\,\Omega_\gamma^{(\text{IOML},1)} + O\left(\epsilon^2\right),$$

where

$$\Omega_\gamma^{(\text{FIML},1)} = \Omega_\gamma^{(\text{IOML},1)} = E\left[\Delta_0\right]\left(2E_{Y_0}\left[\left(\partial\sigma(Y_0,\gamma_0)/\partial\gamma\right)^2\sigma(Y_0,\gamma_0)^{-2}\right]\right)^{-1}.$$

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., $\Omega_\gamma^{(1)} = 2\sigma_0^4 E\left[\Delta_0\right]$.

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*

>From 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) (p. 565) proposes to use ψ given by

$$\frac{\partial\psi(y,\beta)}{\partial y} = \frac{\partial}{\partial\beta}\left(\frac{2\mu(y,\theta) - \partial\sigma^2(y,\gamma)/\partial y}{\sigma^2(y,\gamma)}\right). \tag{45}$$

This choice of ψ yields a C1 estimator $A_\beta \cdot \psi$ which is *test function efficient* in the sense of Appendix C of Conley et al. (1997). In the case of estimating θ for γ known this in the same as as saying that (45) minimizes $\Omega_\theta^{(0)}$. Similarly, in the case of estimating γ for θ known, (45) minimizes $\Omega_\gamma^{(0)}$.

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 $\Omega_\theta^{(0)}$ for the full information maximum likelihood (FIML). In other words, to first order in ϵ, $A_\beta \cdot \psi$ is fully efficient. This fact is easily seen by substituting (45) into (25), and comparing to the corresponding expression in Corollary 2 (p.510) of Aït-Sahalia and Mykland (2003). (The latter expression is only given for $\sigma^2 = \gamma^2$, but the extension to general $\sigma^2(y)$ follows from the development in Aït-Sahalia and Mykland (2004).) The efficient first order asymptotic variance has the form

In view of this efficiency property, it is obvious that this choice of ψ also minimizes the expression (25). This is shown, with different expressions, in Appendix C of Conley et al. (1997).

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 (25). Note that the relevant comparison is not with the full information maximum likelihood (FIML), but rather what we in Aït-Sahalia and Mykland (2003) called the "integrated out maximum likelihood" (IOML). This estimator 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 (44). We then have from Theorem 1 that, on the one hand, for the C1 estimator

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

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

$$\frac{\Omega_\theta^{(\mathrm{C1},1)}}{\Omega_\theta^{(\mathrm{IOML},1)}} = \frac{Var[\Delta_0]}{E[Var[\Delta_0|\chi_1^2\Delta_0]]}. \tag{46}$$

As it should from from the Cramér-Rao lower bound, 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 Section 5.4 (pp. 511-514) of Aït-Sahalia and Mykland (2003), where we showed in particular that

$$E\left[Var\left[\Delta_0|\chi_1^2\Delta_0\right]\right] = E[\Delta_0^2] - E\left[\left(\chi_1^2\Delta_0\frac{m_4(\chi_1^2\Delta_0)}{m_2(\chi_1^2\Delta_0)}\right)^2\right]. \tag{47}$$

where χ_1^2 and Δ_0 are independent random variables and

$$m_q(b) = E_Z\left[Z^{-q}d_0\left(\frac{b}{Z^2}\right)\right], \tag{48}$$

where Z is $N(0,1)$ 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,

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

If one wishes to compare to the FIML estimator rather than IOML, it is shown in Aït-Sahalia and Mykland (2003) 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 $\epsilon \to 0$, that is $\Omega_\gamma^{(\mathrm{C1,0})} > 0$, while the asymptotic variance of both the FIML and the IOML is or order $O(\epsilon)$ (see Aït-Sahalia and Mykland (2003)). 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, will 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, 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

$$\frac{\partial\psi(Y_0,\beta_0)}{\partial y} = -\frac{\partial\psi_0(Y_0,\beta_0)}{\partial y}\psi_1(Y_0,\beta_0) + \frac{\partial\psi_1(Y_0,\beta_0)}{\partial y}\psi_0(Y_0,\beta_0) \tag{49}$$

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 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 above. 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 (49). Thus, noting the form of $T_{\beta,0}^{(\mathrm{C1},0)}$

$$\Omega_\beta^{(\mathrm{C1},1)} = \frac{Var\left[\Delta_0\right]}{E\left[\Delta_0\right]}\frac{S_{\beta,0}^{(0)}}{(D_{\beta,0}^{(0)})^2}. \tag{50}$$

Recall that above, we have shown that

$$\Omega_\beta^{(\mathrm{C2},1)} = E\left[\Delta_0\right]\tilde{\Omega}_\beta^{(1)} + \frac{Var\left[\Delta_0\right]}{E\left[\Delta_0\right]}\frac{S_{\beta,0}^{(0)}}{(D_{\beta,0}^{(0)})^2} \tag{51}$$

This sets the stage for:

Theorem 4. *For optimal choice (45) of ψ, and if ψ_0 and ψ_1 satisfy (49),*

$$\tilde{\Omega}_\theta^{(1)} \geq 0 \tag{52}$$

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

$$E\left[\Delta_0\right]\tilde{\Omega}_\beta^{(1)} + \frac{Var\left[\Delta_0\right]}{E\left[\Delta_0\right]}V \geq \frac{E[Var[\Delta_0|\chi_1^2\Delta_0]]}{E\left[\Delta_0\right]}\ V. \tag{53}$$

This must hold for any distribution of Δ_0 so long as $E\left[\Delta_0\right] > 0$, $E\left[\Delta_0^2\right] < +\infty$, and $Var[\Delta_0] > 0$. Having said that, one can then take a limit of a sequence of distributions of Δ_0 so that $Var[\Delta_0] = 0$, while the two other conditions remain. This proves the result.

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 (49). We also don't 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.

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

$$dX_t = -\theta X_t dt + \sigma dW_t \tag{54}$$

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 $e^{-\delta\theta} y_0$ and variance $\left(1 - e^{-2\delta\theta}\right) \gamma/2\theta$. 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)

$$\begin{aligned}
\Omega_{\theta}^{(\mathrm{FIML})} &= 2\theta_0 + \epsilon^2 \left(\frac{2\theta_0^3 E\left[\Delta_0^3\right]}{3E\left[\Delta_0\right]} \right) + O\left(\epsilon^3\right). \\
\Omega_{\theta}^{(\mathrm{IOML})} &= 2\theta_0 + \epsilon \left(\frac{2\theta_0^2 E\left[Var\left[\Delta_0|\chi^2\Delta_0\right]\right]}{E\left[\Delta_0\right]} \right) + O\left(\epsilon^2\right)
\end{aligned}$$

and

$$\begin{aligned}
\Omega_{\gamma}^{(\mathrm{FIML})} &= \epsilon\left(2\sigma_0^4 E\left[\Delta_0\right]\right) \\
\Omega_{\gamma}^{(\mathrm{IOML})} &= \epsilon \left(\frac{4\sigma_0^4 E\left[\Delta_0\right]}{2 - E\left[Var\left[\chi^2|\chi^2\Delta_0\right]\right]} \right) + O\left(\epsilon^2\right).
\end{aligned}$$

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 $\partial \log \pi / \partial \beta$. This is the choice advocated by Conley et al. (1997): see their Sections 3.2-3.3 and Appendix C where they show that this choice is approximately optimal among the restricted class of moment conditions they consider. In this case, π is the normal density with mean 0 and variance $\kappa^2 = \sigma^2/2\theta$, so one gets

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

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:

$$\begin{aligned}
D_\theta &= \epsilon \left(\frac{1}{\theta_0} \right) \\
S_{\theta,0} &= \epsilon \left(\frac{2}{E[\Delta_0]\,\theta_0} \right) - 2\epsilon^2 \left(\frac{E\left[\Delta_0^2\right]}{(E[\Delta_0])^2} + 1 \right) + O\left(\epsilon^3\right) \\
T_\theta &= 4\epsilon^2 \left(\frac{E\left[\Delta_0^2\right]}{(E[\Delta_0])^2} \right) + O\left(\epsilon^3\right) \\
\Omega_\theta &= 2\theta_0 + \epsilon \left(\frac{2\theta_0^2 Var[\Delta_0]}{E[\Delta_0]} \right) + O\left(\epsilon^2\right),
\end{aligned}$$

while for the estimation of σ^2, one obtains

$$\begin{aligned}
D_{\sigma^2} &= \epsilon \frac{\theta}{\sigma_0^4} \\
S_{\sigma^2,0} &= \epsilon \frac{\theta_0}{\sigma_0^4} \left(\frac{2}{E[\Delta_0]} \right) - 2\epsilon^2 \frac{\theta_0^2}{\sigma_0^4} \left(\frac{E\left[\Delta_0^2\right]}{(E[\Delta_0])^2} + 1 \right) + O\left(\epsilon^3\right) \\
T_{\sigma^2} &= 4\epsilon^2 \left(\frac{\theta_0^2}{\sigma_0^4} \frac{E\left[\Delta_0^2\right]}{(E[\Delta_0])^2} \right) + O\left(\epsilon^3\right) \\
\Omega_{\sigma^2} &= 2\frac{\sigma_0^4}{\theta_0} + \epsilon \left(\frac{2\sigma_0^4 Var[\Delta_0]}{E[\Delta_0]} \right) + O\left(\epsilon^2\right)
\end{aligned}$$

It is noteworthy that this is the only case where Ω_{σ^2} is of order $O(1)$ in ϵ as opposed to order $O(\epsilon)$. 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

$$\sqrt{T}(\hat{\sigma}^2 - \sigma^2) = 2\theta\sqrt{T}(\hat{\kappa}^2 - \kappa^2) = -\frac{\sigma^2}{\theta}\sqrt{T}(\hat{\theta} - \theta) + o_p(1), \tag{55}$$

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,

$$\psi_0(y, \beta_0) = \sum_{i \geq 0} a_i y^i, \quad \psi_1(y, \beta_0) = \sum_{i \geq 0} b_i y^i. \tag{56}$$

Under the optimality constraint (49), with

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

we obtain that

$$\begin{aligned} \frac{y}{\kappa^4} \frac{\partial \kappa^2}{\partial \beta} &= \frac{\partial \psi(Y_0, \beta_0)}{\partial y} \\ &= -\frac{\partial \psi_0(Y_0, \beta_0)}{\partial y} \psi_1(Y_0, \beta_0) + \frac{\partial \psi_1(Y_0, \beta_0)}{\partial y} \psi_0(Y_0, \beta_0) \\ &= -\sum_{i,j \geq 0} (i+1) a_{i+1} b_j y^{i+j} + \sum_{i,j \geq 0} (j+1) a_i b_{j+1} y^{i+j} \\ &= \sum_{n \geq 0} y^n \sum_{i+j=n} [-(i+1) a_{i+1} b_j + (j+1) a_i b_{j+1}]. \end{aligned} \tag{57}$$

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

$$\sum_{i+j=n} [-(i+1) a_{i+1} b_j + (j+1) a_i b_{j+1}] = 0 \tag{58}$$

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

$$J_{\beta_0} \cdot f = A_{\beta_0} \cdot f + \int \{f(y_1 + z, y_0, \delta, \beta, \epsilon) - f(y_1, y_0, \delta, \beta, \epsilon)\} \, \nu(dz, y_0)$$

where A_{β_0}, defined in (12), is the contribution coming from the diffusive part of the stochastic differential equation and $\nu(dz, y_0)$ is the Lévy jump measure specifying the number of jumps of size in $(z, z+dz)$ per unit of time (see e.g., Protter (1992).) In that case, our generalized infinitesimal generator becomes

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

that is, the same expression as (14) 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 $E[\Delta]$ has the usual meaning, but even if this is not the case, by $E[\Delta]$ 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 nonrandom (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}).

Appendix

This appendix contains the technical assumptions under which our general results in Aït-Sahalia and Mykland (2004) hold (hence those of this paper), a summary of the asymptotic distribution expansion for general moment conditions, as well as the proofs of Theorems 1 and 3.

A. Technical Assumptions

We make the following primitive assumptions on (μ, σ):

Assumption 1. *For all values of the parameters* (θ, γ) :

1. *Differentiability: The functions* $\mu(x;\theta)$ *and* $\sigma(x;\gamma)$ *are infinitely differentiable in* x.
2. *Non-degeneracy of the Diffusion: If* $\mathcal{S} = (-\infty, +\infty)$, *there exists a constant* c *such that* $\sigma(x;\gamma) > c > 0$ *for all* x *and* γ. *If* $\mathcal{S} = (0, +\infty)$, $\lim_{x\to 0^+} \sigma^2(x;\gamma) = 0$ *is possible but then there exist constants* $\xi_0 > 0$, $\omega > 0$, $\rho \geq 0$ *such that* $\sigma^2(x;\gamma) \geq \omega x^\rho$ *for all* $0 < x \leq \xi_0$ *and* γ. *Whether or not* $\lim_{x\to 0^+} \sigma^2(x;\gamma) = 0$, σ *is non-degenerate in the interior of* $\mathcal{S}$, *that is: for each* $\xi > 0$, *there exists a constant* c_ξ *such that* $\sigma^2(x;\gamma) \geq c_\xi > 0$ *for all* $x \in [\xi, +\infty)$ *and* γ.
3. *Boundary Behavior:* μ, σ^2 *and their derivatives have at most polynomial growth in* x *near the boundaries,* $\lim_{x\to \underline{x}} S(x;\beta) = -\infty$ *and* $\lim_{x\to \bar{x}} S(x;\beta) = +\infty$,
$$\liminf_{x\to \underline{x}} \tilde{\mu}(x;\beta) > 0 \quad \text{and} \quad \limsup_{x\to \bar{x}} \tilde{\mu}(x;\beta) < 0. \tag{A.1}$$
and
$$\limsup_{x\to \underline{x} \text{ or } x\to \bar{x}} \tilde{\lambda}(x;\beta) < +\infty. \tag{A.2}$$
4. *Identification:* $\mu(x;\theta) = \mu(x;\tilde{\theta})$ *for* $\pi-$*almost all* x *in* $\mathcal{S}$ *implies* $\theta = \tilde{\theta}$ *and* $\sigma^2(x;\gamma) = \sigma^2(x;\tilde{\gamma})$ *for* $\pi-$*almost all* x *in* $\mathcal{S}$ *implies* $\gamma = \tilde{\gamma}$.

The assumptions made ensure that the stochastic differential equation (2) admits a weak solution which is unique in probability law, has a regular transition density, exponentially decaying $\rho-$mixing coefficients and is stationary with stationary density

$$\pi(x,\beta) = \frac{m\left(x;\beta\right)}{\int_{\underline{x}}^{\bar{x}} m\left(y;\beta\right)dy} \tag{A.3}$$

provided that the initial value of the process, X_0, has density π.

We assume the following regarding the data generating process for the sampling intervals:

Assumption 2. *The sampling intervals $\Delta_n = \tau_n - \tau_{n-1}$ are independent and identically distributed. Each Δ_n is drawn from a common distribution which is independent of Y_{n-1} and of the parameter β. Also, $E[\Delta_0^2] < +\infty$.*

Finally, we assume the following regularity condition regarding the vector of moment functions h selected to conduct inference:

Assumption 3. *$h(y_1, y_0, \delta, \beta, \epsilon) \in \mathcal{D}^J$ for some $J \geq 3$, and is of the form*

$$h(y_1, y_0, \delta, \beta, \epsilon) = \tilde{h}(y_1, y_0, \delta, \beta, \epsilon) + \frac{H(y_1, y_0, \delta, \beta, \epsilon)}{\delta}, \tag{A.4}$$

where $\tilde{h} \in \mathcal{D}^J$ and $H \in \mathcal{D}^{J+1}$. When the function H is not identically zero, we add the requirements that

$$H(y_0, y_0, 0, \beta_0, 0) = \frac{\partial H(y_1, y_0, 0, \beta_0, 0)}{\partial y_1} = \dot{H}(y_0, y_0, 0, \beta_0, 0) = 0. \tag{A.5}$$

B. Summary of Results for Generic Moment Functions h

In the following, we summarize the general expansions for the matrices D_β, $S_{\beta,0}$ and $T_\beta = S_\beta - S_{\beta,0}$ from Aït-Sahalia and Mykland (2004). These expansions take the form

$$\begin{aligned} D_\beta &= D_\beta^{(0)} + \epsilon D_\beta^{(1)} + \epsilon^2 D_\beta^{(2)} + O(\epsilon^3) \\ S_{\beta,0} &= S_{\beta,0}^{(0)} + \epsilon S_{\beta,0}^{(1)} + \epsilon^2 S_{\beta,0}^{(2)} + O(\epsilon^3) \\ T_\beta &= \epsilon^{-1} T_\beta^{(-1)} + T_\beta^{(0)} + \epsilon T_\beta^{(1)} + \epsilon^2 T_\beta^{(2)} + O(\epsilon^3) \end{aligned}$$

from which an expansion for the AVAR matrix of the form

$$\Omega_\beta = \Omega_\beta^{(0)} + \epsilon\,\Omega_\beta^{(1)} + \epsilon^2 \Omega_\beta^{(2)} + O\left(\epsilon^3\right). \tag{B.6}$$

B.1. The D_β and $S_{\beta,0}$ Matrices

Lemma 1 in Aït-Sahalia and Mykland (2004) states that for a vector of moment functions $h = (h_1, ..., h_r)'$ where $h = \tilde{h} + \Delta^{-1}H$ as described in (A.4)-(A.5) and $\tilde{h} \in \mathcal{D}^3$, $H \in \mathcal{D}^4$, we have:

In the case where H is identically zero,

$$D_\beta = E_{Y_0}\left[\dot{h}\right] + \epsilon E_{\Delta,Y_0}\left[(\Gamma_{\beta_0} \cdot \dot{h})\right] + \frac{\epsilon^2}{2} E_{\Delta,Y_0}\left[(\Gamma^2_{\beta_0} \cdot \dot{h})\right] + O(\epsilon^3) \tag{B.7}$$

and, with the notation $h \times h'(y_1, y_0, \delta, \beta, \epsilon) \equiv h(y_1, y_0, \delta, \beta, \epsilon)\, h(y_1, y_0, \delta, \beta, \epsilon)'$, we have

$$S_{\beta,0} = E_{Y_0}\left[(h \times h')\right] + \epsilon E_{\Delta,Y_0}\left[(\Gamma_{\beta_0} \cdot (h \times h'))\right] + \frac{\epsilon^2}{2} E_{\Delta,Y_0}\left[(\Gamma^2_{\beta_0} \cdot (h \times h'))\right] + O(\epsilon^3). \tag{B.8}$$

In the case where H is not zero, (B.7) and (B.8) should be evaluated at $\tilde{h}$ rather than h, yielding $D^{\tilde{h}}_\beta$ and $S^{\tilde{h}}_{\beta,0}$ respectively. Then $D_\beta = D^{\tilde{h}}_\beta + D^H_\beta$ and $S_{\beta,0} = S^{\tilde{h}}_{\beta,0} + S^H_{\beta,0}$ where

$$D^H_\beta = E_{\Delta,Y_0}\left[\Delta_0^{-1}(\Gamma_{\beta_0} \cdot \dot{H})\right] + \frac{\epsilon}{2} E_{\Delta,Y_0}\left[\Delta_0^{-1}(\Gamma^2_{\beta_0} \cdot \dot{H})\right] + O(\epsilon^2), \tag{B.9}$$

$$\begin{aligned} S^H_{\beta,0} = {} & E_{\Delta,Y_0}\left[\Delta_0^{-1}(\Gamma_{\beta_0} \cdot (\tilde{h} \times H'))\right] + \frac{\epsilon}{2} E_{\Delta,Y_0}\left[\Delta_0^{-1}(\Gamma^2_{\beta_0} \cdot (\tilde{h} \times H'))\right] \\ & + E_{\Delta,Y_0}\left[\Delta_0^{-1}(\Gamma_{\beta_0} \cdot (H \times \tilde{h}'))\right] + \frac{\epsilon}{2} E_{\Delta,Y_0}\left[\Delta_0^{-1}(\Gamma^2_{\beta_0} \cdot (H \times \tilde{h}'))\right] \\ & + \frac{1}{2} E_{\Delta,Y_0}\left[\Delta_0^{-2}(\Gamma^2_{\beta_0} \cdot (H \times H'))\right] + \frac{\epsilon}{6} E_{\Delta,Y_0}\left[\Delta_0^{-2}(\Gamma^3_{\beta_0} \cdot (H \times H'))\right] + O(\epsilon^2). \end{aligned} \tag{B.10}$$

B.2. The T_β Matrix

The simplest case arises when the moment function is a martingale,

$$E_{\Delta,Y_1}\left[h(Y_1, Y_0, \Delta, \beta_0, \epsilon)|Y_0\right] = 0. \tag{B.11}$$

When (B.11) is satisfied, $S_{\beta,j} = 0$ for all $j \neq 0$, and so $T_\beta = 0$. Denote by h_i the i^{th} element of the vector of moment functions h, and define q_i and α_i by

$$\begin{aligned} E_{\Delta,Y_1}\left[h_i(Y_1, Y_0, \Delta, \bar{\beta}, \epsilon)|Y_0\right] &\equiv \epsilon^{\alpha_i} q_i(Y_0, \beta_0, \epsilon) \\ &= \epsilon^{\alpha_i} q_i(Y_0, \beta_0, 0) + \epsilon^{\alpha_i+1} \frac{\partial q_i(Y_0, \beta_0, 0)}{\partial \epsilon} + O(\epsilon^{\alpha_i+2}) \end{aligned} \tag{B.12}$$

where α_i is an integer greater than or equal to zero for each moment function h_i. α_i is an index of the order at which the moment component h_i deviates from a martingale (note that in a vector h not all components h_i need to have the same index α_i). A martingale moment function corresponds to the limiting case where $\alpha_i = +\infty$, $q_i(Y_0, \beta_0, \epsilon)$ is identically zero, and $S_\beta = S_{\beta,0}$. When the moment functions are not martingales, we show in that the difference $T_\beta \equiv S_\beta - S_{\beta,0}$ is a matrix whose element (i, j) has a leading term of order $O\left(\epsilon^{\min(\alpha_i,\alpha_j)}\right)$ in ϵ that depends on q_i and q_j.

Note that $E_{\Delta,Y_1,Y_0}\left[h_i(Y_1, Y_0, \Delta, \bar{\beta}, \epsilon)\right] = 0$ by definition of $\bar{\beta}$, hence by the law of iterated expectations we have that

$$E_{Y_0}\left[q_i(Y_0, \beta_0, \epsilon)\right] = 0. \tag{B.13}$$

and in particular $E_{Y_0}\left[q_i(Y_0, \beta_0, 0)\right] = 0$. Equation (B.12) is obtained as a consequence of

$$\begin{aligned} E_{\Delta,Y_1}\left[h_i(Y_1, Y_0, \Delta, \bar{\beta}, \epsilon)|Y_0\right] &= E_\Delta\left[E_{Y_1}\left[h_i(Y_1, Y_0, \Delta, \bar{\beta}, \epsilon)|Y_0, \Delta\right]\right] \\ &= \sum_{j=0}^{J} \frac{\epsilon^j}{j!}\left\{E_{\Delta_0}\left[\left(\Gamma_{\beta_0}^{j} \cdot \tilde{h}_i\right)\right] + \frac{1}{(j+1)} E_{\Delta_0}\left[\Delta_0^{-1}\left(\Gamma_{\beta_0}^{j+1} \cdot H_i\right)\right]\right\} + O\left(\epsilon^{J+1}\right) \\ &= \epsilon^{\alpha_i} q_i(Y_0, \beta_0, 0) + O(\epsilon^{\alpha_i+1}), \end{aligned} \tag{B.14}$$

if we let α_i denote an index j at which the sum in the right-hand-side of (B.14) is non-zero. α_i is an index of the order at which the moment component h_i deviates from a martingale (note that in a vector h not all components h_i need to have the same index α_i).

For instance, we have $\alpha_i = 0$ with

$$q_i(Y_0, \beta_0, 0) = \tilde{h}_i(Y_0, Y_0, 0, \beta_0, 0) + E_{\Delta_0}\left[\Delta_0^{-1}\left(\Gamma_{\beta_0} \cdot H_i\right)(Y_0, Y_0, 0, \beta_0, 0)|Y_0\right] \tag{B.15}$$

if the right-hand-side of (B.15) is non-zero, or $\alpha_i = 1$ with

$$q_i(Y_0, \beta_0, 0) \equiv E_{\Delta_0}\left[\Gamma_{\beta_0} \cdot \tilde{h}_i(Y_0, Y_0, 0, \beta_0, 0)|Y_0\right] + \frac{1}{2} E_{\Delta_0}\left[\Delta_0^{-1}\left(\Gamma_{\beta_0}^2 \cdot H_i\right)(Y_0, Y_0, 0, \beta_0, 0)|Y_0\right] \tag{B.16}$$

if the right-hand-side of (B.16) is non-zero, or $\alpha_i = 2$ and

$$q_i(Y_0, \beta_0, 0) \equiv \frac{1}{2} E_{\Delta_0}\left[\Gamma_{\beta_0}^2 \cdot \tilde{h}_i(Y_0, Y_0, 0, \beta_0, 0)|Y_0\right] + \frac{1}{6} E_{\Delta_0}\left[\Delta_0^{-1}\left(\Gamma_{\beta_0}^2 \cdot H_i\right)(Y_0, Y_0, 0, \beta_0, 0)|Y_0\right] \tag{B.17}$$

is that term is non-zero, and so on.

A martingale moment function corresponds to the limiting case where $\alpha_i = +\infty$, $q_i(Y_0, \beta_0, \epsilon)$ is identically zero, and $S_\beta = S_{\beta,0}$. When the moment functions are not martingales, the difference $T_\beta \equiv S_\beta - S_{\beta,0}$ is a matrix whose element (i, j) has a leading term of order $O\left(\epsilon^{\min(\alpha_i,\alpha_j)}\right)$ in ϵ that depends on q_i and q_j. While the index α_i and the function q_i play a crucial role in determining the order in ϵ of the matrix T_β, the function r_i will play an important role in the determination of its coefficients. We define r_i as

$$r_i(y_0, \beta_0, \epsilon) = -\int_0^\infty U_t \cdot A_{\beta_0} \cdot q_i(y_0, \beta_0, \epsilon) E[\tau_{N(t)+1}] dt. \tag{B.18}$$

where $U_\delta \cdot f(y_0, \delta, \beta, \epsilon) \equiv E_{Y_1}[f(Y_1, Y_0, \Delta, \beta, \epsilon) | Y_0 = y_0, \Delta = \delta]$ is the conditional expectations operator.

We then showed that

$$r_i(Y_0, \beta_0, \epsilon) = \breve{r}_i(Y_0, \beta_0, \epsilon) + \frac{1}{2}\epsilon \frac{E[\Delta_0^2]}{E[\Delta_0]} q_i(Y_0, \beta_0, 0) + o_p(\epsilon), \tag{B.19}$$

where the function $\breve{r}_i(y, \beta_0, \epsilon)$ is the solution of the differential equation

$$\frac{\partial}{\partial y}[\frac{\partial \breve{r}_i(y, \beta_0, \epsilon)}{\partial y} \frac{1}{s(y; \beta_0)}] = -\frac{2\ q_i(y, \beta_0, \epsilon)}{\sigma^2(y; \gamma_0) s(y; \beta_0)} \tag{B.20}$$

with the side condition that $E_{Y_0}[\breve{r}_i(Y_0, \beta_0, \epsilon)] = 0$.

As $\epsilon \to 0$, we have $r_i(y, \beta_0, 0) = \breve{r}_i(y, \beta_0, 0)$ and

$$\frac{\partial}{\partial y}[\frac{\partial r_i(y, \beta_0, 0)}{\partial y} \frac{1}{s(y; \beta_0)}] = -\frac{2\ q_i(y, \beta_0, 0)}{\sigma^2(y; \gamma_0) s(y; \beta_0)} \tag{B.21}$$

With

$$\frac{\partial r_i}{\partial \epsilon}(y, \beta_0, 0) = \frac{\partial \breve{r}_i}{\partial \epsilon}(y, \beta_0, 0) + \frac{1}{2} \frac{E[\Delta_0^2]}{E[\Delta_0]} q_i(y, \beta_0, 0) \tag{B.22}$$

and

$$\frac{\partial^k r_i(Y_0, \beta_0, 0)}{\partial y^k} \equiv \frac{\partial^k \breve{r}_i(Y_0, \beta_0, 0)}{\partial y^k} \tag{B.23}$$

for $k = 1, 2$, we have

$$\begin{aligned} r_i(Y_1, \beta_0, \epsilon) &= r_i(Y_0, \beta_0, 0) + (Y_1 - Y_0) \frac{\partial r_i(Y_0, \beta_0, 0)}{\partial y} \\ &\quad + \frac{1}{2}(Y_1 - Y_0)^2 \frac{\partial^2 r_i(Y_0, \beta_0, 0)}{\partial y^2} + \epsilon \frac{\partial r_i(Y_0, \beta_0, 0)}{\partial \epsilon} + o_p(\epsilon). \end{aligned} \tag{B.24}$$

The form (B.21) is convenient to compute the derivatives $\partial^k r_i / \partial y^k$ in applications, such as the one in this paper. If $\sigma^2 = \gamma$ constant, dividing (B.21) by σ_0^2 yields an equivalent form in terms of the stationary density π :

$$\frac{\partial}{\partial y}[\frac{\partial r_i(y, \beta_0, 0)}{\partial y} \pi(y; \beta_0)] = -\frac{2}{\sigma_0^2} q_i(y, \beta_0, 0) \pi(y; \beta_0). \tag{B.25}$$

Lemma 3 of Aït-Sahalia and Mykland (2004) showed that:

If H is zero, the (i,j) term of the time series matrix $T_\beta = S_\beta - S_{\beta,0}$ is $[T_\beta]_{(i,j)}$, given by:

$$[T_\beta]_{(i,j)} = \frac{1}{E[\Delta_0]}\Bigg(\epsilon^{\alpha_j - 1} E_{Y_0}[(h_i \times r_j)] + \epsilon^{\alpha_j} E_{\Delta_0,Y_0}\left[\Gamma_{\beta_0}\cdot(h_i \times r_j)\right] + \frac{\epsilon^{\alpha_j+1}}{2} E_{\Delta_0,Y_0}\left[\Gamma^2_{\beta_0}\cdot(h_i \times r_j)\right]$$
$$+ \epsilon^{\alpha_i - 1} E_{Y_0}[(h_j \times r_i)] + \epsilon^{\alpha_i} E_{\Delta_0,Y_0}\left[(\Gamma_{\beta_0}\cdot(h_j \times r_i))\right] + \frac{\epsilon^{\alpha_i+1}}{2} E_{\Delta_0,Y_0}\left[(\Gamma^2_{\beta_0}\cdot(h_j \times r_i))\right]\Bigg)$$
$$+ O(\epsilon^{\min(\alpha_i,\alpha_j)+2}). \tag{B.26}$$

If H is non-zero, then (B.26) should be evaluated at $\tilde{h}$ rather than h, yielding $T^{\tilde{h}}_\beta$. And $T_\beta = T^{\tilde{h}}_\beta + T^H_\beta$ where

$$\left[T^H_\beta\right]_{(i,j)} = \frac{1}{E[\Delta_0]}\Bigg(\epsilon^{\alpha_j - 1} E_{\Delta,Y_0}\left[\Delta_0^{-1}(\Gamma_{\beta_0}\cdot H_i) \times r_j\right] + \frac{\epsilon^{\alpha_j}}{2} E_{\Delta_0,Y_0}\left[\Delta_0^{-1}\left(\Gamma^2_{\beta_0}\cdot(H_i \times r_j)\right)\right]$$
$$+\epsilon^{\alpha_i - 1} E_{\Delta_0,Y_0}\left[\Delta_0^{-1}(\Gamma_{\beta_0}\cdot H_j) \times r_i\right] + \frac{\epsilon^{\alpha_i}}{2} E_{\Delta_0,Y_0}\left[\Delta_0^{-1}\left(\Gamma^2_{\beta_0}\cdot(H_j \times r_i)\right)\right]\Bigg)$$
$$+ O(\epsilon^{\min(\alpha_i,\alpha_j)+1}). \tag{B.27}$$

C. Proof of Theorem 1

To calculate T_β, 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*

$$E_{Y_0}[\phi q_{C1}] = -\frac{1}{2} E_{Y_0}\left[\sigma^2 \frac{\partial\psi}{\partial y}\frac{\partial\phi}{\partial y}\right] \tag{C.28}$$
$$E_{Y_0}\left[\sigma^2 \phi \frac{\partial r_{C1}}{\partial y}\right] = -E_{Y_0}\left[\sigma^2 \frac{\partial\psi}{\partial y}\phi\right] \tag{C.29}$$
$$E_{Y_0}\left[\sigma^4 \phi \frac{\partial^2 r_{C1}}{\partial y^2}\right] = -E_{Y_0}\left[\sigma^4 \frac{\partial^2\psi}{\partial y^2}\phi\right]. \tag{C.30}$$

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

Proof. From the form of q_{C1} given in (20), we have

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

which proves (C.28).

Next, recall (B.25) and, from (A.3), the fact that $\pi(y;\beta_0) = c/\left(s(y,\gamma_0)\sigma^2(y,\gamma_0)\right)$ where c is the integration constant needed to ensure that $\int \pi(y;\beta_0)dy = 1$. Using integration by parts, we have

$$\begin{aligned}
E_{Y_0}\left[\sigma^2 \phi \frac{\partial r_{C1}}{\partial y}\right] &= \int \frac{\partial r_{C1}}{\partial y}\sigma^2 \phi \pi dy \\
&= c\int \frac{\partial r_{C1}}{\partial y}\frac{1}{s}\phi dy \\
&= -c\int \frac{\partial}{\partial y}[\frac{\partial r_{C1}}{\partial y}\frac{1}{s}]\left(\int^y \phi dz_0\right)dy \\
&= c\int \frac{2q_{C1}}{\sigma^2 s}\left(\int^y \phi dz_0\right)dy \qquad \text{(C.31)} \\
&= 2\int q_{C1}\left(\int^y \phi dz_0\right)\pi dy \\
&= 2E_{Y_0}\left[\left(\int^{Y_0}\phi dz_0\right)q_{C1}\right] \\
&= -E_{Y_0}\left[\sigma^2 \frac{\partial \psi}{\partial y}\phi\right]
\end{aligned}$$

where in the second-to-last equality, the integration constant in $\int^{Y_0}\phi dz_0$ is irrelevant because $E_{Y_0}[q_{C1}(Y_0,\beta_0,0)] = 0$. The last equality follows from (C.28).

Using again (B.21), that is

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

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

$$\begin{aligned} E_{Y_0}\left[\sigma^4\phi\frac{\partial^2 r_{C1}}{\partial y^2}\right] &= -2E_{Y_0}\left[\sigma^2\phi\mu\frac{\partial r_{C1}}{\partial y}\right] - 2E_{Y_0}\left[\sigma^2\phi q_{C1}\right] \\ &= 2E_{Y_0}\left[\sigma^2\frac{\partial\psi}{\partial y}\phi\mu\right] + E_{Y_0}\left[\sigma^2\frac{\partial\psi}{\partial y}\frac{\partial\left(\sigma^2\phi\right)}{\partial y}\right] \\ &= 2E_{Y_0}\left[\sigma^2\frac{\partial\psi}{\partial y}\phi\mu\right] + E_{Y_0}\left[\left(\sigma^4\frac{\partial\phi}{\partial y} + \sigma^2\frac{\partial\sigma^2}{\partial y}\phi\right)\frac{\partial\psi}{\partial y}\right] \end{aligned} \tag{C.32}$$

by applying (C.28) and (C.29).

But recall now that

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

so that

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

and the lemma is proved. □

Returning to T_β, we have from (B.26):

$$\begin{aligned} T_\beta &= \frac{2}{E\left[\Delta_0\right]}\left(\epsilon^{-1}E_{Y_0}\left[(h_{C1}\times r_{C1})\right] + E_{\Delta_0,Y_0}\left[\left(\Gamma_{\beta_0}\cdot(h_{C1}\times r_{C1})\right)\right]\right) + O(\epsilon) \\ &\equiv \epsilon^{-1}T_\beta^{(-1)} + T_\beta^{(0)} + O(\epsilon) \end{aligned} \tag{C.33}$$

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

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

because of (C.28).

Next, we apply (C.29) to get

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

and thus

$$\begin{aligned}
T_\beta^{(-1)} &= \frac{2}{E\left[\Delta_0\right]} E_{Y_0}\left[\left(h_{C1} \times r_{C1}\right)\right] \\
&= \frac{2}{E\left[\Delta_0\right]}\left(-\frac{1}{2}\right)\left(-E_{Y_0}\left[\sigma^2(Y_0, \gamma_0)\left(\frac{\partial \psi(Y_0, \beta_0)}{\partial y}\right)^2\right]\right) \\
&= \frac{1}{E\left[\Delta_0\right]} E_{Y_0}\left[\sigma^2(Y_0, \gamma_0)\left(\frac{\partial \psi(Y_0, \beta_0)}{\partial y}\right)^2\right]. \qquad \text{(C.34)}
\end{aligned}$$

Regarding the next order term in T_β, we have

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

since by (18) 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

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

Using the definition of the operator Γ_{β_0}, we have $\Gamma_{\beta_0} \cdot r_{C1} = \Delta_0\left(A_{\beta_0} \cdot r_{C1}\right) + \partial r_{C1}/\partial \epsilon$ (since $\partial \beta/\partial \epsilon = 0$ as the estimator is unbiased), with

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

from (B.22). But $q_{C1}(y, \beta_0, \epsilon) \equiv q_{C1}(y, \beta_0, 0)$ identically, i.e., q_{C1} does not depend on ϵ and therefore $\breve{r}_{C1}$ does not depend on ϵ either. Hence $\left(\partial \breve{r}_{C1}/\partial \epsilon\right)(y, \beta_0, 0) = 0$.

Therefore

$$\begin{aligned}
E_{\Delta_0, Y_0}\left[h_{C1} \times \left(\Gamma_{\beta_0} \cdot r_{C1}\right)\right] &= E_{\Delta_0, Y_0}\left[\left(B_{\beta_0} \cdot \psi\right) \times \left(\Delta_0 A_{\beta_0} \cdot r_{C1} + \frac{1}{2}\frac{E[\Delta_0^2]}{E[\Delta_0]} q_{C1}\right)\right] \qquad \text{(C.35)} \\
&= E\left[\Delta_0\right] \underbrace{E_{Y_0}\left[\left(B_{\beta_0} \cdot \psi\right)\left(A_{\beta_0} \cdot r_{C1}\right)\right]}_{\equiv\; K_1} + \frac{1}{2}\frac{E[\Delta_0^2]}{E[\Delta_0]} \underbrace{E_{Y_0}\left[\left(B_{\beta_0} \cdot \psi\right) q_{C1}\right]}_{\equiv\; K_2}
\end{aligned}$$

Consider first the term K_2 in (C.35). Applying (C.28), we have

$$\begin{aligned}
K_2 &= E_{Y_0}\left[(B_{\beta_0}\cdot\psi)q_{C1})\right] = E_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\frac{\partial(B_{\beta_0}\cdot\psi)}{\partial y}\right] \\
&= E_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\frac{\partial}{\partial y}\left(\mu\frac{\partial\psi}{\partial y}+\frac{\sigma^2}{2}\frac{\partial^2\psi}{\partial y^2}\right)\right] \\
&= E_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\left(\mu\frac{\partial^2\psi}{\partial y^2}+\frac{\partial\mu}{\partial y}\frac{\partial\psi}{\partial y}+\frac{1}{2}\frac{\partial\sigma^2}{\partial y}\frac{\partial^2\psi}{\partial y^2}+\frac{\sigma^2}{2}\frac{\partial^3\psi}{\partial y^3}\right)\right] \\
&= E_{Y_0}\left[\frac{\sigma^2}{2}\frac{\partial}{\partial y}\left(\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial y^2}\right)-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\left(\frac{\partial\mu}{\partial y}\frac{\partial\psi}{\partial y}+\frac{1}{2}\frac{\partial\sigma^2}{\partial y}\frac{\partial^2\psi}{\partial y^2}+\frac{\sigma^2}{2}\frac{\partial^3\psi}{\partial y^3}\right)\right] \\
&= E_{Y_0}\left[\frac{\sigma^2}{4}\frac{\partial\sigma^2}{\partial y}\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial y^2}+\frac{\sigma^4}{4}\left(\left(\frac{\partial^2\psi}{\partial y^2}\right)^2+\frac{\partial\psi}{\partial y}\frac{\partial^3\psi}{\partial y^3}\right)\right. \\
&\qquad\left.-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\left(\frac{\partial\mu}{\partial y}\frac{\partial\psi}{\partial y}+\frac{1}{2}\frac{\partial\sigma^2}{\partial y}\frac{\partial^2\psi}{\partial y^2}\right)-\frac{\sigma^4}{4}\frac{\partial\psi}{\partial y}\frac{\partial^3\psi}{\partial y^3}\right] \\
&= E_{Y_0}\left[\frac{\sigma^4}{4}\left(\frac{\partial^2\psi}{\partial y^2}\right)^2-\frac{\sigma^2}{2}\frac{\partial\mu}{\partial y}\left(\frac{\partial\psi}{\partial y}\right)^2\right] \qquad \text{(C.36)}
\end{aligned}$$

Regarding the term K_1 in (C.35), we apply similarly (C.28) to obtain

$$\begin{aligned}
K_1 &= E_{Y_0}\left[(B_{\beta_0}\cdot\psi)(A_{\beta_0}\cdot r_{C1})\right] = E_{Y_0}\left[-\frac{\sigma^2}{2}\frac{\partial\psi}{\partial y}\frac{\partial(A_{\beta_0}\cdot r_{C1})}{\partial y}\right] \\
&= 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] \\
&= 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] \\
&= 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] \\
&= 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)\right. \\
&\qquad\left.-\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] \\
&= 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 (C.29) to get

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

Then we apply (C.30) to obtain

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

Therefore

$$
\begin{aligned}
K_1 &= E_{Y_0}\left[(B_{\beta_0}\cdot\psi)(A_{\beta_0}\cdot r_{C1})\right] \\
&= E_{Y_0}\left[\frac{\sigma^4}{4}\frac{\partial^2\psi}{\partial y^2}\frac{\partial^2 r_{C1}}{\partial y^2} - \frac{\sigma^2}{2}\frac{\partial\mu}{\partial y}\frac{\partial\psi}{\partial y}\frac{\partial r_{C1}}{\partial y}\right] \\
&= \frac{1}{2}E_{Y_0}\left[\sigma^2\left(\frac{\partial\psi}{\partial y}\right)^2\frac{\partial\mu}{\partial y}\right] - \frac{1}{4}E_{Y_0}\left[\sigma^4\frac{\partial^2\psi}{\partial y^2}\frac{\partial^2\psi}{\partial y^2}\right] \qquad \text{(C.37)}
\end{aligned}
$$

Replacing (C.36) and (C.37) into (C.35), we therefore have:

$$
\begin{aligned}
T_\beta^{(0)} &= \frac{2}{E\left[\Delta_0\right]}\left\{E_{\Delta_0,Y_0}\left[h_{C1}\times(\Gamma_{\beta_0}\cdot r_{C1})\right]\right\} \\
&= 2\ K_1 + \frac{E[\Delta_0^2]}{E[\Delta_0]^2}\ K_2 \\
&= E_{Y_0}\left[\sigma^2\left(\frac{\partial\psi}{\partial y}\right)^2\frac{\partial\mu}{\partial y}\right] - E_{Y_0}\left[\frac{\sigma^4}{2}\frac{\partial^2\psi}{\partial y^2}\frac{\partial^2\psi}{\partial y^2}\right] \\
&\quad + \frac{E[\Delta_0^2]}{E[\Delta_0]^2}E_{Y_0}\left[\frac{\sigma^4}{4}\left(\frac{\partial^2\psi}{\partial y^2}\right)^2 - \frac{\sigma^2}{2}\frac{\partial\mu}{\partial y}\left(\frac{\partial\psi}{\partial y}\right)^2\right] \\
&= \frac{\left(E[\Delta_0^2]-2E[\Delta_0]^2\right)}{4E[\Delta_0]^2}\left\{E_{Y_0}\left[\sigma^4\left(\frac{\partial^2\psi}{\partial y^2}\right)^2\right] - 2E_{Y_0}\left[\sigma^2\left(\frac{\partial\psi}{\partial y}\right)^2\frac{\partial\mu}{\partial y}\right]\right\}. \qquad \text{(C.38)}
\end{aligned}
$$

We then put everything together: (C.34) and (C.38) give the expansion of T_β in (C.33); (21)-(22) for D_β and $S_{\beta,0}$; and from there follows the expansion for Ω_β given in (24).

D. Proof of Theorem 2

The moment function h_{C2} depends on (y_1, y_0, β), but not on (δ, ϵ). Also, as above, $\bar{\beta}(\beta_0, \epsilon) \equiv \beta_0$ identically. As in the h_{C1} case, the only difference between estimating θ and estimating σ^2 appears in D_β. Recall that q_{C2} is defined by

$$
\begin{aligned}
E_{\Delta,Y_1}\left[h_{C2}(Y_1,Y_0,\Delta,\bar{\beta},\epsilon)|Y_0\right] &\equiv q_{C2}(Y_0,\beta_0,\epsilon) \\
&= q_2(Y_0,\beta_0,0) + \epsilon\frac{\partial q_2(Y_0,\beta_0,0)}{\partial\epsilon} + O(\epsilon^2)
\end{aligned}
$$

and note that, even though h_{C2} does not depend on ϵ, $q_{C2}(y_0, \beta_0, \epsilon)$ does depend on ϵ (unlike q_{C1}): the dependence of h_{C2} on y_1 implies that $q_{C2}(y_0, \beta_0, \epsilon)$ does not reduce to $q_{C2}(y_0, \beta_0, 0)$. The specific expressions for $q_{C2}(y_0, \beta_0, 0)$ and $\partial q_{C2}(Y_0, \beta_0, 0)/\partial\epsilon$ are

$$\begin{aligned} q_{C2}(y_0, \beta_0, 0) &= h_{C2}(y_0, y_0, 0, \beta_0, 0) \\ &= \left\{A_{\beta_0} \cdot \psi_1(y_0, \beta_0)\right\} \times \psi_0(y_0, \beta_0) - \left\{B_{\beta_0} \cdot \psi_0(y_0, \beta_0)\right\} \times \psi_1(y_0, \beta_0) \end{aligned} \quad \text{(D.39)}$$

and

$$\begin{aligned} \frac{\partial q_{C2}(y_0, \beta_0, 0)}{\partial \epsilon} &= E[\Delta_0] \left(A_{\beta_0} \cdot h_{C2}\right)(y_0, y_0, 0, \beta_0, 0) \\ &= E[\Delta_0] \Big\{\Big\{A^2_{\beta_0} \cdot \psi_1(y_0, \beta_0)\Big\} \times \psi_0(y_0, \beta_0) \\ &\qquad - \left\{B_{\beta_0} \cdot \psi_0(y_0, \beta_0)\right\} \times \left\{A_{\beta_0} \cdot \psi_1(y_0, \beta_0)\right\}\Big). \end{aligned} \quad \text{(D.40)}$$

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

$$D^{(0)}_\beta = \begin{cases} D^{(0)}_\theta = E_{Y_0}\left[\frac{\partial\mu}{\partial\theta}\left(\frac{\partial\psi_1}{\partial y}\psi_0 - \psi_1\frac{\partial\psi_0}{\partial y}\right)\right] & \text{when estimating } \theta \\ D^{(0)}_\gamma = \frac{1}{2}E_{Y_0}\left[\frac{\partial\sigma^2}{\partial\gamma}\left(\frac{\partial^2\psi_1}{\partial y^2}\psi_0 - \psi_1\frac{\partial^2\psi_0}{\partial y^2}\right)\right] & \text{when estimating } \sigma^2 \end{cases} \quad \text{(D.41)}$$

$$D^{(1)}_\beta = \begin{cases} D^{(1)}_\theta = \frac{1}{2}E[\Delta_0]E_{Y_0}\left[\sigma^2\frac{\partial\mu}{\partial\theta}\left(\frac{\partial^2\psi_0}{\partial y^2}\frac{\partial\psi_1}{\partial y} - \frac{\partial\psi_0}{\partial y}\frac{\partial^2\psi_1}{\partial y^2}\right)\right] & \text{when estimating } \theta \\ D^{(1)}_\gamma = \frac{1}{4}E[\Delta_0]E_{Y_0}\left[\sigma^2\frac{\partial\sigma^2}{\partial\gamma}\left(\frac{\partial^3\psi_0}{\partial y^3}\frac{\partial\psi_1}{\partial y} - \frac{\partial\psi_0}{\partial y}\frac{\partial^3\psi_1}{\partial y^3}\right)\right. & \\ \qquad\left. + \sigma^2\frac{\partial^2\sigma^2}{\partial y\partial\gamma}\left(\frac{\partial^2\psi_0}{\partial y^2}\frac{\partial\psi_1}{\partial y} - \frac{\partial\psi_0}{\partial y}\frac{\partial^2\psi_1}{\partial y^2}\right)\right] & \text{when estimating } \sigma^2 \end{cases} \quad \text{(D.42)}$$

$$S^{(0)}_{\beta,0} = E_{Y_0}\left[\left(\left\{A_{\beta_0} \cdot \psi_1(Y_0, \beta_0)\right\} \times \psi_0(Y_0, \beta_0) - \left\{B_{\beta_0} \cdot \psi_0(Y_0, \beta_0)\right\} \times \psi_1(Y_0, \beta_0)\right)^2\right] \quad \text{(D.43)}$$

As for T_β, we have from (B.26):

$$\begin{aligned} T_\beta &= \frac{2}{E\left[\Delta_0\right]}\left(\epsilon^{-1}E_{Y_0}\left[(h_{C2} \times r_{C2})\right] + E_{\Delta_0, Y_0}\left[\left(\Gamma_{\beta_0} \cdot (h_{C2} \times r_{C2})\right)\right]\right) + O(\epsilon) \\ &\equiv \epsilon^{-1}T^{(-1)}_\beta + T^{(0)}_\beta + O(\epsilon) \end{aligned} \quad \text{(D.44)}$$

The first term is

$$\begin{aligned} T^{(-1)}_\beta &= \frac{2}{E\left[\Delta_0\right]}E_{Y_0}\left[h_{C2}(Y_0, Y_0, 0, \beta_0, 0) \times r_{C2}(Y_0, \beta_0, 0))\right] \\ &= \frac{2}{E\left[\Delta_0\right]}E_{Y_0}\left[q_{C2}(Y_0, \beta_0, 0) \times r_{C2}(Y_0, \beta_0, 0)\right] \\ &= \frac{1}{E\left[\Delta_0\right]}E_{Y_0}\left[\sigma^2\left(\psi_1\frac{\partial\psi_0}{\partial y} - \psi_0\frac{\partial\psi_1}{\partial y}\right)\frac{\partial r_{C2}}{\partial y}\right] \\ &= \frac{1}{E\left[\Delta_0\right]}E_{Y_0}\left[\sigma^2\left(\psi_1\frac{\partial\psi_0}{\partial y} - \psi_0\frac{\partial\psi_1}{\partial y}\right)^2\right] \end{aligned} \quad \text{(D.45)}$$

with the second equality following from (D.39), the third from (D.50) and the last from (D.51).

Regarding the next order term in T_β, we have

$$T_\beta^{(0)} = \frac{2}{E\left[\Delta_0\right]} E_{\Delta_0,Y_0}\left[\Gamma_{\beta_0}\cdot\left(h_{C2}\times r_{C2}\right)\right] \tag{D.46}$$

Since h_{C2} does not depend on ϵ, and in view of (B.19),

$$\begin{aligned} T_\beta^{(0)} &= \frac{2}{E\left[\Delta_0\right]} E_{\Delta_0,Y_0}\left[\Gamma_{\beta_0}\cdot\left(h_{C2}\times \breve{r}_{C2}\right)\right] + \frac{E[\Delta_0^2]}{E\left[\Delta_0\right]^2} E_{\Delta_0,Y_0}\left[h_{C2}q\right] \\ &= 2E_{Y_0}\left[A_{\beta_0}\cdot\left(h_{C2}\times \breve{r}_{C2}\right)\right] + \frac{E[\Delta_0^2]}{E\left[\Delta_0\right]^2} S_{\beta,0}^{(0)} \end{aligned} \tag{D.47}$$

since, by iterated conditional expectations,

$$\begin{aligned} E_{\Delta_0,Y_0}\left[h_{C2}q\right] &= E_{\Delta_0,Y_0}\left[h_{C2}^2\right] \\ &= S_{\beta,0}^{(0)} \end{aligned} \tag{D.48}$$

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

$$\begin{aligned} S_\beta^{(0)} &= S_{\beta,0}^{(0)} + T_{\beta,0}^{(0)} \\ &= 2\left(E_{Y_0}\left[A_{\beta_0}\cdot\left(h_{C2}\times \breve{r}_{C2}\right)\right] + S_{\beta,0}^{(0)}\right) + \frac{Var\left[\Delta_0\right]}{E\left[\Delta_0\right]^2} S_{\beta,0}^{(0)} \end{aligned} \tag{D.49}$$

We then put together the expansions of D_β, $S_{\beta,0}$ and T_β to obtain the expansion for Ω_β given in (36). The terms of order ϵ^0 are given in the statement of the Theorem, while the terms of order ϵ^1 are:

$$\begin{aligned} \Omega_\theta^{(1)} &= \frac{E\left[\Delta_0\right]\left(D_\theta^{(0)}\left(S_{\beta,0}^{(0)} + T_\beta^{(0)}\right) - 2D_\theta^{(1)}T_\beta^{(-1)}\right)}{\left(D_\theta^{(0)}\right)^3} \\ \Omega_{\sigma^2}^{(1)} &= \frac{E\left[\Delta_0\right]\left(D_{\sigma^2}^{(0)}\left(S_{\beta,0}^{(0)} + T_\beta^{(0)}\right) - 2D_{\sigma^2}^{(1)}T_\beta^{(-1)}\right)}{\left(D_{\sigma^2}^{(0)}\right)^3} \end{aligned}$$

when estimating θ or σ^2 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*

$$
\begin{aligned}
E_{Y_0}\left[\phi q_{C2}\right] &= \frac{1}{2}E_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\left(\frac{\partial\phi}{\partial y_1} + \frac{\partial\phi}{\partial y_0}\right)\right] && \text{(D.50)}\\
E_{Y_0}\left[\sigma^2\phi\frac{\partial r_{C2}}{\partial y}\right] &= E_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\right] && \text{(D.51)}\\
E_{Y_0}\left[\sigma^4\phi\frac{\partial^2 r_{C2}}{\partial y^2}\right] &= E_{Y_0}\left[\sigma^4\left(\frac{\partial\psi_0}{\partial y}\psi_1 - \psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\right]. && \text{(D.52)}
\end{aligned}
$$

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

Proof. From the form of q_{C2} given in (D.39), we have

$$
\begin{aligned}
E_{Y_0}\left[\phi q_{C2}\right] &= E_{Y_0}\left[\phi \times \left(\left\{A_{\beta_0}\cdot\psi_1(y_0,\beta_0)\right\}\times\psi_0(y_0,\beta_0) - \left\{B_{\beta_0}\cdot\psi_0(y_0,\beta_0)\right\}\times\psi_1(y_0,\beta_0)\right)\right]\\
&= E_{Y_0}\left[\phi\times\left(\left\{\mu\frac{\partial\psi_1}{\partial y} + \frac{\sigma^2}{2}\frac{\partial^2\psi_1}{\partial y^2}\right\}\psi_0 - \left\{\mu\frac{\partial\psi_0}{\partial y} + \frac{\sigma^2}{2}\frac{\partial^2\psi_0}{\partial y^2}\right\}\times\psi_1\right)\right]\\
&= E_{Y_0}\left[\mu\phi\left\{\frac{\partial\psi_1}{\partial y}\psi_0 - \frac{\partial\psi_0}{\partial y}\psi_1\right\}\psi_0 + \frac{\sigma^2}{2}\phi\left\{\frac{\partial^2\psi_1}{\partial y^2}\psi_0 - \frac{\partial^2\psi_0}{\partial y^2}\psi_1\right\}\right]\\
&= -E_{Y_0}\left[\frac{\sigma^2}{2}\frac{d}{dy_0}\left(\phi\left\{\frac{\partial\psi_1}{\partial y}\psi_0 - \frac{\partial\psi_0}{\partial y}\psi_1\right\}\right)\right] + E_{Y_0}\left[\frac{\sigma^2}{2}\phi\left\{\frac{\partial^2\psi_1}{\partial y^2}\psi_0 - \frac{\partial^2\psi_0}{\partial y^2}\psi_1\right\}\right]\\
&= -E_{Y_0}\left[\frac{\sigma^2}{2}\left\{\frac{\partial\psi_1}{\partial y}\psi_0 - \frac{\partial\psi_0}{\partial y}\psi_1\right\}\frac{d\phi}{dy_0}\right] - E_{Y_0}\left[\frac{\sigma^2}{2}\left\{\frac{\partial^2\psi_1}{\partial y^2}\psi_0 - \frac{\partial^2\psi_0}{\partial y^2}\psi_1\right\}\phi\right]\\
&\quad + E_{Y_0}\left[\frac{\sigma^2}{2}\phi\left\{\frac{\partial^2\psi_1}{\partial y^2}\psi_0 - \frac{\partial^2\psi_0}{\partial y^2}\psi_1\right\}\right]\\
&= \frac{1}{2}E_{Y_0}\left[\sigma^2\left\{\frac{\partial\psi_1}{\partial y}\psi_0 - \frac{\partial\psi_0}{\partial y}\psi_1\right\}\frac{d\phi}{dy_0}\right]
\end{aligned}
$$

with the fourth equality following from

$$
E_{Y_0}\left[\mu(Y_0,\theta_0)f(Y_0)\right] = -E_{Y_0}\left[\frac{\sigma^2(Y_0,\gamma_0)}{2}\frac{df(Y_0)}{dy_0}\right]. \quad \text{(D.53)}
$$

Then we have

$$
\frac{d\phi}{dy_0} = \frac{d\phi\left(y_0, y_0, \beta_0\right)}{dy_0} = \frac{\partial\phi}{\partial y_1} + \frac{\partial\phi}{\partial y_0}.
$$

Next, as in the development (C.31) in Lemma 1, we have

$$
\begin{aligned}
E_{Y_0}\left[\sigma^2\phi\frac{\partial r_{C2}}{\partial y}\right] &= \int\frac{\partial r_{C2}}{\partial y}\sigma^2\phi\pi dy\\
&= 2E_{Y_0}\left[\left(\int^{Y_0}\phi dz_0\right)q_{C2}\right]
\end{aligned}
$$

from which it follows that

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

by applying (D.50).

Next, the same development as (C.32) in Lemma 1 now gives

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

and we apply (D.50) and (D.51) to obtain:

$$E_{Y_0}\left[\sigma^4\phi\frac{\partial^2 r_{C2}}{\partial y^2}\right]=-2E_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1-\psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\mu\right]-E_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1-\psi_0\frac{\partial\psi_1}{\partial y}\right)\frac{d\left(\sigma^2\phi\right)}{dy_0}\right]$$

But from (D.53) it follows that

$$\begin{aligned}-2E_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1-\psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\mu\right] &= E_{Y_0}\left[\sigma^2\frac{d}{dy_0}\left(\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1-\psi_0\frac{\partial\psi_1}{\partial y}\right)\phi\right)\right]\\ &= E_{Y_0}\left[\sigma^2\left(\frac{\partial\psi_0}{\partial y}\psi_1-\psi_0\frac{\partial\psi_1}{\partial y}\right)\frac{d\left(\sigma^2\phi\right)}{dy_0}\right]\\ &\quad +E_{Y_0}\left[\sigma^4\phi\frac{\partial}{\partial y}\left(\left(\frac{\partial\psi_0}{\partial y}\psi_1-\psi_0\frac{\partial\psi_1}{\partial y}\right)\right)\right]\end{aligned}$$

and therefore

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

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.

Godambe, V. P., 1960. An optimum property of regular maximum likelihood estimation. Annals of Mathematical Statistics 31, 1208–1211.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

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.

Heyde, C. C., 1997. Quasi-Likelihood and Its Application. Springer-Verlag, 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.

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