

NBER WORKING PAPER SERIES

HOW OFTEN TO SAMPLE A CONTINUOUS-TIME
PROCESS IN THE PRESENCE OF
MARKET MICROSTRUCTURE NOISE

Yacine Aït-Sahalia
Per A. Mykland

Working Paper 9611
<http://www.nber.org/papers/w9611>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2003

Financial support from the NSF under grants SBR-0111140 (Aït-Sahalia) and DMS-0204639 (Mykland) is gratefully acknowledged. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

©2003 by Yacine Aït-Sahalia and Per A. Mykland. All rights reserved. Short sections of text not to exceed two paragraphs, may be quoted without explicit permission provided that full credit including ©notice, is given to the source.

How Often to Sample a Continuous-Time Process in the Presence of
Market Microstructure Noise

Yacine Aït-Sahalia and Per A. Mykland

NBER Working Paper No. 9611

April 2003

JEL No. G12, C22

ABSTRACT

Classical statistics suggest that for inference purposes one should always use as much data as is available. We study how the presence of market microstructure noise in high-frequency financial data can change that result. We show that the optimal sampling frequency at which to estimate the parameters of a discretely sampled continuous-time model can be finite when the observations are contaminated by market microstructure effects. We then address the question of what to do about the presence of the noise. We show that modelling the noise term explicitly restores the first order statistical effect that sampling as often as possible is optimal. But, more surprisingly, we also demonstrate that this is true even if one misspecifies the assumed distribution of the noise term. Not only is it still optimal to sample as often as possible, but the estimator has the same variance as if the noise distribution had been correctly specified, implying that attempts to incorporate the noise into the analysis cannot do more harm than good. Finally, we study the same questions when the observations are sampled at random time intervals, which are an essential feature of transaction-level data.

Yacine Aït-Sahalia
Bendheim Center for Finance
Princeton University
Princeton, NJ 08544-1021
and NBER
yacine@princeton.edu

Per A. Mykland
Department of Statistics
University of Chicago
Chicago, IL 60637-1514
Mykland@galton.uchicago.edu

The notion that the observed transaction price in high frequency financial data is the unobservable efficient price plus some noise component due to the imperfections of the trading process is a well established concept in the market microstructure literature (see for instance Black (1986)). In this paper, we study the implications of such a data generating process for the estimation of the parameters of the continuous-time efficient price process, using discretely sampled data on the transaction price process. In particular, we focus on the effects of the presence of the noise for the estimation of the variance of asset returns, σ^2 . In the absence of noise, it is well known that the quadratic variation of the process (i.e., the average sum of squares of log-returns measured at high frequency) estimates σ^2 . In theory, sampling as often as possible will produce in the limit a perfect estimate of σ^2 . We show, however, that the situation changes radically in the presence of market microstructure noise that is not taken into account in the analysis.

We start by asking whether it remains optimal to sample the price process as often as possible in the presence of market microstructure noise, consistently with the basic statistical principle that, *ceteris paribus*, more data is preferred to less. We show that, if noise is present but unaccounted for, then the optimal sampling frequency is finite. The intuition for this result is as follows. The volatility of the underlying efficient price process and the market microstructure noise tend to behave differently at different frequencies. Thinking in terms of signal-to-noise ratio, a log-return observed from transaction prices over a tiny time interval is mostly composed of market microstructure noise and brings little information regarding the volatility of the price process since the latter is (at least in the Brownian case) proportional to the time interval separating successive observations. As the time interval separating the two prices in the log-return increases, the amount of market microstructure noise remains constant, since each price is measured with error, while the informational content of volatility increases. Hence very high frequency data are mostly composed of market microstructure noise, while the volatility of the price process is more apparent in longer horizon returns. Running counter to this effect is the basic statistical principle mentioned above: in an idealized setting where the data are observed without error, sampling more frequently cannot hurt. What we show is that these two effects compensate each other and result in a finite optimal sampling frequency (in the root mean squared error sense).

We then address the question of what to do about the presence of the noise. If, convinced by either the empirical evidence and/or the theoretical market microstructure models, one decides to account for the presence of the noise, how should one go about doing it? We show that modelling

the noise term explicitly restores the first order statistical effect that sampling as often as possible is optimal. But, more surprisingly, we also demonstrate that this is true *even if* one misspecifies the assumed distribution of the noise term. If the econometrician assumes that the noise terms are normally distributed when in fact they are not, not only is it still optimal to sample as often as possible (unlike the result when no allowance is made for the presence of noise), but the estimator has the *same variance* as if the noise distribution had been correctly specified. Put differently, attempts to include a noise term in the econometric analysis *cannot* do more harm than good. This robustness result, we think, is a major argument in favor of incorporating the presence of the noise when estimating continuous time models with high frequency financial data, even if one is unsure about what is the true distribution of the noise term. Finally, we study the same questions when the observations are sampled at random time intervals, which are an essential feature of transaction-level data.

Our results also have implications for the two parallel tracks that have developed in the recent financial econometrics literature dealing with discretely observed continuous-time processes. One strand of the literature has argued that estimation methods should be robust to the potential issues arising in the presence of high frequency data and, consequently, be asymptotically valid without requiring that the sampling interval Δ separating successive observations tend to zero (see, e.g., Hansen and Scheinkman (1995), Aït-Sahalia (1996) and Aït-Sahalia (2002)). Another strand of the literature has dispensed with that constraint, and the asymptotic validity of these methods requires that Δ tend to zero instead of or in addition to, an increasing length of time T over which these observations are recorded (see, e.g., Andersen, Bollerslev, Diebold, and Labys (2003), Bandi and Phillips (2003) and Barndorff-Nielsen and Shephard (2002)).

The first strand of literature has been informally warning about the potential dangers of using high frequency financial data without accounting for their inherent noise (see e.g., page 529 of Aït-Sahalia (1996)), and we propose a formal modelization of that phenomenon. The implications of our analysis are most salient for the second strand of the literature, which is predicated on the use of high frequency data but does not account for the presence of market microstructure noise. Our results show that the properties of estimators based on the local sample path properties of the process (such as the quadratic variation to estimate σ^2) change dramatically in the presence of noise, while at the same time we suggest a robust approach to correcting for the presence of market microstructure noise.

The paper is organized as follows. We start by describing in Section 1 our reduced form setup and the underlying structural models that support it. We then review in Section 2 the base case where no noise is present, before analyzing in Section 3 the situation where the presence of the noise is ignored. Next, we show in Section 4 that accounting for the presence of the noise restores the optimality of high frequency sampling. Our robustness results are presented in Section 5 and interpreted in Section 6. We incorporate random sampling intervals into the analysis in Section 7, and a drift term in 8. Sections 9 and 10 present two further relaxation of our assumptions, to serially correlated and cross-correlated noise respectively. Section 11 concludes. All proofs are in the Appendix.

1. Setup

Our basic setup is as follows. We assume that the underlying process of interest, typically the log-price of a security, is a time-homogenous diffusion on the real line

$$dX_t = \mu(X_t; \theta)dt + \sigma dW_t \tag{1.1}$$

where $X_0 = 0$, W_t is a Brownian motion, $\mu(\cdot, \cdot)$ is the drift function, σ^2 the diffusion coefficient and θ the drift parameters, $\theta \in \Theta$ and $\sigma > 0$. The parameter space is an open and bounded set. As discussed in Aït-Sahalia and Mykland (2003), the properties of parametric estimators in this model are quite different depending upon we estimate θ alone, σ^2 alone, or both parameters together. When the data are noisy, the main effects that we describe are already present in the simpler of these three cases, where σ^2 alone is estimated, and so we focus on that case. Moreover, in the high frequency context we have in mind, the diffusive component of (1.1) is of order $(dt)^{1/2}$ while the drift component is of order dt only, so the drift component is mathematically negligible at high frequencies. This is validated empirically: including a drift actually deteriorates the performance of variance estimates from high frequency data since the drift is estimated with a large standard error. Not centering the log returns for the purpose of variance estimation produces more accurate results (see Merton (1980)). So we simplify the analysis one step further by setting $\mu = 0$, which we do until Section 8, where we then show that adding a drift term does not alter our results.

In that case,

$$X_t = \sigma W_t. \tag{1.2}$$

Until Section 7, we treat the case where we observations occur at equidistant time intervals Δ , in

which case the parameter σ^2 is therefore estimated at time T on the basis of $N+1$ discrete observations recorded at times $\tau_0 = 0, \tau_1 = \Delta, \dots, \tau_N = N\Delta = T$. In Section 7, we let the sampling intervals be themselves random variables, since this feature is an essential characteristic of high frequency transaction data.

Where we depart from the inference setup previously studied in Aït-Sahalia and Mykland (2003) is that we now assume that, instead of observing the process X at dates τ_i , we observe X with error:

$$\tilde{X}_{\tau_i} = X_{\tau_i} + U_{\tau_i} \tag{1.3}$$

where the U_{τ_i} 's are iid noise with mean zero and variance a^2 and are independent of the W process. In that context, we view X as the efficient log-price, while the observed \tilde{X} is the transaction log-price. In an efficient market, X_t is the log of the expectation of the final value of the security conditional on all publicly available information at time t . It corresponds to the log-price that would be in effect in a perfect market with no trading imperfections, frictions, or informational effects. The Brownian motion W is the process representing the arrival of new information, which in this idealized setting is immediately impounded in X .

By contrast, U_t summarizes the noise generated by the mechanics of the trading process. What we have in mind as the source of noise is a diverse array of market microstructure effects, either information or non-information related, such as the presence of a bid-ask spread and the corresponding bounces, the differences in trade sizes and the corresponding differences in representativeness of the prices, the different informational content of price changes due to informational asymmetries of traders, the gradual response of prices to a block trade, the strategic component of the order flow, inventory control effects, the discreteness of price changes in markets that are not decimalized, etc. That these phenomena are real and important is an accepted fact in the market microstructure literature, both theoretical and empirical. One can in fact argue that these phenomena justify this literature.

We view (1.3) as the simplest possible reduced form of structural market microstructure models. Our specification coincides with that of Hasbrouck (1993), who discusses the theoretical market microstructure underpinnings of such a model, estimates the value of the parameter a to be 0.33% (since the model is set in terms of log prices, a is a percentage of the asset price) and argues that the parameter a is a summary measure of market quality. Simple structural market microstructure models will indeed generate (1.3). For instance, Roll (1984) proposes a model where U is due entirely to the bid-ask spread, and takes the form $U = \pm \text{spread}/2$. A disturbance U can also be generated

by adverse selection effects as in Glosten (1987), where the spread has two components: one that is due to monopoly power, clearing costs, inventory carrying costs, etc., as previously, and a second one that arises because of adverse selection whereby the specialist is concerned that the investor on the other side of the transaction has superior information. In that situation, the disturbance U would no longer be uncorrelated with the W process and would exhibit autocorrelation at the first order, which would complicate our analysis without fundamentally altering it: see Sections 9 and 10 where we relax the assumptions that the U 's are serially uncorrelated and independent of the W process, respectively. The situation where the measurement error is primarily due to the fact that transaction prices are multiples of a tick size (i.e., $\tilde{X}_{\tau_i} = m_i \kappa$ where κ is the tick size and m_i is the integer closest to X_{τ_i}/κ) can be modeled as a rounding off problem (see Jacod (1996) and Delattre and Jacod (1997)). Finally, more complex structural models, such as that of Madhavan, Richardson, and Roomans (1997), also give rise to reduced forms where the observed transaction price \tilde{X} takes the form of an unobserved fundamental value plus error.

With (1.3) as our basic data generating process, we now turn to the questions we address in this paper: how often should one sample a continuous-time process when the data are subject to market microstructure noise, what are the implications of the noise for the estimation of the parameters of the X process, and how should one correct for the presence of the noise, allowing for the possibility that the econometrician misspecifies the assumed distribution of the noise term, and finally allowing for the sampling to occur at random points in time? We proceed from the simplest to the most complex situation by adding one extra layer of complexity at a time: Figure 1 shows the three sampling schemes we consider, starting with fixed sampling without market microstructure noise, then moving to fixed sampling with noise and concluding with an analysis of the situation where transaction prices are not only subject to microstructure noise but are also recorded at random time intervals.

2. The Baseline Case: No Microstructure Noise

We start by briefly reviewing what would happen in the absence of market microstructure noise, that is when $a = 0$. With X denoting the log-price, the first differences of the observations are the log-returns $Y_i = \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}}$, $i = 1, \dots, N$. The observations $Y_i = \sigma (W_{\tau_{i+1}} - W_{\tau_i})$ are then iid $N(0, \sigma^2 \Delta)$ so the likelihood function is

$$l(\sigma^2) = -N \ln(2\pi\sigma^2\Delta)/2 - (2\sigma^2\Delta)^{-1} Y'Y, \quad (2.1)$$

where $Y = (Y_1, \dots, Y_N)'$. The maximum-likelihood estimator of σ^2 coincides with the discrete approximation to the quadratic variation of the process

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^N Y_i^2 \quad (2.2)$$

which has the following exact small sample moments:

$$E[\hat{\sigma}^2] = \frac{1}{T} \sum_{i=1}^N E[Y_i^2] = \frac{N(\sigma^2 \Delta)}{T} = \sigma^2,$$

$$Var[\hat{\sigma}^2] = \frac{1}{T^2} Var\left[\sum_{i=1}^N Y_i^2\right] = \frac{1}{T^2} \left(\sum_{i=1}^N Var[Y_i^2]\right) = \frac{N}{T^2} (2\sigma^4 \Delta^2) = \frac{2\sigma^4 \Delta}{T}$$

and the following asymptotic distribution

$$T^{1/2}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{T \rightarrow \infty} N(0, \omega) \quad (2.3)$$

where

$$\omega = AVAR(\hat{\sigma}^2) = \Delta E\left[-\ddot{l}(\sigma^2)\right]^{-1} = 2\sigma^4 \Delta. \quad (2.4)$$

Thus selecting Δ as small as possible is optimal for the purpose of estimating σ^2 .

3. When the Observations Are Noisy But the Noise Is Ignored

Suppose now that market microstructure noise is present but the presence of the U 's is ignored when estimating σ^2 . In other words, we use the log-likelihood (2.1) even though the true structure of the observed log-returns Y_i 's is given by an MA(1) process since

$$\begin{aligned} Y_i &= \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}} + U_{\tau_i} - U_{\tau_{i-1}} \\ &= \sigma(W_{\tau_i} - W_{\tau_{i-1}}) + U_{\tau_i} - U_{\tau_{i-1}} \\ &\equiv \varepsilon_i + \eta \varepsilon_{i-1} \end{aligned} \quad (3.1)$$

where the ε_i 's are iid with mean zero and variance γ^2 . The relationship to the original parametrization (σ^2, a^2) is given by

$$\gamma^2(1 + \eta^2) = Var[Y_i] = \sigma^2 \Delta + 2a^2 \quad (3.2)$$

$$\gamma^2 \eta = cov(Y_i, Y_{i-1}) = -a^2 \quad (3.3)$$

or equivalently

$$\gamma^2 = \frac{1}{2} \left\{ 2a^2 + \sigma^2 \Delta + \sqrt{\sigma^2 \Delta (4a^2 + \sigma^2 \Delta)} \right\} \quad (3.4)$$

$$\eta = \frac{1}{2a^2} \left\{ -2a^2 - \sigma^2 \Delta + \sqrt{\sigma^2 \Delta (4a^2 + \sigma^2 \Delta)} \right\} \quad (3.5)$$

which implies, as required, that $-1 < \eta < 0$.

The estimator $\hat{\sigma}^2$ obtained from maximizing the misspecified log-likelihood has the following properties:

Proposition 1. *In small samples (finite T), the RMSE of the estimator $\hat{\sigma}^2$ is given by*

$$\begin{aligned} RMSE [\hat{\sigma}^2] &= \left((E[\hat{\sigma}^2] - \sigma^2)^2 + Var[\hat{\sigma}^2] \right)^{1/2} \\ &= \left(\frac{4a^4}{\Delta^2} + \frac{2(\sigma^4 \Delta^2 + 4\sigma^2 \Delta a^2 + 6a^4)}{T\Delta} - \frac{4a^4}{T^2} \right)^{1/2} \end{aligned} \quad (3.6)$$

and has a minimum (provided that T is greater than $2^{1/2}a^2/\sigma^2$) which is reached at

$$\Delta^* = \frac{2^{1/3}a^{4/3}}{\sigma^2} \left(\left(\sigma^2 T + (\sigma^4 T^2 - 2a^4)^{1/2} \right)^{1/3} + \frac{2^{1/3}a^{4/3}}{\left(\sigma^2 T + (\sigma^4 T^2 - 2a^4)^{1/2} \right)^{1/3}} \right). \quad (3.7)$$

As T grows, we have

$$\Delta^* = \frac{2^{2/3}a^{4/3}}{\sigma^{4/3}} T^{1/3} + O\left(\frac{1}{T^{1/3}}\right).$$

Figure 2 displays the RMSE of the estimator as a function of Δ and T , with parameter values $\sigma = 0.5$ and $a = 0.05$. Complementary to this are the results of Gloter and Jacod (2000) which show that the presence of even increasingly negligible noise is sufficient to adversely affect the identification of σ^2 . They study the asymptotic distribution of $\hat{\sigma}^2$ when the standard deviation a_N of the noise term goes to zero as the sample size increases, showing that $\hat{\sigma}^2$ is consistent if and only if Na_N goes to zero, and characterizing the asymptotic distribution of $\hat{\sigma}^2$ as a function of whether $N^{3/2}a_N$ goes to zero, a finite constant, or infinity.

4. Incorporating Market Microstructure Noise Explicitly

Now we show that if we explicitly incorporate the U 's into the likelihood function, then we are back into the standard case where the optimal sampling scheme consists in sampling as often as

possible. Suppose that the microstructure noise is normally distributed, an assumption we will relax below in Section 5. The likelihood function for the Y 's is then given by

$$l(\eta, \gamma^2) = -\ln \det(V)/2 - N \ln(2\pi\gamma^2)/2 - (2\gamma^2)^{-1} Y' V^{-1} Y, \quad (4.1)$$

where the covariance matrix for the vector $Y = (Y_1, \dots, Y_N)'$ is given by $\gamma^2 V$, where

$$V = [v_{ij}]_{i,j=1,\dots,N} = \begin{pmatrix} 1 + \eta^2 & \eta & 0 & \dots & 0 \\ \eta & 1 + \eta^2 & \eta & \ddots & \vdots \\ 0 & \eta & 1 + \eta^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \eta \\ 0 & \dots & 0 & \eta & 1 + \eta^2 \end{pmatrix} \quad (4.2)$$

Further,

$$\det(V) = \frac{1 - \eta^{2N+2}}{1 - \eta^2} \quad (4.3)$$

and, neglecting the end effects, an approximate inverse of V is the matrix $\Omega = [\omega_{ij}]_{i,j=1,\dots,N}$ where

$$\omega_{ij} = (1 - \eta^2)^{-1} (-\eta)^{|i-j|}$$

(see Durbin (1959)). The product $V\Omega$ differs from the identity matrix only on the first and last rows.

The exact inverse is $V^{-1} = [v^{ij}]_{i,j=1,\dots,N}$ where

$$v^{ij} = (1 - \eta^2)^{-1} (1 - \eta^{2N+2})^{-1} \left\{ (-\eta)^{|i-j|} - (-\eta)^{i+j} - (-\eta)^{2N-i-j+2} - (-\eta)^{2N+|i-j|+2} + (-\eta)^{2N+i-j+2} + (-\eta)^{2N-i+j+2} \right\}. \quad (4.4)$$

(see Shaman (1969) and Haddad (1995)).

We then obtain the following for the MLE estimators of σ^2 and a^2 :

Proposition 2. *The MLE $(\hat{\sigma}^2, \hat{a}^2)$ is consistent and its asymptotic variance is given by*

$$AVAR_{normal}(\hat{\sigma}^2, \hat{a}^2) = \begin{pmatrix} 4\sqrt{\sigma^6 \Delta (4a^2 + \sigma^2 \Delta)} + 2\sigma^4 \Delta & -\sigma^2 \Delta h(\Delta, \sigma^2, a^2) \\ \bullet & \frac{\Delta}{2} (2a^2 + \sigma^2 \Delta) h(\Delta, \sigma^2, a^2) \end{pmatrix}.$$

with

$$h(\Delta, \sigma^2, a^2) \equiv 2a^2 + \sqrt{\sigma^2 \Delta (4a^2 + \sigma^2 \Delta)} + \sigma^2 \Delta. \quad (4.5)$$

Since $AVAR_{\text{normal}}(\hat{\sigma}^2)$ is increasing in Δ , it is optimal to sample as often as possible. Further, since

$$AVAR_{\text{normal}}(\hat{\sigma}^2) = 8\sigma^3 a \Delta^{1/2} + 2\sigma^2 \Delta + o(\Delta), \quad (4.6)$$

the loss of efficiency relative to the case where no market microstructure noise is present (and $AVAR(\hat{\sigma}^2) = 2\sigma^2 \Delta$ as given in (2.4)) is at order $\Delta^{1/2}$. Figure 3 plots the asymptotic variances of $\hat{\sigma}^2$ as functions of Δ with and without noise (the parameter values are again $\sigma = 0.5$ and $a = 0.05$).

5. The Effect of Misspecifying the Distribution of the Microstructure Noise

We now study the situation where one attempts to incorporate the presence of the U 's into the analysis, as in Section 4, but assumes a misspecified model for them. Specifically, we consider the case where the U 's are assumed to be normally distributed when in reality they have a different distribution. We still suppose that the U 's are iid with mean zero and variance a^2 .

Since the econometrician assumes the U 's to have a normal distribution, inference is still done with the log-likelihood $l(\sigma^2, a^2)$, or equivalently $l(\eta, \gamma^2)$ given in (4.1), using (3.2)-(3.3). This means that the scores \dot{l}_{σ^2} and \dot{l}_{a^2} , or equivalently (B.1) and (B.2), are used as moment functions (or “estimating equations”). Since the first order moments of the moment functions only depend on the second order moment structure of the log-returns (Y_1, \dots, Y_N) , which is unchanged by the absence of normality, the moment functions are unbiased

$$E_{\text{true}}[\dot{l}_{\eta}] = E_{\text{true}}[\dot{l}_{\gamma^2}] = 0 \quad (5.1)$$

and similarly for \dot{l}_{σ^2} and \dot{l}_{a^2} . Hence the estimator $(\hat{\sigma}^2, \hat{a}^2)$ based on these moment functions is consistent and asymptotically unbiased (even though the likelihood function is misspecified.)

The effect of misspecification lies in the asymptotic variance matrix. We use a technical trick to simplify calculations that would otherwise be daunting. By using the cumulants of the distribution of U , we express the asymptotic variance of these estimators in terms of deviations from normality.

We have that

$$\begin{aligned}
E_{\text{true}} \left[\dot{l}_\eta \dot{l}_{\gamma^2} \right] &= \text{Cov}_{\text{true}}(\dot{l}_\eta, \dot{l}_{\gamma^2}) \\
&= \text{Cov}_{\text{true}} \left(-\frac{1}{2\gamma^2} \sum_{i,j=1}^N Y_i Y_j \frac{\partial v^{ij}}{\partial \eta}, \frac{1}{2\gamma^4} \sum_{k,l=1}^N Y_k Y_l v^{kl} \right) \\
&= -\frac{1}{4\gamma^6} \sum_{i,j,k,l=1}^N \frac{\partial v^{ij}}{\partial \eta} v^{kl} \text{Cov}_{\text{true}}(Y_i Y_j, Y_k Y_l) \\
&= -\frac{1}{4\gamma^6} \sum_{i,j,k,l=1}^N \frac{\partial v^{ij}}{\partial \eta} v^{kl} [\text{Cum}_{\text{true}}(Y_i, Y_j, Y_k, Y_l) + 2\text{Cov}_{\text{true}}(Y_i, Y_j) \text{Cov}_{\text{true}}(Y_k, Y_l)].
\end{aligned} \tag{5.2}$$

where “true” denotes the true distribution of the Y 's, not the incorrectly specified one, and Cum denotes the cumulants. The last transition is because

$$\begin{aligned}
\text{Cov}_{\text{true}}(Y_i Y_j, Y_k Y_l) &= E_{\text{true}} [Y_i Y_j Y_k Y_l] - E_{\text{true}} [Y_i Y_j] E_{\text{true}} [Y_k Y_l] \\
&= \kappa^{ijkl} - \kappa^{ij} \kappa^{kl} \\
&= \kappa^{i,j,k,l} + \kappa^{i,j} \kappa^{k,l} [3] - \kappa^{i,j} \kappa^{k,l} \\
&= \kappa^{i,j,k,l} + \kappa^{i,k} \kappa^{j,l} + \kappa^{i,l} \kappa^{j,k} \\
&= \text{Cum}_{\text{true}}(Y_i, Y_j, Y_k, Y_l) + \text{Cov}_{\text{true}}(Y_i, Y_k) \text{Cov}_{\text{true}}(Y_j, Y_l) \\
&\quad + \text{Cov}_{\text{true}}(Y_i, Y_l) \text{Cov}_{\text{true}}(Y_j, Y_k)
\end{aligned}$$

since Y has mean zero (see e.g., Section 2.3 of McCullagh (1987)). The need for permutation goes away due to the summing over all indices (i, j, k, l) , and since $V^{-1} = [v^{ij}]$ is symmetric.

When looking at (5.2), note that $\text{Cum}_{\text{normal}}(Y_i, Y_j, Y_k, Y_l) = 0$, where “normal” denotes a Normal distribution with the same first and second order moments as the true distribution. That is, if the Y 's were normal we would have

$$E_{\text{normal}} \left[\dot{l}_\eta \dot{l}_{\gamma^2} \right] = -\frac{1}{4\gamma^6} \sum_{i,j,k,l=1}^N \frac{\partial v^{ij}}{\partial \eta} v^{kl} [2\text{Cov}_{\text{normal}}(Y_i, Y_j) \text{Cov}_{\text{normal}}(Y_k, Y_l)].$$

Also, since the covariance structure does not depend on Gaussianity, $\text{Cov}_{\text{true}}(Y_i, Y_j) = \text{Cov}_{\text{normal}}(Y_i, Y_j)$.

Next, we have

$$E_{\text{normal}} \left[\dot{l}_\eta \dot{l}_{\gamma^2} \right] = -E_{\text{normal}} \left[\ddot{l}_{\eta\gamma^2} \right] = -E_{\text{true}} \left[\ddot{l}_{\eta\gamma^2} \right] \tag{5.3}$$

with the last equality following from the fact that $\ddot{l}_{\eta\gamma^2}$ depends only on the second moments of the Y 's. (Note that in general $E_{\text{true}} \left[\dot{l}_\eta \dot{l}_{\gamma^2} \right] \neq -E_{\text{true}} \left[\ddot{l}_{\eta\gamma^2} \right]$ because the likelihood may be misspecified.)

Thus, it follows from (5.2) that

$$\begin{aligned}
E_{\text{true}} \left[\dot{l}_\eta \dot{l}_{\gamma^2} \right] &= E_{\text{normal}} \left[\dot{l}_\eta \dot{l}_{\gamma^2} \right] - \frac{1}{4\gamma^6} \sum_{i,j,k,l=1}^N \frac{\partial v^{ij}}{\partial \eta} v^{kl} \text{Cum}_{\text{true}}(Y_i, Y_j, Y_k, Y_l) \\
&= -E_{\text{true}} \left[\ddot{l}_{\eta\gamma^2} \right] - \frac{1}{4\gamma^6} \sum_{i,j,k,l=1}^N \frac{\partial v^{ij}}{\partial \eta} v^{kl} \text{Cum}_{\text{true}}(Y_i, Y_j, Y_k, Y_l) \tag{5.4}
\end{aligned}$$

It follows similarly that

$$\begin{aligned}
E_{\text{true}} \left[\left(\dot{l}_\eta \right)^2 \right] &= \text{Var}_{\text{true}}(\dot{l}_\eta) \\
&= -E_{\text{true}} \left[\ddot{l}_{\eta\eta} \right] + \frac{1}{4\gamma^4} \sum_{i,j,k,l=1}^N \frac{\partial v^{ij}}{\partial \eta} \frac{\partial v^{kl}}{\partial \eta} \text{Cum}_{\text{true}}(Y_i, Y_j, Y_k, Y_l) \tag{5.5}
\end{aligned}$$

and

$$\begin{aligned}
E_{\text{true}} \left[\left(\dot{l}_{\gamma^2} \right)^2 \right] &= \text{Var}_{\text{true}}(\dot{l}_{\gamma^2}) \\
&= -E_{\text{true}} \left[\ddot{l}_{\gamma^2\gamma^2} \right] + \frac{1}{4\gamma^8} \sum_{i,j,k,l=1}^N v^{ij} v^{kl} \text{Cum}_{\text{true}}(Y_i, Y_j, Y_k, Y_l). \tag{5.6}
\end{aligned}$$

To calculate the fourth cumulant $\text{Cum}_{\text{true}}(Y_i, Y_j, Y_k, Y_l)$, recall from (3.1) that the observed log-returns are

$$Y_i = \sigma (W_{\tau_i} - W_{\tau_{i-1}}) + U_{\tau_i} - U_{\tau_{i-1}}.$$

First, note that the τ_i are nonrandom, and since W is independent of the U 's, and has Gaussian increments. Second, the cumulants are multilinear, so

$$\begin{aligned}
\text{Cum}_{\text{true}}(Y_i, Y_j, Y_k, Y_l) &= \text{Cum}_{\text{true}} \left(\sigma (W_{\tau_i} - W_{\tau_{i-1}}) + U_{\tau_i} - U_{\tau_{i-1}}, \sigma (W_{\tau_j} - W_{\tau_{j-1}}) + U_{\tau_j} - U_{\tau_{j-1}}, \right. \\
&\quad \left. \sigma (W_{\tau_k} - W_{\tau_{k-1}}) + U_{\tau_k} - U_{\tau_{k-1}}, \sigma (W_{\tau_l} - W_{\tau_{l-1}}) + U_{\tau_l} - U_{\tau_{l-1}} \right) \\
&= \sigma^4 \text{Cum}_{\text{true}}(W_{\tau_i} - W_{\tau_{i-1}}, W_{\tau_j} - W_{\tau_{j-1}}, W_{\tau_k} - W_{\tau_{k-1}}, W_{\tau_l} - W_{\tau_{l-1}}) \\
&\quad + \sigma^3 \text{Cum}_{\text{true}}(W_{\tau_i} - W_{\tau_{i-1}}, W_{\tau_j} - W_{\tau_{j-1}}, W_{\tau_k} - W_{\tau_{k-1}}, U_{\tau_l} - U_{\tau_{l-1}})[4] \\
&\quad + \sigma^2 \text{Cum}_{\text{true}}(W_{\tau_i} - W_{\tau_{i-1}}, W_{\tau_j} - W_{\tau_{j-1}}, U_{\tau_k} - U_{\tau_{k-1}}, U_{\tau_l} - U_{\tau_{l-1}})[6] \\
&\quad + \sigma \text{Cum}_{\text{true}}(W_{\tau_i} - W_{\tau_{i-1}}, U_{\tau_j} - U_{\tau_{j-1}}, U_{\tau_k} - U_{\tau_{k-1}}, U_{\tau_l} - U_{\tau_{l-1}})[4] \\
&\quad + \text{Cum}_{\text{true}}(U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_j} - U_{\tau_{j-1}}, U_{\tau_k} - U_{\tau_{k-1}}, U_{\tau_l} - U_{\tau_{l-1}})
\end{aligned}$$

Out of these terms, only the last is nonzero because W has Gaussian increments (so all cumulants of its increments of order greater than two are zero), and is independent of the U 's (so all cumulants

involving increments of both W and U are also zero.) Therefore,

$$Cum_{\text{true}}(Y_i, Y_j, Y_k, Y_l) = Cum_{\text{true}}(U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_j} - U_{\tau_{j-1}}, U_{\tau_k} - U_{\tau_{k-1}}, U_{\tau_l} - U_{\tau_{l-1}})$$

where U is a generic random variable with distribution U_{τ_i} .

If $i = j = k = l$, we have:

$$\begin{aligned} Cum_{\text{true}}(U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_i} - U_{\tau_{i-1}}) &= Cum_4(U_{\tau_i} - U_{\tau_{i-1}}) \\ &= Cum_4(U_{\tau_i}) + Cum_4(-U_{\tau_{i-1}}) \\ &= 2 Cum_4(U) \end{aligned}$$

with the second equality following from the independence of U_{τ_i} and $U_{\tau_{i-1}}$, and the third from the fact that the cumulant is of even order. $Cum_4(U)$ denotes the fourth cumulant of the random variable U , which has mean zero, so, in terms of the moments of U we have

$$Cum_4(U) = E[U^4] - 3(E[U^2])^2. \quad (5.7)$$

If $\max(i, j, k, l) = \min(i, j, k, l) + 1$, two situations arise. Set $m = \min(i, j, k, l)$ and $M = \max(i, j, k, l)$. Also set $s = s(i, j, k, l) = \#\{i, j, k, l = m\}$. If s is odd, say $s = 1$ with $i = m$, and $j, k, l = M = m + 1$, we get a term of the form

$$Cum_{\text{true}}(U_{\tau_m} - U_{\tau_{m-1}}, U_{\tau_{m+1}} - U_{\tau_m}, U_{\tau_{m+1}} - U_{\tau_m}, U_{\tau_{m+1}} - U_{\tau_m}) = -Cum_4(U_{\tau_m}).$$

By permutation, the same situation arises if $s = 3$. If instead s is even, i.e., $s = 2$, then we have terms of the form

$$Cum_{\text{true}}(U_{\tau_m} - U_{\tau_{m-1}}, U_{\tau_m} - U_{\tau_{m-1}}, U_{\tau_{m+1}} - U_{\tau_m}, U_{\tau_{m+1}} - U_{\tau_m}) = Cum_4(U_{\tau_m}).$$

Finally, if at least one pair of indices in the quadruple (i, j, k, l) is more than one integer apart, then

$$Cum_{\text{true}}(U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_j} - U_{\tau_{j-1}}, U_{\tau_k} - U_{\tau_{k-1}}, U_{\tau_l} - U_{\tau_{l-1}}) = 0$$

by independence of the U 's.

Putting it all together, we have

$$Cum_{\text{true}}(Y_i, Y_j, Y_k, Y_l) = \begin{cases} 2 Cum_4(U) & \text{if } i = j = k = l \\ (-1)^s Cum_4(U) & \text{if } \max(i, j, k, l) = \min(i, j, k, l) + 1 \\ 0 & \text{otherwise} \end{cases} \quad (5.8)$$

We now need to evaluate the sums that appear on the right hand sides of (5.4), (5.5) and (5.6).

We obtain:

Theorem 1. *The estimators $(\hat{\sigma}^2, \hat{a}^2)$ obtained by maximizing the log-likelihood (4.1) are consistent and their asymptotic variance is given by*

$$AVAR_{true}(\hat{\sigma}^2, \hat{a}^2) = AVAR_{normal}(\hat{\sigma}^2, \hat{a}^2) + Cum_4(U) \begin{pmatrix} 0 & 0 \\ 0 & \Delta \end{pmatrix} \quad (5.9)$$

where $AVAR_{normal}(\hat{\sigma}^2, \hat{a}^2)$ is the asymptotic variance in the case where the distribution of U is Normal, that is, the expression given in Proposition 2.

6. Robustness to Misspecification of the Noise Distribution

The above Theorem 1 has implications for the use of the Gaussian likelihood l that go beyond consistency, namely that this likelihood can also be used to estimate the distribution of $\hat{\sigma}^2$ under misspecification. With l denoting the log-likelihood assuming that the U 's are Gaussian, given in (4.1), $-\ddot{l}(\hat{\sigma}^2, \hat{a}^2)$ denote the observed information matrix in the original parameters σ^2 and a^2 . Then

$$\hat{V} = \widehat{AVAR}_{normal} = \left(-\frac{1}{T} \ddot{l}(\hat{\sigma}^2, \hat{a}^2) \right)^{-1}$$

is the usual estimate of asymptotic variance when the distribution is correctly specified as Gaussian. Also note, however, that otherwise, so long as $(\hat{\sigma}^2, \hat{a}^2)$ is consistent, \hat{V} is also a consistent estimate of the matrix $AVAR_{normal}(\hat{\sigma}^2, \hat{a}^2)$. Since this matrix coincides with $AVAR_{true}(\hat{\sigma}^2, \hat{a}^2)$ for all but the (a^2, a^2) term (see (5.9)), the asymptotic variance of $T^{1/2}(\hat{\sigma}^2 - \sigma^2)$ is consistently estimated by $\hat{V}_{\sigma^2 \sigma^2}$. The similar statement is true for the covariances, but not, obviously, for the asymptotic variance of $T^{1/2}(\hat{a}^2 - a^2)$.

In the likelihood context, the possibility of estimating the asymptotic variance by the observed information is due to the second Bartlett identity. For a general log likelihood l , if $S \equiv E_{true}[\dot{l}']/N$ and $D \equiv -E_{true}[\ddot{l}]/N$ (differentiation refers to the original parameters (σ^2, a^2) , not the transformed parameters (γ^2, η)) this identity says that

$$S - D = 0. \quad (6.1)$$

It implies that the asymptotic variance takes the form

$$AVAR = \Delta(DS^{-1}D)^{-1} = \Delta D^{-1}. \quad (6.2)$$

It is clear that (6.2) remains valid if the second Bartlett identity holds only to first order, i.e.,

$$S - D = o(1) \quad (6.3)$$

as $N \rightarrow \infty$, for a general criterion function l which satisfies $E_{\text{true}}[\dot{l}] = o(N)$.

However, in view of Theorem 1, equation (6.3) cannot be satisfied. In fact, we show in Appendix D that

$$S - D = Cum_4(U)gg' + o(1), \quad (6.4)$$

where

$$g = \begin{pmatrix} g_{\sigma^2} \\ g_{a^2} \end{pmatrix} = \begin{pmatrix} \frac{\Delta^{1/2}}{\sigma(4a^2 + \sigma^2\Delta)^{3/2}} \\ \frac{1}{2a^4} \left(1 - \frac{\Delta^{1/2}\sigma(6a^2 + \sigma^2\Delta)}{(4a^2 + \sigma^2\Delta)^{3/2}} \right) \end{pmatrix}. \quad (6.5)$$

From (6.5), we see that $g \neq 0$ whenever $\sigma^2 > 0$. This is consistent with the result in Theorem 1 that the true asymptotic variance matrix, $AVAR_{\text{true}}(\hat{\sigma}^2, \hat{a}^2)$, does not coincide with the one for Gaussian noise, $AVAR_{\text{normal}}(\hat{\sigma}^2, \hat{a}^2)$. On the other hand, the 2×2 matrix gg' is of rank 1, signaling that there exist linear combinations that will cancel out the first column of $S - D$. From what we already know of the form of the correction matrix, D^{-1} gives such a combination, so that the asymptotic variance of the original parameters (σ^2, a^2) will have the property that its first column is not subject to correction in the absence of normality.

A curious consequence of (6.4) is that while the observed information can be used to estimate the asymptotic variance of $\hat{\sigma}^2$ when a^2 is not known, this is not the case when a^2 is known. This is because the second Bartlett identity also fails to first order when considering a^2 to be known, i.e., when differentiating with respect to σ^2 only. Indeed, in that case we have from the upper left component in the matrix equation (6.4)

$$\begin{aligned} S_{\sigma^2\sigma^2} - D_{\sigma^2\sigma^2} &= N^{-1}E_{\text{true}} \left[\dot{l}_{\sigma^2\sigma^2}(\sigma^2, a^2)^2 \right] + N^{-1}E_{\text{true}} \left[\ddot{l}_{\sigma^2\sigma^2}(\sigma^2, a^2) \right] \\ &= Cum_4(U) (g_{\sigma^2})^2 + o(1) \end{aligned}$$

which is not $o(1)$ unless $Cum_4(U) = 0$.

To make the connection between Theorem 1 and the second Bartlett identity, one needs to go to the log profile likelihood

$$\lambda(\sigma^2) \equiv \sup_{a^2} l(\sigma^2, a^2). \quad (6.6)$$

Obviously, maximizing the likelihood $l(\sigma^2, a^2)$ is the same as maximizing $\lambda(\sigma^2)$. Thus one can think of σ^2 as being estimated (when a^2 is unknown) by maximizing the criterion function $\lambda(\sigma^2)$, or by solving $\dot{\lambda}(\hat{\sigma}^2) = 0$. Also, the observed profile information is related to the original observed information by

$$\ddot{\lambda}(\hat{\sigma}^2)^{-1} = \left[\ddot{l}(\hat{\sigma}^2, \hat{a}^2)^{-1} \right]_{\sigma^2\sigma^2}, \quad (6.7)$$

i.e., the first (upper left hand corner) component of the inverse observed information in the original problem. We recall the rationale for equation (6.7) in Appendix D, where we also show that $E_{\text{true}}[\dot{\lambda}] = o(N)$. In view of Theorem 1, $\ddot{\lambda}(\hat{\sigma}^2)$ can be used to estimate the asymptotic variance of $\hat{\sigma}^2$ under the true (possibly non-Gaussian) distribution of the U 's, and so it must be that the criterion function λ satisfies (6.3), that is

$$N^{-1}E_{\text{true}}[\dot{\lambda}(\sigma^2)^2] + N^{-1}E_{\text{true}}[\ddot{\lambda}(\sigma^2)] = o(1). \quad (6.8)$$

This is indeed the case, as shown in Appendix D.

This phenomenon is related, although not identical, to what occurs in the context of quasi-likelihood (for comprehensive treatments of quasi-likelihood theory, see the books by McCullagh and Nelder (1989) and Heyde (1997), and the references therein, and for early econometrics examples see Macurdy (1982) and White (1982)). In quasi-likelihood situations, one uses a possibly incorrectly specified score vector which is nevertheless required to satisfy the second Bartlett identity. What makes our situation unusual relative to quasi-likelihood is that the interest parameter σ^2 and the nuisance parameter a^2 are entangled in the same estimating equations (\dot{l}_{σ^2} and \dot{l}_{a^2} from the Gaussian likelihood) in such a way that the estimate of σ^2 depends, to first order, on whether a^2 is known or not. This is unlike the typical development of quasi-likelihood, where the nuisance parameter separates out (see, e.g., Table 9.1 (p. 326) of McCullagh and Nelder (1989)). Thus only by going to the profile likelihood λ can one make the usual comparison to quasi-likelihood.

7. Randomly Spaced Sampling Intervals

One could of course argue that we have made many simplifying or special assumptions. We now show that none of these assumptions drive the results, and that the effects we describe are all present in more complex (and realistic) setups. We start by relaxing the assumption that Δ is constant.

Indeed, one essential feature of transaction data in finance is that the time that separates successive observations is random, or at least time-varying. So, as in Aït-Sahalia and Mykland (2003), we are led to consider the case where $\Delta_i = \tau_i - \tau_{i-1}$ are either deterministic and time-varying, or random in which case we assume for simplicity that they are iid, independent of the W process. We denote by N_T the number of observations recorded by time T . N_T is random if the Δ 's are. We also suppose that U_{τ_i} can be written U_i , where the U_i are iid and independent of the W process and the Δ_i 's. Thus, the observation noise is the same at all observation times, whether random or nonrandom. If we define the Y_i s as before, in the first two lines of (3.1), though the MA(1) representation

is not valid in the same form.

We can do inference conditionally on the observed sampling times, in light of the fact that the likelihood function using all the available information is

$$L(Y_N, \Delta_N, \dots, Y_1, \Delta_1; \beta, \psi) = L(Y_N, \dots, Y_1 | \Delta_N, \dots, \Delta_1; \beta) \times L(\Delta_N, \dots, \Delta_1; \psi)$$

where β are the parameters of the state process, that is (σ^2, a^2) , and ψ are the parameters of the sampling process, if any (the density of the sampling intervals density $L(\Delta_N, \dots, \Delta_1; \psi)$ may have its own nuisance parameters ψ , such as an unknown arrival rate, but we assume that it does not depend on the parameters β of the state process.) The corresponding log-likelihood function is

$$\sum_{n=1}^N \ln L(Y_N, \dots, Y_1 | \Delta_N, \dots, \Delta_1; \beta) + \sum_{n=1}^{N-1} \ln L(\Delta_N, \dots, \Delta_1; \psi) \quad (7.1)$$

and since we only care about β , we only need to maximize the first term in that sum.

We operate on the covariance matrix Σ of the log-returns Y^l s, now given by

$$\Sigma = \begin{pmatrix} \sigma^2 \Delta_1 + 2a^2 & -a^2 & 0 & \dots & 0 \\ -a^2 & \sigma^2 \Delta_2 + 2a^2 & -a^2 & \ddots & \vdots \\ 0 & -a^2 & \sigma^2 \Delta_3 + 2a^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a^2 \\ 0 & \dots & 0 & -a^2 & \sigma^2 \Delta_n + 2a^2 \end{pmatrix} \quad (7.2)$$

Note that in the equally spaced case, $\Sigma = \gamma^2 V$. The log-likelihood function is given by

$$\begin{aligned} \ln L(Y_N, \dots, Y_1 | \Delta_N, \dots, \Delta_1; \beta) &\equiv l(\sigma^2, a^2) \\ &= -\ln \det(\Sigma)/2 - N \ln(2\pi)/2 - Y' \Sigma^{-1} Y/2, \end{aligned} \quad (7.3)$$

Suppose in the following that β_1 and β_2 can represent either σ^2 or a^2 . We start with:

Lemma 1. *Fisher's Conditional Information is given by*

$$E \left[-\ddot{l}_{\beta_2 \beta_1} \middle| \Delta \right] = -\frac{1}{2} \frac{\partial^2 \ln \det \Sigma}{\partial \beta_2 \beta_1}. \quad (7.4)$$

7.1 Expansion around a fixed value of Δ

To continue further with the calculations, we now expand around a fixed value of Δ , namely $\Delta_0 = E[\Delta]$. Specifically, suppose now that

$$\Delta_i = \Delta_0 (1 + \epsilon \xi_i) \quad (7.5)$$

where ϵ and Δ_0 are nonrandom, the ξ_i^l 's are iid random variables with mean zero. We will Taylor-expand the expressions above around $\epsilon = 0$, i.e., around the non-random sampling case. For simplicity, we take the ξ_i^l 's to be bounded. Denote by Σ_0 the value of Σ when Δ is replaced by Δ_0 , and let Ξ denote the matrix whose diagonal elements are the terms $\Delta_0 \xi_i$, and whose off-diagonal elements are zero. We obtain:

Theorem 2. *The MLE $(\hat{\sigma}^2, \hat{a}^2)$ is again consistent, this time with asymptotic variance*

$$AVAR(\hat{\sigma}^2, \hat{a}^2) = A^{(0)} + \epsilon^2 A^{(2)} + O(\epsilon^3) \quad (7.6)$$

where

$$A^{(0)} = \begin{pmatrix} 4\sqrt{\sigma^6 \Delta_0 (4a^2 + \sigma^2 \Delta_0)} + 2\sigma^4 \Delta_0 & -\sigma^2 \Delta_0 h(\Delta_0, \sigma^2, a^2) \\ \bullet & \frac{\Delta_0}{2} (2a^2 + \sigma^2 \Delta_0) h(\Delta_0, \sigma^2, a^2) \end{pmatrix}$$

and

$$A^{(2)} = \frac{Var[\xi]}{(4a^2 + \Delta_0 \sigma^2)} \begin{pmatrix} A_{\sigma^2 \sigma^2}^{(2)} & A_{\sigma^2 a^2}^{(2)} \\ \bullet & A_{a^2 a^2}^{(2)} \end{pmatrix}$$

with

$$\begin{aligned} A_{\sigma^2 \sigma^2}^{(2)} &= -4 \left(\Delta_0^2 \sigma^6 + \Delta_0^{3/2} \sigma^5 \sqrt{4a^2 + \Delta_0 \sigma^2} \right) \\ A_{\sigma^2 a^2}^{(2)} &= \Delta_0^{3/2} \sigma^3 \sqrt{4a^2 + \Delta_0 \sigma^2} (2a^2 + 3\Delta_0 \sigma^2) + \Delta_0^2 \sigma^4 (8a^2 + 3\Delta_0 \sigma^2) \\ A_{a^2 a^2}^{(2)} &= -\Delta_0^2 \sigma^2 \left(2a^2 + \sigma \sqrt{\Delta_0} \sqrt{4a^2 + \Delta_0 \sigma^2} + \Delta_0 \sigma^2 \right)^2 \end{aligned}$$

Note that $A^{(0)}$ is the asymptotic variance matrix already present in Proposition 2, except that it is evaluated at $\Delta_0 = E[\Delta]$. Note also that the second order correction term is proportional to $Var[\xi]$, and is therefore zero in the absence of sampling randomness. When that happens, $\Delta = \Delta_0$ with probability one and the asymptotic variance of the estimator reduces to the leading term $A^{(0)}$, i.e., to the result in the fixed sampling case given in Proposition 2..

7.2 Randomly Spaced Sampling Intervals and Misspecified Microstructure Noise

Suppose now, as in Section 5, that the U 's are iid, have mean zero and variance a^2 , but are otherwise not necessarily Gaussian. We adopt the same approach as in Section 5, namely to express the estimator's properties in terms of deviations from the deterministic and Gaussian case. The additional correction terms in the asymptotic variance are given in the following result.

Theorem 3. *The asymptotic variance is given by*

$$AVAR_{true}(\hat{\sigma}^2, \hat{a}^2) = \left(A^{(0)} + Cum_4(U)B^{(0)} \right) + \epsilon^2 \left(A^{(2)} + Cum_4(U)B^{(2)} \right) + O(\epsilon^3) \quad (7.7)$$

where $A^{(0)}$ and $A^{(2)}$ are given in the statement of Theorem 2 and

$$B^{(0)} = \begin{pmatrix} 0 & 0 \\ 0 & \Delta_0 \end{pmatrix}$$

while

$$B^{(2)} = Var[\xi] \begin{pmatrix} B_{\sigma^2\sigma^2}^{(2)} & B_{\sigma^2a^2}^{(2)} \\ \bullet & B_{a^2a^2}^{(2)} \end{pmatrix}$$

$$B_{\sigma^2\sigma^2}^{(2)} = \frac{10\Delta_0^{3/2}\sigma^5}{(4a^2 + \Delta_0\sigma^2)^{5/2}} + \frac{4\Delta_0^2\sigma^6(16a^4 + 11a^2\Delta_0\sigma^2 + 2\Delta_0^2\sigma^4)}{(2a^2 + \Delta_0\sigma^2)^3(4a^2 + \Delta_0\sigma^2)^2}$$

$$B_{\sigma^2a^2}^{(2)} = \frac{-\Delta_0^2\sigma^4}{(2a^2 + \Delta_0\sigma^2)^3(4a^2 + \Delta_0\sigma^2)^{5/2}} \left(\sqrt{4a^2 + \Delta_0\sigma^2} (32a^6 + 64a^4\Delta_0\sigma^2 + 35a^2\Delta_0^2\sigma^4 + 6\Delta_0^3\sigma^6) \right. \\ \left. + \Delta_0^{1/2}\sigma (116a^6 + 126a^4\Delta_0\sigma^2 + 47a^2\Delta_0^2\sigma^4 + 6\Delta_0^3\sigma^6) \right)$$

$$B_{a^2a^2}^{(2)} = \frac{16a^8\Delta_0^{5/2}\sigma^3(13a^4 + 10a^2\Delta_0\sigma^2 + 2\Delta_0^2\sigma^4)}{(2a^2 + \Delta_0\sigma^2)^3(4a^2 + \Delta_0\sigma^2)^{5/2} \left(2a^2 + \sigma^2\Delta - \sqrt{\sigma^2\Delta(4a^2 + \sigma^2\Delta)} \right)^2}$$

The term $A^{(0)}$ is the base asymptotic variance of the estimator, already present with fixed sampling and Gaussian noise. The term $Cum_4(U)B^{(0)}$ is the correction due to the misspecification of the error distribution. These two terms are identical to those present in Theorem 1. The terms proportional to ϵ^2 are the further correction terms introduced by the randomness of the sampling. $A^{(2)}$ is the base correction term present even with Gaussian noise in Theorem 2, and $Cum_4(U)B^{(2)}$ is the further correction due to the sampling randomness. Both $A^{(2)}$ and $B^{(2)}$ are proportional to $Var[\xi]$ and hence vanish in the absence of sampling randomness.

8. Presence of a Drift Coefficient

What happens to our conclusions when the underlying X process has a drift? We shall see in this case that the presence of the drift does not alter our earlier conclusions. As a simple example, consider linear drift, i.e., replace (1.2) with

$$X_t = \mu t + \sigma W_t. \quad (8.1)$$

The contamination by market microstructure noise is as before: the observed process is given by (1.3).

As before, we first-difference to get the log-returns $Y_i = \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}} + U_{\tau_i} - U_{\tau_{i-1}}$. The likelihood function is now

$$\begin{aligned} \ln L(Y_N, \dots, Y_1 | \Delta_N, \dots, \Delta_1; \beta) &\equiv l(\sigma^2, a^2, \mu) \\ &= -\ln \det(\Sigma)/2 - N \ln(2\pi)/2 - (Y - \mu\Delta)' \Sigma^{-1} (Y - \mu\Delta)/2, \end{aligned}$$

where the covariance matrix is given in (7.2), and where $\Delta = (\Delta_1, \dots, \Delta_N)'$. If β denotes either σ^2 or a^2 , one obtains

$$\ddot{l}_{\mu\beta} = \Delta' \frac{\partial \Sigma^{-1}}{\partial \beta} (Y - \mu\Delta),$$

so that $E[\ddot{l}_{\mu\beta} | \Delta] = 0$ no matter whether the U 's are normally distributed or have another distribution with mean 0 and variance a^2 . In particular,

$$E[\ddot{l}_{\mu\beta}] = 0. \tag{8.2}$$

Now let $E[\ddot{l}]$ be the 3×3 matrix of expected second likelihood derivatives. Let $E[\ddot{l}] = -TE[\Delta]D + o(T)$. Similarly define $Cov(\dot{l}, \dot{l}) = TE[\Delta]S + o(T)$. As before, when the U 's have a normal distribution, $S = D$, and otherwise that is not the case. The asymptotic variance matrix of the estimators is of the form $AVAR = E[\Delta]D^{-1}SD^{-1}$.

Let D_{σ^2, a^2} be the corresponding 2×2 matrix when estimation is carried out on σ^2 and a^2 for known μ , and D_μ is the asymptotic information on μ for known σ^2 and a^2 . Similarly define S_{σ^2, a^2} and $AVAR_{\sigma^2, a^2}$. Since D is block diagonal by (8.2),

$$D = \begin{pmatrix} D_{\sigma^2, a^2} & 0 \\ 0' & D_\mu \end{pmatrix},$$

it follows that

$$D^{-1} = \begin{pmatrix} D_{\sigma^2, a^2}^{-1} & 0 \\ 0' & D_\mu^{-1} \end{pmatrix}.$$

Hence

$$AVAR(\hat{\sigma}^2, \hat{a}^2) = E[\Delta]D_{\sigma^2, a^2}^{-1}S_{\sigma^2, a^2}D_{\sigma^2, a^2}^{-1}. \tag{8.3}$$

The asymptotic variance of $(\hat{\sigma}^2, \hat{a}^2)$ is thus the same as if μ were known, in other words, as if $\mu = 0$, which is the case that we focused on in all the previous sections.

9. Serially Correlated Noise

We now examine what happens if we relax the assumption that the market microstructure noise is serially independent. Suppose that, instead of being iid with mean 0 and variance a^2 , the market microstructure noise follows

$$dU_t = -bU_t dt + cdZ_t$$

where $b > 0$, $c > 0$ and Z is a Brownian motion independent of W . $U_\Delta|U_0$ has a Gaussian distribution with mean $e^{-b\Delta}U_0$ and variance $\frac{c^2}{2b}(1 - e^{-2b\Delta})$. The unconditional mean and variance of U are 0 and $a^2 = \frac{c^2}{2b}$. The main consequence of this model is that the variance contributed by the noise to a log-return observed over an interval of time Δ is now of order $O(\Delta)$, that is of the same order as the variance of the efficient price process $\sigma^2\Delta$, instead of being of order $O(1)$ as previously. In other words, log-prices observed close together have very highly correlated noise terms. Because of this feature, this model for the microstructure noise would not be appropriate if the primary source of the noise consists of bid-ask bounces. In such a situation, the fact that a transaction is on the bid or ask side has little predictive power for the next transaction, or at least not enough to predict that two successive transactions are on the same side with very high probability. On the other hand, this model can better capture effects such as the gradual adjustment of prices in response to a shock such as a large trade. In practice, the noise term probably encompasses both of these examples, resulting in a situation where the variance contributed by the noise has both types of components, some of order $O(1)$, some of lower orders in Δ .

The observed log-returns take the form

$$\begin{aligned} Y_i &= \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}} + U_{\tau_i} - U_{\tau_{i-1}} \\ &= \sigma(W_{\tau_i} - W_{\tau_{i-1}}) + U_{\tau_i} - U_{\tau_{i-1}} \\ &\equiv w_i + u_i \end{aligned}$$

where the w'_i s are iid $N(0, \sigma^2\Delta)$ and the u'_i s are independent of the w'_i s, are Gaussian with mean zero and variance

$$E[u_i^2] = E[(U_{\tau_i} - U_{\tau_{i-1}})^2] = \frac{c^2(1 - e^{-b\Delta})}{b} = c^2\Delta + o(\Delta) \quad (9.1)$$

instead of $2a^2$.

In addition, the u'_i s are now serially correlated. In particular, we have

$$E[U_{\tau_i}U_{\tau_k}] = \frac{c^2(1 - e^{-b\Delta(i-k)})}{2b}$$

for $i \geq k$. If one ignores the presence of this type of serially correlated noise when estimating σ^2 , then:

Proposition 3. *In small samples (finite T), the RMSE of the estimator $\hat{\sigma}^2$ is given by*

$$\begin{aligned} RMSE[\hat{\sigma}^2] &= \left(\frac{c^4 (1 - e^{-b\Delta})^2}{b^2 \Delta^2} + \frac{c^4 (1 - e^{-b\Delta})^2 \left(\frac{T}{\Delta} e^{-2b\Delta} - 1 + e^{-2Tb} \right)}{T^2 b^2 (1 + e^{-b\Delta})^2} \right. \\ &\quad \left. + \frac{2}{T\Delta} \left(\sigma^2 \Delta + \frac{c^2 (1 - e^{-b\Delta})}{b} \right)^2 \right)^{1/2} \\ &= c^2 - \frac{bc^2}{2} \Delta + \frac{(\sigma^2 + c^2)^2 \Delta}{c^2 T} + O(\Delta^2) + O\left(\frac{1}{T^2}\right) \end{aligned} \quad (9.2)$$

so that for large T , starting from a value of c^2 in the limit where $\Delta \rightarrow 0$, increasing Δ first reduces $RMSE[\hat{\sigma}^2]$. Hence the optimal sampling frequency is finite.

Figure 4 displays the RMSE of the estimator as a function of Δ and T , with parameter values $\sigma = 0.15$, $b = 1$ and $c = 0.1$. With realistic parameter values, T must be quite large before $RMSE[\hat{\sigma}^2]$ exhibits a minimum in Δ : asymptotically in small Δ , this will only occur if

$$-\frac{bc^2}{2} + \frac{(\sigma^2 + c^2)^2}{c^2 T} < 0$$

that is,

$$T > \frac{2(\sigma^2 + c^2)^2}{bc^4}$$

or $T > 21.125$ years with these parameter values. This is due to the fact that the existence of a minimum in Δ comes from the bias component, which becomes predominant as T gets large. This is another way of seeing that this type of noise is not nearly as bad as iid noise for the purpose of inferring σ^2 from high frequency data. Recall from (9.1) that the variance of the noise is of the same order $O(\Delta)$ as the variance of the efficient price process. Thus log returns computed from transaction prices sampled close together are not subject to a lot of noise ($O(\Delta)$ vs. $O(1)$). Figure 4 shows the shape of the curve for two values of T , one too small for a minimum to occur, one large enough.

As for the rest of the analysis of the paper, the covariance matrix of the log-returns, $\gamma^2 V$ in (4.2), should be replaced by the matrix whose diagonal elements are

$$Var[Y_i^2] = E[w_i^2] + E[u_i^2] = \sigma^2 \Delta + \frac{c^2 (1 - e^{-b\Delta})}{b}$$

and off-diagonal elements $i > j$ are:

$$\begin{aligned}
cov(Y_i, Y_j) &= E[Y_i Y_j] = E[(w_i + u_i)(w_j + u_j)] \\
&= E[u_i u_j] = E[(U_{\tau_i} - U_{\tau_{i-1}})(U_{\tau_j} - U_{\tau_{j-1}})] \\
&= E[U_{\tau_i} U_{\tau_j}] - E[U_{\tau_i} U_{\tau_{j-1}}] - E[U_{\tau_{i-1}} U_{\tau_j}] + E[U_{\tau_{i-1}} U_{\tau_{j-1}}] \\
&= -\frac{c^2 (1 - e^{-b\Delta})^2 e^{-b\Delta(i-j-1)}}{2b}
\end{aligned}$$

Note that the theorems in the previous sections do not apply to this new situation because, having modified the matrix $\gamma^2 V$, the artificial “normal” distribution that assumes iid U 's that are $N(0, \alpha^2)$ would no longer use the correct second moment structure of the data. Thus the analysis would have to be repeated for this new scenario.

10. Noise Correlated with the Price Process

We have assumed so far that the U process was uncorrelated with the W process. Microstructure noise attributable to informational effects is likely to be correlated with the efficient price process, since it is generated by the response of market participants to information signals (i.e., to the efficient price process). This would be the case for instance in the bid-ask model with adverse selection of Glosten (1987). When the U process is no longer uncorrelated from the W process, the form of the variance matrix of the observed log-returns Y must be altered, replacing $\gamma^2 v_{ij}$ in (4.2) with

$$\begin{aligned}
cov(Y_i, Y_j) &= cov(\sigma(W_{\tau_i} - W_{\tau_{i-1}}) + U_{\tau_i} - U_{\tau_{i-1}}, \sigma(W_{\tau_j} - W_{\tau_{j-1}}) + U_{\tau_j} - U_{\tau_{j-1}}) \\
&= \sigma^2 \Delta \delta_{ij} + cov(\sigma(W_{\tau_i} - W_{\tau_{i-1}}), U_{\tau_j} - U_{\tau_{j-1}}) \\
&\quad + cov(\sigma(W_{\tau_j} - W_{\tau_{j-1}}), U_{\tau_i} - U_{\tau_{i-1}}) + cov(U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_j} - U_{\tau_{j-1}})
\end{aligned}$$

where δ_{ij} is the Kronecker symbol.

The small sample properties of the misspecified MLE for σ^2 , including its RMSE, can be obtained from

$$\begin{aligned}
E[\hat{\sigma}^2] &= \frac{1}{T} \sum_{i=1}^N E[Y_i^2] \\
Var[\hat{\sigma}^2] &= \frac{1}{T^2} \sum_{i=1}^N Var[Y_i^2] + \frac{2}{T^2} \sum_{i=1}^N \sum_{j=1}^{i-1} cov(Y_i^2, Y_j^2).
\end{aligned}$$

Specific expressions for all these quantities depend upon the assumptions of the particular structural model under consideration: for instance, in the Glosten (1987) model (see his Proposition 6), the U 's

remain stationary, the transaction noise U_{τ_i} is uncorrelated with the return noise during the previous observation period, i.e., $U_{\tau_{i-1}} - U_{\tau_{i-2}}$, and the efficient return $\sigma(W_{\tau_i} - W_{\tau_{i-1}})$ is also uncorrelated with the transaction noises $U_{\tau_{i+1}}$ and $U_{\tau_{i-2}}$.

With these in hand, the analysis can then proceed as above. The same caveat as in serially correlated U case applies: having modified the matrix $\gamma^2 V$, the artificial “normal” distribution would no longer use the correct second moment structure of the data. Thus the theorems should be modified accordingly.

11. Conclusions

Our first finding in the paper is that there are situations where the presence of market microstructure noise makes it optimal to sample less often than would otherwise be the case in the absence of noise. We then addressed the issue of what to do about it, and showed that modelling the noise term explicitly restores the first order statistical effect that sampling as often as possible is optimal. We also demonstrated that this is true even if one misspecifies the assumed distribution of the noise term. If the econometrician assumes that the noise terms are normally distributed when in fact they are not, not only is it still optimal to sample as often as possible (unlike the result when no allowance is made for the presence of noise), but the estimator has the same asymptotic variance as if the noise distribution had been correctly specified.

We purposefully adopted the simplest possible setup to demonstrate that our results are not driven by the complexity of the model, but rather are likely to be genuine features facing the econometrics of high frequency data. Our robustness results suggest that attempts to incorporate the market microstructure noise when estimating continuous-time models based on high frequency data should have beneficial effects.

Appendix A: Proof of Proposition 1

The estimator (2.2) has the following mean

$$E[\hat{\sigma}^2] = \frac{1}{T} \sum_{i=1}^N E[Y_i^2] = \frac{N(\sigma^2\Delta + 2a^2)}{T} = \sigma^2 + \frac{2a^2}{\Delta}. \quad (\text{A.1})$$

Given the form of the bias, one would in fact want to select the largest Δ possible to minimize the bias (as opposed to the smallest one as in the no-noise case of Section 2).

The estimator's variance is

$$\text{Var}[\hat{\sigma}^2] = \frac{1}{T^2} \text{Var} \left[\sum_{i=1}^N Y_i^2 \right] = \frac{1}{T^2} \left(\sum_{i=1}^N \text{Var}[Y_i^2] + 2 \sum_{i=2}^N \text{cov}(Y_i^2, Y_{i-1}^2) \right).$$

Since the Y_i 's are normal with mean zero, $\text{Var}[Y_i^2] = 2\text{Var}[Y_i]^2 = 2\gamma^4(1 + \eta^2)^2$ and

$$\begin{aligned} \text{cov}(Y_i^2, Y_{i-1}^2) &= \text{cov}(\varepsilon_i^2 + 2\eta\varepsilon_i\varepsilon_{i-1} + \eta^2\varepsilon_{i-1}^2, \varepsilon_{i-1}^2 + 2\eta\varepsilon_{i-1}\varepsilon_{i-2} + \eta^2\varepsilon_{i-2}^2) \\ &= \eta^2 \text{Var}[\varepsilon_{i-1}^2] \\ &= 2\eta^2\gamma^4 \end{aligned}$$

so that

$$\begin{aligned} \text{Var}[\hat{\sigma}^2] &= \frac{1}{T^2} \{2N\gamma^4(1 + \eta^2)^2 + 4(N-1)\eta^2\gamma^4\} \\ &= \frac{1}{T^2} \{2N\gamma^4(1 + 4\eta^2 + \eta^4)\} - \frac{1}{T^2} \{4\eta^2\gamma^4\} \\ &= \frac{2\gamma^4(1 + 4\eta^2 + \eta^4)}{T\Delta} - \frac{4\eta^2\gamma^4}{T^2} \\ &= \frac{2(\sigma^4\Delta^2 + 4\sigma^2\Delta a^2 + 6a^4)}{T\Delta} - \frac{4a^4}{T^2} \end{aligned} \quad (\text{A.2})$$

Note that these are exact small sample expressions, valid for all (T, Δ) . Asymptotically in T , $\text{Var}[\hat{\sigma}^2] \rightarrow 0$, and hence the RMSE of the estimator is dominated by the bias term.

In finite samples, the expression for the RMSE given in (3.6) follows from those for the expected value and variance (A.1) and (A.2). The optimal value Δ^* of the sampling interval given in (3.7) is obtained by minimizing the RMSE (3.6) over Δ .

Appendix B: Proof of Proposition 2

The partial derivatives of the log-likelihood function (4.1) have the form

$$i_\eta = -\frac{1}{2} \frac{\partial \ln \det(V)}{\partial \eta} - \frac{1}{2\gamma^2} Y' \frac{\partial V^{-1}}{\partial \eta} Y, \quad (\text{B.1})$$

and

$$i_{\gamma^2} = -\frac{N}{2\gamma^2} + \frac{1}{2\gamma^4} Y' V^{-1} Y. \quad (\text{B.2})$$

so that the MLE for γ^2 is

$$\hat{\gamma}^2 = \frac{1}{N} Y' V^{-1} Y. \quad (\text{B.3})$$

At the true parameters, the expected value of the score vector is zero: $E[\dot{l}_\eta] = E[\dot{l}_{\gamma^2}] = 0$. Hence it follows from (B.1) that

$$E\left[Y' \frac{\partial V^{-1}}{\partial \eta} Y\right] = -\gamma^2 \frac{\partial \ln \det(V)}{\partial \eta} = -\gamma^2 \frac{2\eta(1 - (1+N)\eta^{2N} + N\eta^{2(1+N)})}{(1-\eta^2)(1-\eta^{2(1+N)})}$$

thus as $N \rightarrow \infty$

$$E\left[Y' \frac{\partial V^{-1}}{\partial \eta} Y\right] = -\frac{2\eta\gamma^2}{(1-\eta^2)} + o(1).$$

Similarly, it follows from (B.2) that

$$E[Y'V^{-1}Y] = N\gamma^2.$$

Turning now to Fisher's information, we have

$$E[-\ddot{l}_{\gamma^2\gamma^2}] = -\frac{N}{2\gamma^4} + \frac{1}{\gamma^6} E[Y'V^{-1}Y] = \frac{N}{2\gamma^4}, \quad (\text{B.4})$$

whence the asymptotic variance of $T^{1/2}(\hat{\gamma}^2 - \gamma^2)$ is $2\gamma^4\Delta$. We also have that

$$E[-\ddot{l}_{\gamma^2\eta}] = \frac{1}{2\gamma^4} E\left[Y' \frac{\partial V^{-1}}{\partial \eta} Y\right] = -\frac{\eta}{\gamma^2(1-\eta^2)} + o(1), \quad (\text{B.5})$$

whence the asymptotic covariance of $T^{1/2}(\hat{\gamma}^2 - \gamma^2)$ and $T^{1/2}(\hat{\eta} - \eta)$ is zero.

To evaluate $E[-\ddot{l}_{\eta\eta}]$, we compute

$$E[-\ddot{l}_{\eta\eta}] = \frac{1}{2} \frac{\partial^2 \ln \det(V)}{\partial \eta^2} + \frac{1}{2\gamma^2} E\left[Y' \frac{\partial^2 V^{-1}}{\partial \eta^2} Y\right] \quad (\text{B.6})$$

and evaluate both terms. For the first term in (B.6), we have from (4.3):

$$\begin{aligned} \frac{\partial^2 \ln \det(V)}{\partial \eta^2} &= \frac{1}{(1-\eta^{2+2N})^2} \left\{ \frac{2(1+\eta^2 + \eta^{2+2N}(1-3\eta^2))(1-\eta^{2N})}{(1-\eta^2)^2} \right. \\ &\quad \left. - 2N\eta^{2N}(3+\eta^{2+2N}) - 4N^2\eta^{2N} \right\} \\ &= \frac{2(1+\eta^2)}{(1-\eta^2)^2} + o(1) \end{aligned} \quad (\text{B.7})$$

For the second term, we have for any non-random $N \times N$ matrix Q :

$$\begin{aligned} E[Y'QY] &= E[\text{Tr}[Y'QY]] = E[\text{Tr}[QYY']] = \text{Tr}[E[QYY']] \\ &= \text{Tr}[QE[YY']] = \text{Tr}[Q\gamma^2V] = \gamma^2\text{Tr}[QV] \end{aligned}$$

where Tr denotes the matrix trace, which satisfies $\text{Tr}[AB] = \text{Tr}[BA]$. Therefore

$$\begin{aligned} E\left[Y' \frac{\partial^2 V^{-1}}{\partial \eta^2} Y\right] &= \gamma^2 \text{Tr}\left[\frac{\partial^2 V^{-1}}{\partial \eta^2} V\right] = \gamma^2 \left(\sum_{i=1}^N \sum_{j=1}^N \frac{\partial^2 v^{ij}}{\partial \eta^2} v_{ij} \right) \\ &= \gamma^2 \left(\sum_{i=1}^N \frac{\partial^2 v^{ii}}{\partial \eta^2} (1+\eta^2) + \sum_{i=1}^{N-1} \frac{\partial^2 v^{i,i+1}}{\partial \eta^2} \eta + \sum_{i=2}^N \frac{\partial^2 v^{i,i-1}}{\partial \eta^2} \eta \right) \\ &= \frac{\gamma^2}{(1-\eta^{2+2N})^2} \left\{ -\frac{4(1+2\eta^2 + \eta^{2+2N}(1-4\eta^2))(1-\eta^{2N})}{(1-\eta^2)^2} \right. \\ &\quad \left. + \frac{2N(1+\eta^{2N}(6-6\eta^2 + 2\eta^{2+2N} - 3\eta^{4+2N}))}{(1-\eta^2)} + 8N^2\eta^{2N} \right\} \\ &= \frac{2\gamma^2 N}{(1-\eta^2)} + o(N) \end{aligned} \quad (\text{B.8})$$

Combining (B.7) and (B.8) into (B.6), it follows that

$$E \left[-\ddot{l}_{\eta\eta} \right] = \frac{1}{2} \frac{\partial^2 \ln \det(V_N)}{\partial \eta^2} + \frac{1}{2\gamma^2} E \left[Y' \frac{\partial^2 V_N^{-1}}{\partial \eta^2} Y \right] \underset{N \rightarrow \infty}{\sim} \frac{N}{(1-\eta^2)} + o(N) \quad (\text{B.9})$$

In light of that and (B.5), the asymptotic variance of $T^{1/2}(\hat{\eta} - \eta)$ is the same as in the γ^2 known case, that is, $(1-\eta^2)\Delta$ (which of course confirms the result of Durbin (1959) for this parameter).

We can now retrieve the asymptotic covariance matrix for the original parameters (σ^2, a^2) from that of the parameters (γ^2, η) . This follows from the delta method applied to the change of variable (3.2)-(3.3):

$$\begin{pmatrix} \sigma^2 \\ a^2 \end{pmatrix} = f(\gamma^2, \eta) = \begin{pmatrix} \Delta^{-1}\gamma^2(1+\eta)^2 \\ -\gamma^2\eta \end{pmatrix}. \quad (\text{B.10})$$

Hence

$$T^{1/2} \left(\begin{pmatrix} \hat{\sigma}^2 \\ \hat{a}^2 \end{pmatrix} - \begin{pmatrix} \sigma^2 \\ a^2 \end{pmatrix} \right) \underset{T \rightarrow \infty}{\rightarrow} N(0, AVAR(\hat{\sigma}^2, \hat{a}^2))$$

where

$$\begin{aligned} AVAR(\hat{\sigma}^2, \hat{a}^2) &= \nabla f(\gamma^2, \eta) \cdot AVAR(\hat{\gamma}^2, \hat{\eta}) \cdot \nabla f(\gamma^2, \eta)' \\ &= \begin{pmatrix} \frac{(1+\eta)^2}{\Delta} & \frac{2\gamma^2(1+\eta)}{\Delta} \\ -\eta & -\gamma^2 \end{pmatrix} \begin{pmatrix} 2\gamma^4\Delta & 0 \\ 0 & (1-\eta^2)\Delta \end{pmatrix} \begin{pmatrix} \frac{(1+\eta)^2}{\Delta} & -\eta \\ \frac{2\gamma^2(1+\eta)}{\Delta} & -\gamma^2 \end{pmatrix} \\ &= \begin{pmatrix} 4\sqrt{\sigma^6\Delta(4a^2 + \sigma^2\Delta)} + 2\sigma^4\Delta & -\sigma^2\Delta h(\Delta, \sigma^2, a^2) \\ \bullet & \frac{\Delta}{2}(2a^2 + \sigma^2\Delta) h(\Delta, \sigma^2, a^2) \end{pmatrix}. \end{aligned}$$

Appendix C: Proof of Theorem 1

Consider two generic symmetric $N \times N$ matrices $[\nu^{i,j}]$ and $[\omega^{i,j}]$. We are interested in expressions of the form

$$\begin{aligned} \sum_{i,j,k,l:M=m+1} (-1)^s \nu^{i,j} \omega^{k,l} &= \sum_{h=1}^{N-1} \sum_{i,j,k,l:m=h, M=h+1} (-1)^s \nu^{i,j} \omega^{k,l} \\ &= \sum_{h=1}^{N-1} \sum_{r=1}^3 \sum_{i,j,k,l:m=h, M=h+1, s=r} (-1)^r \nu^{i,j} \omega^{k,l} \\ &= \sum_{h=1}^{N-1} \left\{ -2\nu^{h,h+1} \omega^{h+1,h+1} - 2\nu^{h+1,h+1} \omega^{h,h+1} \right. \\ &\quad \left. + \nu^{h,h} \omega^{h+1,h+1} + \nu^{h+1,h+1} \omega^{h,h} + 4\nu^{h,h+1} \omega^{h,h+1} \right. \\ &\quad \left. - 2\nu^{h+1,h} \omega^{h,h} - 2\nu^{h,h} \omega^{h+1,h} \right\} \end{aligned} \quad (\text{C.1})$$

It follows that if we set

$$\Upsilon(\nu, \omega) = \sum_{i,j,k,l=1}^N \nu^{i,j} \omega^{k,l} Cum_{\text{true}}(Y_i, Y_j, Y_k, Y_l) \quad (\text{C.2})$$

then $\Upsilon(\nu, \omega) = Cum_4(U) \Psi(\nu, \omega)$ where

$$\begin{aligned} \psi(\nu, \omega) &= 2 \sum_{h=1}^N \nu^{h,h} \omega^{h,h} + \sum_{h=1}^{N-1} \left\{ -2\nu^{h,h+1} \omega^{h+1,h+1} - 2\nu^{h+1,h+1} \omega^{h,h+1} \right. \\ &\quad \left. + \nu^{h,h} \omega^{h+1,h+1} + \nu^{h+1,h+1} \omega^{h,h} + 4\nu^{h,h+1} \omega^{h,h+1} \right. \\ &\quad \left. - 2\nu^{h+1,h} \omega^{h,h} - 2\nu^{h,h} \omega^{h+1,h} \right\} \end{aligned} \quad (\text{C.3})$$

If the two matrices $[\nu^{i,j}]$ and $[\omega^{i,j}]$ satisfy the following reversibility property: $\nu^{N+1-i,N+1-j} = \nu^{i,j}$ and $\omega^{N+1-i,N+1-j} = \omega^{i,j}$ (so long as one is within the index set), then (C.3) simplifies to:

$$\begin{aligned} \psi(\nu, \omega) &= 2 \sum_{h=1}^N \nu^{h,h} \omega^{h,h} + \sum_{h=1}^{N-1} \{ -4\nu^{h,h+1} \omega^{h+1,h+1} - 4\nu^{h+1,h+1} \omega^{h,h+1} \\ &\quad + 2\nu^{h,h} \omega^{h+1,h+1} + 4\nu^{h,h+1} \omega^{h,h+1} \} \end{aligned}$$

This is the case for V^{-1} and its derivative $\partial V^{-1}/\partial\eta$, as can be seen from the expression for $v^{i,j}$ given in (4.4), and consequently for $\partial v^{i,j}/\partial\eta$.

If we wish to compute the sums in equations (5.4), (5.5), and (5.6), therefore, we need, respectively, to find the three quantities $\psi(\partial v/\partial\eta, v)$, $\psi(\partial v/\partial\eta, \partial v/\partial\eta)$, and $\psi(v, v)$ respectively. All are of order $O(N)$, and only the first term is needed. Replacing the terms $v^{i,j}$ and $\partial v^{i,j}/\partial\eta$ by their expressions from (4.4), we obtain:

$$\begin{aligned} \psi(v, v) &= \frac{2}{(1+\eta^2)(1-\eta)^3(1-\eta^{2(1+N)})^2} \left\{ -(1+\eta)(1-\eta^{2N}) \left(1 + 2\eta^2 + 2\eta^{2(1+N)} + \eta^{2(2+N)} \right) \right. \\ &\quad \left. + N(1-\eta)(1+\eta^2) \left(2 + \eta^{2N} + \eta^{2(1+N)} + 6\eta^{1+2N} + 2\eta^{2+4N} \right) \right\} \\ &= \frac{4N}{(1-\eta)^2} + o(N) \end{aligned} \tag{C.4}$$

$$\begin{aligned} \psi\left(\frac{\partial v}{\partial\eta}, v\right) &= \frac{2(O(1) + 2N(1-\eta)(1+\eta^2)\eta(1+\eta^2 + O(\eta^{2N})) + N^2O(\eta^{2N}))}{\eta(1-\eta)^4(1+\eta^2)^2(1-\eta^{2(1+N)})^3} \\ &= \frac{4N}{(1-\eta)^3} + o(N) \end{aligned} \tag{C.5}$$

$$\begin{aligned} \psi\left(\frac{\partial v}{\partial\eta}, \frac{\partial v}{\partial\eta}\right) &= \frac{4(O(1) + 3N(1-\eta^4)\eta^2((1+\eta^2)^2 + O(\eta^{2N})) + N^2O(\eta^{2N}) + N^3O(\eta^{2N}))}{3\eta^2(1+\eta)(1+\eta^2)^3(1-\eta)^5(1-\eta^{2(1+N)})^4} \\ &= \frac{4N}{(1-\eta)^4} + o(N) \end{aligned} \tag{C.6}$$

The asymptotic variance of the estimator $(\hat{\gamma}^2, \hat{\eta})$ obtained by maximizing the (incorrectly-specified) log-likelihood (4.1) that assumes Gaussianity of the U 's is given by

$$AVAR_{\text{true}}(\hat{\gamma}^2, \hat{\eta}) = \Delta (D'S^{-1}D)^{-1}$$

where, from (B.4), (B.5) and (B.9) we have

$$\begin{aligned} D &= D' = -\frac{1}{N}E_{\text{true}}[\ddot{l}] = -\frac{1}{N}E_{\text{normal}}[\ddot{l}] = \frac{1}{N}E_{\text{normal}}[\ddot{l}l'] \\ &= \begin{pmatrix} \frac{1}{2\gamma^4} & -\frac{\eta}{N\gamma^2(1-\eta^2)} + o\left(\frac{1}{N}\right) \\ \bullet & \frac{1}{(1-\eta^2)} + o(1) \end{pmatrix} \end{aligned} \tag{C.7}$$

and, in light of (5.4), (5.5), and (5.6),

$$S = \frac{1}{N}E_{\text{true}}[\ddot{l}l'] = -\frac{1}{N}E_{\text{true}}[\ddot{l}] + Cum_4(U) \Psi = D + Cum_4(U) \Psi \tag{C.8}$$

where

$$\begin{aligned}
\Psi &= \frac{1}{4N} \begin{pmatrix} \frac{1}{\gamma^8} \psi(v, v) & -\frac{1}{\gamma^6} \psi\left(\frac{\partial v}{\partial \eta}, v\right) \\ \bullet & \frac{1}{\gamma^4} \psi\left(\frac{\partial v}{\partial \eta}, \frac{\partial v}{\partial \eta}\right) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\gamma^8(1-\eta)^2} + o(1) & \frac{-1}{\gamma^6(1-\eta)^3} + o(1) \\ \bullet & \frac{1}{\gamma^4(1-\eta)^4} + o(1) \end{pmatrix}
\end{aligned} \tag{C.9}$$

from the expressions just computed. It follows that

$$\begin{aligned}
AVAR_{\text{true}}(\hat{\gamma}^2, \hat{\eta}) &= \Delta \left(D(D + Cum_4(U) \Psi)^{-1} D \right)^{-1} \\
&= \Delta \left(D(Id + Cum_4(U) D^{-1} \Psi)^{-1} \right)^{-1} \\
&= \Delta (Id + Cum_4(U) D^{-1} \Psi) D^{-1} \\
&= \Delta (Id + Cum_4(U) D^{-1} \Psi) D^{-1} \\
&= AVAR_{\text{normal}}(\hat{\gamma}^2, \hat{\eta}) + \Delta Cum_4(U) D^{-1} \Psi D^{-1}
\end{aligned}$$

where Id denotes the identity matrix and

$$AVAR_{\text{normal}}(\hat{\gamma}^2, \hat{\eta}) = \begin{pmatrix} 2\gamma^4 \Delta & 0 \\ 0 & (1 - \eta^2) \Delta \end{pmatrix}, \quad D^{-1} \Psi D^{-1} = \begin{pmatrix} \frac{4}{(1-\eta)^2} & \frac{-2(1+\eta)}{\gamma^2(1-\eta)^2} \\ \bullet & \frac{(1+\eta)^2}{\gamma^4(1-\eta)^2} \end{pmatrix}$$

so that

$$AVAR_{\text{true}}(\hat{\gamma}^2, \hat{\eta}) = \Delta \begin{pmatrix} 2\gamma^4 & 0 \\ 0 & (1 - \eta^2) \end{pmatrix} + \Delta Cum_4(U) \begin{pmatrix} \frac{4}{(1-\eta)^2} & \frac{-2(1+\eta)}{\gamma^2(1-\eta)^2} \\ \bullet & \frac{(1+\eta)^2}{\gamma^4(1-\eta)^2} \end{pmatrix}.$$

By applying the delta method as in the previous section, we now recover the asymptotic variance of the estimates of the original parameters

$$\begin{aligned}
AVAR_{\text{true}}(\hat{\sigma}^2, \hat{a}^2) &= \nabla f(\gamma^2, \eta) \cdot AVAR_{\text{true}}(\hat{\gamma}^2, \hat{\eta}) \cdot \nabla f(\gamma^2, \eta)' \\
&= \begin{pmatrix} 4\sqrt{\sigma^6 \Delta (4a^2 + \sigma^2 \Delta)} + 2\sigma^4 \Delta & -\sigma^2 \Delta h(\Delta, \sigma^2, a^2) \\ \bullet & \frac{\Delta}{2} (2a^2 + \sigma^2 \Delta) h(\Delta, \sigma^2, a^2) + \Delta Cum_4(U) \end{pmatrix}.
\end{aligned}$$

Appendix D: Derivations for Section 6

To see (6.4), let “orig” (D.7) denote parametrization in (and differentiation with respect to) the original parameters σ^2 and a^2 , while “transf” denotes parametrization and differentiation in γ^2 and η , and f_{inv} denotes the inverse of the change of variable function defined in (B.10), namely

$$\begin{pmatrix} \gamma^2 \\ \eta \end{pmatrix} = f_{\text{inv}}(\sigma^2, \alpha^2) = \begin{pmatrix} \frac{1}{2} \left\{ 2a^2 + \sigma^2 \Delta + \sqrt{\sigma^2 \Delta (4a^2 + \sigma^2 \Delta)} \right\} \\ \frac{1}{2a^2} \left\{ -2a^2 - \sigma^2 \Delta + \sqrt{\sigma^2 \Delta (4a^2 + \sigma^2 \Delta)} \right\} \end{pmatrix}. \tag{D.1}$$

and ∇f_{inv} its Jacobian matrix. Then, from $\dot{l}_{\text{orig}} = \nabla f_{\text{inv}}(\sigma^2, \alpha^2)' \cdot \dot{l}_{\text{transf}}$, we have

$$\ddot{l}_{\text{orig}} = \nabla f_{\text{inv}}(\sigma^2, \alpha^2)' \cdot \ddot{l}_{\text{transf}} \cdot \nabla f_{\text{inv}}(\sigma^2, \alpha^2) + H[\dot{l}_{\text{transf}}]$$

where $H[\dot{l}_{\text{transf}}]$ is a 2×2 matrix whose terms are linear in \dot{l}_{transf} and the second partial derivatives of f_{inv} . Now $E_{\text{true}}[\dot{l}_{\text{orig}}] = E_{\text{true}}[\dot{l}_{\text{transf}}] = 0$, and so $E_{\text{true}}[H[\dot{l}_{\text{transf}}]] = 0$ from which it follows that

$$\begin{aligned} D_{\text{orig}} &= N^{-1}E_{\text{true}}[-\ddot{l}_{\text{orig}}] \\ &= \nabla f_{\text{inv}}(\sigma^2, \alpha^2)' \cdot D_{\text{transf}} \cdot \nabla f_{\text{inv}}(\sigma^2, \alpha^2) \\ &= \left(\begin{array}{c} \frac{\Delta^{1/2}(2a^2 + \sigma^2\Delta)}{2\sigma^3(4a^2 + \sigma^2\Delta)^{3/2}} \\ \bullet \\ \frac{1}{2a^4} \left(1 - \frac{\Delta^{1/2}\sigma(6a^2 + \sigma^2\Delta)}{(4a^2 + \sigma^2\Delta)^{3/2}} \right) \end{array} \right) + o(1) \end{aligned} \quad (\text{D.2})$$

with $D_{\text{transf}} = N^{-1}E_{\text{true}}[-\ddot{l}_{\text{transf}}]$ given in (C.7). Similarly, $\dot{l}_{\text{orig}}\dot{l}'_{\text{orig}} = \nabla f_{\text{inv}}(\sigma^2, \alpha^2)' \cdot \dot{l}_{\text{transf}}\dot{l}'_{\text{transf}} \cdot \nabla f_{\text{inv}}(\sigma^2, \alpha^2)$ and so

$$\begin{aligned} S_{\text{orig}} &= \nabla f_{\text{inv}}(\sigma^2, \alpha^2)' \cdot S_{\text{transf}} \cdot \nabla f_{\text{inv}}(\sigma^2, \alpha^2) \\ &= \nabla f_{\text{inv}}(\sigma^2, \alpha^2)' \cdot (D_{\text{transf}} + \text{Cum}_4(U) \Psi) \cdot \nabla f_{\text{inv}}(\sigma^2, \alpha^2) \\ &= D_{\text{orig}} + \text{Cum}_4(U) \nabla f_{\text{inv}}(\sigma^2, \alpha^2)' \cdot \Psi \cdot \nabla f_{\text{inv}}(\sigma^2, \alpha^2) \end{aligned} \quad (\text{D.3})$$

with the second equality following from the expression for S_{transf} given in (C.8).

To complete the calculation, note from (C.9) that

$$\Psi = g_{\text{transf}} \cdot g'_{\text{transf}} + o(1),$$

where

$$g_{\text{transf}} = \begin{pmatrix} \gamma^{-4} (1 - \eta)^{-1} \\ -\gamma^{-2} (1 - \eta)^{-2} \end{pmatrix}.$$

Thus

$$\nabla f_{\text{inv}}(\sigma^2, \alpha^2)' \cdot \Psi \cdot \nabla f_{\text{inv}}(\sigma^2, \alpha^2) = g_{\text{orig}} \cdot g'_{\text{orig}} + o(1), \quad (\text{D.4})$$

where

$$g = g_{\text{orig}} = \nabla f_{\text{inv}}(\sigma^2, \alpha^2)' \cdot g_{\text{transf}} = \begin{pmatrix} \frac{\Delta^{1/2}}{\sigma(4a^2 + \sigma^2\Delta)^{3/2}} \\ \frac{1}{2a^4} \left(1 - \frac{\Delta^{1/2}\sigma(6a^2 + \sigma^2\Delta)}{(4a^2 + \sigma^2\Delta)^{3/2}} \right) \end{pmatrix} \quad (\text{D.5})$$

which is the result (6.5). Inserting (D.4) into (D.3) yields the result (6.4).

For the profile likelihood λ , let $\hat{a}_{\sigma^2}^2$ denote the maximizer of $l(\sigma^2, a^2)$ for given σ^2 . Thus by definition $\lambda(\sigma^2) = l(\sigma^2, \hat{a}_{\sigma^2}^2)$. From now on, all differentiation takes place with respect to the original parameters, and we will omit the subscript ‘‘orig’’ in what follows. Since $0 = \dot{l}_{a^2}(\sigma^2, \hat{a}_{\sigma^2}^2)$, it follows that

$$\begin{aligned} 0 &= \frac{\partial}{\partial \sigma^2} \dot{l}_{a^2}(\sigma^2, \hat{a}_{\sigma^2}^2) \\ &= \ddot{l}_{\sigma^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2) + \ddot{l}_{a^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2) \frac{\partial \hat{a}_{\sigma^2}^2}{\partial \sigma^2}, \end{aligned}$$

so that

$$\frac{\partial \hat{a}_{\sigma^2}^2}{\partial \sigma^2} = -\frac{\ddot{l}_{\sigma^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2)}{\ddot{l}_{a^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2)} \quad (\text{D.6})$$

The profile score then follows

$$\dot{\lambda}(\sigma^2) = \dot{l}_{\sigma^2}(\sigma^2, \hat{a}_{\sigma^2}^2) + \dot{l}_{a^2}(\sigma^2, \hat{a}_{\sigma^2}^2) \frac{\partial \hat{a}_{\sigma^2}^2}{\partial \sigma^2} \quad (\text{D.7})$$

so that at the true value of (σ^2, a^2) ,

$$\dot{\lambda}(\sigma^2) = \dot{l}_{\sigma^2}(\sigma^2, a^2) - \frac{E_{\text{true}}[\ddot{l}_{\sigma^2 a^2}]}{E_{\text{true}}[\ddot{l}_{a^2 a^2}]} \dot{l}_{a^2}(\sigma^2, a^2) + O_p(1), \quad (\text{D.8})$$

since $\hat{a}^2 = a^2 + O_p(N^{-1/2})$ and

$$\begin{aligned} \Delta \ddot{l}_{\sigma^2 a^2} &\equiv N^{-1} \ddot{l}_{\sigma^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2) - N^{-1} E_{\text{true}}[\ddot{l}_{\sigma^2 a^2}] = O_p(N^{-1/2}) \\ \Delta \ddot{l}_{a^2 a^2} &\equiv N^{-1} \ddot{l}_{a^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2) - N^{-1} E_{\text{true}}[\ddot{l}_{a^2 a^2}] = O_p(N^{-1/2}) \end{aligned}$$

as sums of random variables with expected value zero, so that

$$\begin{aligned} -\frac{\partial \hat{a}_{\sigma^2}^2}{\partial \sigma^2} &= \frac{N^{-1} \ddot{l}_{\sigma^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2)}{N^{-1} \ddot{l}_{a^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2)} \\ &= \frac{N^{-1} E_{\text{true}}[\ddot{l}_{\sigma^2 a^2}] + \Delta \ddot{l}_{\sigma^2 a^2}}{N^{-1} E_{\text{true}}[\ddot{l}_{a^2 a^2}] + \Delta \ddot{l}_{a^2 a^2}} \\ &= \frac{E_{\text{true}}[\ddot{l}_{\sigma^2 a^2}]}{E_{\text{true}}[\ddot{l}_{a^2 a^2}]} + \left(\frac{\Delta \ddot{l}_{\sigma^2 a^2}}{E_{\text{true}}[\ddot{l}_{a^2 a^2}]} - \frac{\Delta \ddot{l}_{a^2 a^2}}{E_{\text{true}}[\ddot{l}_{a^2 a^2}]} \right) + o_p(N^{-1/2}) \\ &= \frac{E_{\text{true}}[\ddot{l}_{\sigma^2 a^2}]}{E_{\text{true}}[\ddot{l}_{a^2 a^2}]} + O_p(N^{-1/2}) \end{aligned}$$

while

$$\dot{l}_{a^2}(\sigma^2, a^2) = O_p(N^{1/2})$$

also as a sum of random variables with expected value zero.

Therefore

$$\begin{aligned} E_{\text{true}}[\dot{\lambda}(\sigma^2)] &= E_{\text{true}}[\dot{l}_{\sigma^2}(\sigma^2, a^2)] - \frac{E_{\text{true}}[\ddot{l}_{\sigma^2 a^2}]}{E_{\text{true}}[\ddot{l}_{a^2 a^2}]} E_{\text{true}}[\dot{l}_{a^2}(\sigma^2, a^2)] + O(1) \\ &= O(1) \end{aligned}$$

since $E_{\text{true}}[\dot{l}_{\sigma^2}(\sigma^2, a^2)] = E_{\text{true}}[\dot{l}_{a^2}(\sigma^2, a^2)] = 0$. In particular, $E_{\text{true}}[\dot{\lambda}(\sigma^2)] = o(N)$ as claimed.

Further differentiating (D.7), one obtains

$$\begin{aligned} \ddot{\lambda}(\sigma^2) &= \ddot{l}_{\sigma^2 \sigma^2}(\sigma^2, \hat{a}_{\sigma^2}^2) + \ddot{l}_{a^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2) \left(\frac{\partial \hat{a}_{\sigma^2}^2}{\partial \sigma^2} \right)^2 \\ &\quad + 2 \ddot{l}_{\sigma^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2) \frac{\partial \hat{a}_{\sigma^2}^2}{\partial \sigma^2} + \dot{l}_{a^2}(\sigma^2, \hat{a}_{\sigma^2}^2) \frac{\partial^2 \hat{a}_{\sigma^2}^2}{\partial^2 \sigma^2} \\ &= \ddot{l}_{\sigma^2 \sigma^2}(\sigma^2, \hat{a}_{\sigma^2}^2) - \frac{\ddot{l}_{\sigma^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2)^2}{\ddot{l}_{a^2 a^2}(\sigma^2, \hat{a}_{\sigma^2}^2)} + \dot{l}_{a^2}(\sigma^2, \hat{a}_{\sigma^2}^2) \frac{\partial^2 \hat{a}_{\sigma^2}^2}{\partial^2 \sigma^2} \end{aligned}$$

from (D.6). Evaluated at $\sigma^2 = \hat{\sigma}^2$, one gets $\hat{a}_{\sigma^2}^2 = \hat{a}^2$ and $\dot{l}_{a^2}(\hat{\sigma}^2, \hat{a}^2) = 0$, and so

$$\begin{aligned} \ddot{\lambda}(\hat{\sigma}^2) &= \ddot{l}_{\sigma^2 \sigma^2}(\hat{\sigma}^2, \hat{a}^2) - \frac{\ddot{l}_{\sigma^2 a^2}(\hat{\sigma}^2, \hat{a}^2)^2}{\ddot{l}_{a^2 a^2}(\hat{\sigma}^2, \hat{a}^2)} \\ &= \frac{1}{\left[\ddot{l}(\hat{\sigma}^2, \hat{a}^2)^{-1} \right]_{\sigma^2 \sigma^2}} \end{aligned} \quad (\text{D.9})$$

where $\left[\ddot{l}(\hat{\sigma}^2, \hat{a}^2)^{-1} \right]_{\sigma^2 \sigma^2}$ is the upper left element of the matrix $\ddot{l}(\hat{\sigma}^2, \hat{a}^2)^{-1}$. Thus (6.7) is valid.

Alternatively, we can see that the profile likelihood λ satisfies the Bartlett identity to first order, i.e., (6.8). Note that by (D.8),

$$\begin{aligned}
N^{-1}E_{\text{true}}[\dot{\lambda}(\sigma^2)^2] &= N^{-1}E_{\text{true}} \left[\left(i_{\sigma^2}(\sigma^2, a^2) - \frac{E_{\text{true}}[\ddot{l}_{\sigma^2 a^2}]}{E_{\text{true}}[\ddot{l}_{a^2 a^2}]} i_{a^2}(\sigma^2, a^2) + O_p(1) \right)^2 \right] \\
&= N^{-1}E_{\text{true}} \left[\left(i_{\sigma^2}(\sigma^2, a^2) - \frac{E_{\text{true}}[\ddot{l}_{\sigma^2 a^2}]}{E_{\text{true}}[\ddot{l}_{a^2 a^2}]} i_{a^2}(\sigma^2, a^2) \right)^2 \right] + o(1) \\
&= N^{-1}E_{\text{true}} \left[i_{\sigma^2}(\sigma^2, a^2)^2 + \left(\frac{E_{\text{true}}[\ddot{l}_{\sigma^2 a^2}]}{E_{\text{true}}[\ddot{l}_{a^2 a^2}]} i_{a^2}(\sigma^2, a^2) \right)^2 \right. \\
&\quad \left. - 2 \frac{E_{\text{true}}[\ddot{l}_{\sigma^2 a^2}]}{E_{\text{true}}[\ddot{l}_{a^2 a^2}]} i_{a^2}(\sigma^2, a^2) i_{\sigma^2}(\sigma^2, a^2) \right] + o(1)
\end{aligned}$$

so that

$$\begin{aligned}
N^{-1}E_{\text{true}}[\dot{\lambda}(\sigma^2)^2] &= S_{\sigma^2 \sigma^2} + \left(\frac{D_{\sigma^2 a^2}}{D_{a^2 a^2}} \right)^2 S_{a^2 a^2} - 2 \frac{D_{\sigma^2 a^2}}{D_{a^2 a^2}} S_{a^2 \sigma^2} + o_p(1) \\
&= \left(D_{\sigma^2 \sigma^2} + \left(\frac{D_{\sigma^2 a^2}}{D_{a^2 a^2}} \right)^2 D_{a^2 a^2} - 2 \frac{D_{\sigma^2 a^2}}{D_{a^2 a^2}} D_{a^2 \sigma^2} \right) \\
&\quad + \text{Cum}_4(U) \left(g_{\sigma^2}^2 + \left(\frac{D_{\sigma^2 a^2}}{D_{a^2 a^2}} \right)^2 g_{a^2}^2 - 2 \frac{D_{\sigma^2 a^2}}{D_{a^2 a^2}} g_{\sigma^2} g_{a^2} \right) + o_p(1)
\end{aligned}$$

by invoking (6.4).

Continuing the calculation,

$$\begin{aligned}
N^{-1}E_{\text{true}}[\dot{\lambda}(\sigma^2)^2] &= \left(D_{\sigma^2 \sigma^2} - \frac{D_{\sigma^2 a^2}^2}{D_{a^2 a^2}} \right) + \text{Cum}_4(U) \left(g_{\sigma^2} - \frac{D_{\sigma^2 a^2}}{D_{a^2 a^2}} g_{a^2} \right)^2 + o(1) \\
&= 1/[D^{-1}]_{\sigma^2 \sigma^2} + o(1)
\end{aligned} \tag{D.10}$$

since from the expressions for D_{orig} and g_{orig} in (D.2) and (D.5) we have

$$g_{\sigma^2} - \frac{D_{\sigma^2 a^2}}{D_{a^2 a^2}} g_{a^2} = 0. \tag{D.11}$$

Then by (D.9) and the law of large numbers, we have

$$N^{-1}E_{\text{true}}[\ddot{\lambda}(\sigma^2)] = -1/[D^{-1}]_{\sigma^2 \sigma^2} + o(1), \tag{D.12}$$

and (6.8) follows from combining (D.10) with (D.12).

Appendix E: Proof of Lemma 1

$\Sigma \Sigma^{-1} \equiv Id$ implies that

$$\frac{\partial \Sigma^{-1}}{\partial \beta_1} = -\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \Sigma^{-1} \tag{E.1}$$

and, since Σ is linear in the parameters σ^2 and a^2 (see (7.2)) we have

$$\frac{\partial^2 \Sigma}{\partial \beta_2 \partial \beta_1} = 0 \tag{E.2}$$

so that

$$\begin{aligned}
\frac{\partial^2 \Sigma^{-1}}{\partial \beta_2 \partial \beta_1} &= \frac{\partial}{\partial \beta_2} \left(\frac{\partial \Sigma^{-1}}{\partial \beta_1} \right) \\
&= \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \Sigma^{-1} + \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1} - \Sigma^{-1} \frac{\partial^2 \Sigma}{\partial \beta_1 \partial \beta_2} \Sigma^{-1} \\
&= \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \Sigma^{-1} + \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1}
\end{aligned} \tag{E.3}$$

In the rest of this lemma, let expectations be conditional on the Δ 's. We use the notation $E[\cdot | \Delta]$ as a shortcut for $E[\cdot | \Delta_N, \dots, \Delta_1]$. At the true value of the parameter vector, we have,

$$\begin{aligned}
0 &= E[\dot{l}_{\beta_1} | \Delta] \\
&= -\frac{1}{2} \frac{\partial \ln \det \Sigma}{\partial \beta_1} - \frac{1}{2} E \left[Y' \frac{\partial \Sigma^{-1}}{\partial \beta_1} Y | \Delta \right].
\end{aligned} \tag{E.4}$$

with the second equality following from (7.3). Then, for any nonrandom Q , we have

$$E[Y' Q Y] = \text{Tr} [Q E[YY']] = \text{Tr} [Q \Sigma]. \tag{E.5}$$

This can be applied to Q that depends on the Δ 's, even when they are random, because the expected value is conditional on the Δ 's. Therefore it follows from (E.4) that

$$\frac{\partial \ln \det \Sigma}{\partial \beta_1} = -E \left[Y' \frac{\partial \Sigma^{-1}}{\partial \beta_1} Y | \Delta \right] = -\text{Tr} \left[\frac{\partial \Sigma^{-1}}{\partial \beta_1} \Sigma \right] = \text{Tr} \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \right], \tag{E.6}$$

with the last equality following from (E.1) and so

$$\begin{aligned}
\frac{\partial^2 \ln \det \Sigma}{\partial \beta_2 \partial \beta_1} &= \frac{\partial}{\partial \beta_2} \text{Tr} \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \right] \\
&= \text{Tr} \left[\frac{\partial}{\partial \beta_2} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \right) \right] \\
&= -\text{Tr} \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} + \Sigma^{-1} \frac{\partial^2 \Sigma}{\partial \beta_2 \partial \beta_1} \right] \\
&= -\text{Tr} \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \right],
\end{aligned} \tag{E.7}$$

again because of (E.2).

In light of (7.3), the expected information (conditional on the Δ 's) is given by

$$E \left[-\ddot{l}_{\beta_2 \beta_1} | \Delta \right] = \frac{1}{2} \frac{\partial^2 \ln \det \Sigma}{\partial \beta_2 \partial \beta_1} + \frac{1}{2} E \left[Y' \frac{\partial^2 \Sigma^{-1}}{\partial \beta_2 \partial \beta_1} Y | \Delta \right].$$

Then,

$$\begin{aligned}
E \left[Y' \frac{\partial^2 \Sigma^{-1}}{\partial \beta_2 \partial \beta_1} Y | \Delta \right] &= \text{Tr} \left[\frac{\partial^2 \Sigma^{-1}}{\partial \beta_2 \partial \beta_1} \Sigma \right] \\
&= \text{Tr} \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} + \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \right] \\
&= 2\text{Tr} \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \right]
\end{aligned}$$

with the first equality following from (E.5) applied to $Q = \partial^2 \Sigma^{-1} / \partial \beta_2 \beta_1$, the second from (E.3) and the third from the fact that $Tr[AB] = Tr[BA]$. It follows that

$$\begin{aligned} E \left[-\ddot{l}_{\beta_2 \beta_1} \mid \Delta \right] &= -\frac{1}{2} Tr \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \right] + Tr \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \right] \\ &= \frac{1}{2} Tr \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \right] \\ &= -\frac{1}{2} \frac{\partial^2 \ln \det \Sigma}{\partial \beta_2 \beta_1}. \end{aligned}$$

Appendix F: Proof of Theorem 2

In light of (7.2) and (7.5),

$$\Sigma = \Sigma_0 + \epsilon \sigma^2 \Xi \tag{F.1}$$

from which it follows that

$$\begin{aligned} \Sigma^{-1} &= (\Sigma_0 (Id + \epsilon \sigma^2 \Sigma_0^{-1} \Xi))^{-1} \\ &= (Id + \epsilon \sigma^2 \Sigma_0^{-1} \Xi)^{-1} \Sigma_0^{-1} \\ &= \Sigma_0^{-1} - \epsilon \sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1} + \epsilon^2 \sigma^4 (\Sigma_0^{-1} \Xi)^2 \Sigma_0^{-1} + O(\epsilon^3) \end{aligned} \tag{F.2}$$

since

$$(Id + \epsilon A)^{-1} = Id - \epsilon A + \epsilon^2 A^2 + O(\epsilon^3).$$

Also,

$$\frac{\partial \Sigma}{\partial \beta_1} = \frac{\partial \Sigma_0}{\partial \beta_1} + \epsilon \frac{\partial \sigma^2}{\partial \beta_1} \Xi.$$

Therefore, recalling (E.6), we have

$$\begin{aligned} \frac{\partial \ln \det \Sigma}{\partial \beta_1} &= Tr \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta_1} \right] \\ &= Tr \left[(\Sigma_0^{-1} - \epsilon \sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1} + \epsilon^2 \sigma^4 (\Sigma_0^{-1} \Xi)^2 \Sigma_0^{-1} + O(\epsilon^3)) \left(\frac{\partial \Sigma_0}{\partial \beta_1} + \epsilon \frac{\partial \sigma^2}{\partial \beta_1} \Xi \right) \right] \\ &= Tr \left[\Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] + \epsilon Tr \left[-\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} + \frac{\partial \sigma^2}{\partial \beta_1} \Sigma_0^{-1} \Xi \right] \\ &\quad + \epsilon^2 Tr \left[\sigma^4 (\Sigma_0^{-1} \Xi)^2 \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} - \sigma^2 \frac{\partial \sigma^2}{\partial \beta_1} \Sigma_0^{-1} \Xi \Sigma_0^{-1} \Xi \right] \\ &\quad + O_p(\epsilon^3) \end{aligned} \tag{F.3}$$

Two things can be determined from this expansion. Since the ξ'_i s are iid with mean 0, $E[\Xi] = 0$, and so, taking unconditional expectations with respect to the law of the Δ'_i s, we obtain that the coefficient of order ϵ is

$$\begin{aligned} &E \left[Tr \left[-\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} + \frac{\partial \sigma^2}{\partial \beta_1} \Sigma_0^{-1} \Xi \right] \right] \\ &= Tr \left[E \left[-\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} + \frac{\partial \sigma^2}{\partial \beta_1} \Sigma_0^{-1} \Xi \right] \right] \\ &= Tr \left[-\sigma^2 \Sigma_0^{-1} E[\Xi] \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} + \frac{\partial \sigma^2}{\partial \beta_1} \Sigma_0^{-1} E[\Xi] \right] \\ &= 0. \end{aligned}$$

Similarly, the coefficient of order ϵ^2 is

$$\begin{aligned}
& E \left[\text{Tr} \left[\sigma^4 (\Sigma_0^{-1} \Xi)^2 \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} - \sigma^2 \frac{\partial \sigma^2}{\partial \beta_1} (\Sigma_0^{-1} \Xi)^2 \right] \right] \\
&= \text{Tr} \left[\sigma^4 E \left[(\Sigma_0^{-1} \Xi)^2 \right] \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} - \sigma^2 \frac{\partial \sigma^2}{\partial \beta_1} E \left[(\Sigma_0^{-1} \Xi)^2 \right] \right] \\
&= \text{Tr} \left[\sigma^2 E \left[(\Sigma_0^{-1} \Xi)^2 \right] \left(\sigma^2 \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} - \frac{\partial \sigma^2}{\partial \beta_1} \text{Id} \right) \right] \\
&= \text{Tr} \left[\sigma^2 \Sigma_0^{-1} E \left[\Xi \Sigma_0^{-1} \Xi \right] \left(\sigma^2 \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} - \frac{\partial \sigma^2}{\partial \beta_1} \text{Id} \right) \right].
\end{aligned}$$

The matrix $E \left[\Xi \Sigma_0^{-1} \Xi \right]$ has the following terms

$$\left[\Xi \Sigma_0^{-1} \Xi \right]_{i,j} = \sum_{k=1}^N \sum_{l=1}^N \Xi_{ik} \left[\Sigma_0^{-1} \right]_{kl} \Xi_{lj} = \Delta_0^2 \xi_i \xi_j \left[\Sigma_0^{-1} \right]_{ij}$$

and since $E \left[\xi_i \xi_j \right] = \delta_{ij} \text{Var}[\xi]$ (where δ_{ij} denotes the Kronecker symbol), it follows that

$$E \left[\Xi \Sigma_0^{-1} \Xi \right] = \Delta_0^2 \text{Var}[\xi] \text{diag} \left[\Sigma_0^{-1} \right] \tag{F.4}$$

where $\text{diag} \left[\Sigma_0^{-1} \right]$ is the diagonal matrix formed with the diagonal elements of Σ_0^{-1} . From this, we obtain that

$$\begin{aligned}
E \left[\frac{\partial \ln \det \Sigma}{\partial \beta_1} \right] &= \text{Tr} \left[\Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] \\
&\quad + \epsilon^2 \text{Tr} \left[\sigma^2 \Sigma_0^{-1} E \left[\Xi \Sigma_0^{-1} \Xi \right] \left(\sigma^2 \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} - \frac{\partial \sigma^2}{\partial \beta_1} \text{Id} \right) \right] + O(\epsilon^3) \\
&= \text{Tr} \left[\Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] \\
&\quad + \epsilon^2 \Delta_0^2 \text{Var}[\xi] \text{Tr} \left[\sigma^2 \Sigma_0^{-1} \text{diag} \left[\Sigma_0^{-1} \right] \left(\sigma^2 \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} - \frac{\partial \sigma^2}{\partial \beta_1} \text{Id} \right) \right] + O(\epsilon^3).
\end{aligned} \tag{F.5}$$

To calculate $E \left[\ddot{l}_{\beta_2 \beta_1} \right]$, in light of (7.4), we need to differentiate $E \left[\partial \ln \det \Sigma / \partial \beta_1 \right]$ with respect to β_2 . Indeed

$$E \left[-\ddot{l}_{\beta_2 \beta_1} \right] = E \left[E \left[-\ddot{l}_{\beta_2 \beta_1} \mid \Delta \right] \right] = -\frac{1}{2} E \left[\frac{\partial^2 \ln \det \Sigma}{\partial \beta_2 \partial \beta_1} \right] = -\frac{1}{2} \frac{\partial}{\partial \beta_2} \left(E \left[\frac{\partial \ln \det \Sigma}{\partial \beta_1} \right] \right)$$

where we can interchange the unconditional expectation and the differentiation with respect to β_2 because the unconditional expectation is taken with respect to the law of the Δ_i 's, which is independent of the β parameters (i.e., σ^2 and a^2). Therefore, differentiating (F.5) with respect to β_2 will produce the result we need. (The reader may wonder why we take the expected value before differentiating, rather than the other way around. As just discussed, the results are identical. However, it turns out that taking expectations first reduces the computational burden quite substantially.)

Combining with (F.5), we therefore have

$$\begin{aligned}
E \left[-\ddot{l}_{\beta_2 \beta_1} \right] &= -\frac{1}{2} \frac{\partial}{\partial \beta_2} \left(E \left[\frac{\partial \ln \det \Sigma}{\partial \beta_1} \right] \right) \\
&= -\frac{1}{2} \frac{\partial}{\partial \beta_2} \text{Tr} \left[\Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] \\
&\quad - \frac{1}{2} \epsilon^2 \Delta_0^2 \text{Var}[\xi] \frac{\partial}{\partial \beta_2} \left(\text{Tr} \left[\sigma^2 \Sigma_0^{-1} \text{diag} \left[\Sigma_0^{-1} \right] \left(\sigma^2 \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} - \frac{\partial \sigma^2}{\partial \beta_1} \text{Id} \right) \right] \right) \\
&\quad + O(\epsilon^3) \\
&\equiv \phi^{(0)} + \epsilon^2 \phi^{(2)} + O(\epsilon^3)
\end{aligned} \tag{F.6}$$

It is useful now to introduce the same transformed parameters (γ^2, η) as in previous sections and write $\Sigma_0 = \gamma^2 V$ with the parameters and V defined as in (3.2)-(3.3) and (4.2), except that Δ is replaced by Δ_0 in these expressions. To compute $\phi^{(0)}$, we start with

$$\begin{aligned}
Tr \left[\Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] &= Tr \left[\gamma^{-2} V^{-1} \frac{\partial (\gamma^2 V)}{\partial \beta_1} \right] \\
&= Tr \left[V^{-1} \frac{\partial V}{\partial \beta_1} \right] + Tr \left[\gamma^{-2} V^{-1} V \frac{\partial \gamma^2}{\partial \beta_1} \right] \\
&= Tr \left[V^{-1} \frac{\partial V}{\partial \eta} \right] \frac{\partial \eta}{\partial \beta_1} + N \gamma^{-2} \frac{\partial \gamma^2}{\partial \beta_1}
\end{aligned} \tag{F.7}$$

with $\partial \gamma^2 / \partial \beta_1$ and $\partial \eta / \partial \beta_1$ to be computed from (3.4)-(3.5). If Id denotes the identity matrix and J the matrix with 1 on the infra and supra-diagonal lines and 0 everywhere else, we have $V = \eta^2 Id + \eta J$, so that $\partial V / \partial \eta = 2\eta Id + J$. Therefore

$$\begin{aligned}
Tr \left[V^{-1} \frac{\partial V}{\partial \eta} \right] &= 2\eta Tr [V^{-1}] + Tr [V^{-1} J] \\
&= 2\eta \sum_{i=1}^N v^{i,i} + \sum_{i=2}^{N-1} \{v^{i,i-1} + v^{i,i+1}\} + v^{1,2} + v^{N,N-1} \\
&= \frac{2\eta (1 - \eta^{2N} (N(1 - \eta^2) + 1))}{(1 - \eta^2) (1 - \eta^{2(1+N)})} \\
&= \frac{2\eta}{(1 - \eta^2)} + o(1)
\end{aligned} \tag{F.8}$$

Therefore the first term in (F.7) is $O(1)$ while the second term is $O(N)$ and hence

$$Tr \left[\Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] = N \gamma^{-2} \frac{\partial \gamma^2}{\partial \beta_1} + O(1).$$

This holds also for the partial derivative of (F.7) with respect to β_2 . Indeed, given the form of (F.8), we have that

$$\frac{\partial}{\partial \beta_2} \left(Tr \left[V^{-1} \frac{\partial V}{\partial \eta} \right] \right) = \frac{\partial}{\partial \beta_2} \left(\frac{2\eta}{(1 - \eta^2)} \right) + o(1) = O(1)$$

since the remainder term in (F.8) is of the form $p(N)\eta^q(N)$, where p and q are polynomials in N or order greater than or equal to 0 and 1 respectively, whose differentiation with respect to η will produce terms that are of order $o(N)$. Thus it follows that

$$\begin{aligned}
\frac{\partial}{\partial \beta_2} \left(Tr \left[\Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] \right) &= N \frac{\partial}{\partial \beta_2} \left(\gamma^{-2} \frac{\partial \gamma^2}{\partial \beta_1} \right) + o(N) \\
&= N \left\{ \frac{\partial \gamma^{-2}}{\partial \beta_2} \frac{\partial \gamma^2}{\partial \beta_1} + \gamma^{-2} \frac{\partial^2 \gamma^2}{\partial \beta_2 \partial \beta_1} \right\} + o(N)
\end{aligned} \tag{F.9}$$

Writing the result in matrix form, where the (1, 1) element corresponds to $(\beta_1, \beta_2) = (\sigma^2, \sigma^2)$, the (1, 2) and (2, 1) elements to $(\beta_1, \beta_2) = (\sigma^2, a^2)$ and the (2, 2) element to $(\beta_1, \beta_2) = (a^2, a^2)$, and computing the partial derivatives in (F.9), we have

$$\begin{aligned}
\phi^{(0)} &= -\frac{1}{2} \frac{\partial}{\partial \beta_2} \left(Tr \left[\Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] \right) \\
&= N \left(\begin{array}{c} \frac{\Delta_0^{1/2} (2a^2 + \sigma^2 \Delta_0)}{2\sigma^3 (4a^2 + \sigma^2 \Delta_0)^{3/2}} \\ \bullet \\ \frac{1}{2a^4} \left(1 - \frac{\Delta_0^{1/2} \sigma (6a^2 + \sigma^2 \Delta_0)}{(4a^2 + \sigma^2 \Delta_0)^{3/2}} \right) \end{array} \right) + o(N).
\end{aligned} \tag{F.10}$$

As for the coefficient of order ϵ^2 , that is $\phi^{(2)}$ in (F.6), define

$$\alpha \equiv Tr \left[\sigma^2 \Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}] \left(\sigma^2 \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} - \frac{\partial \sigma^2}{\partial \beta_1} Id \right) \right] \quad (\text{F.11})$$

so that

$$\phi^{(2)} = -\frac{1}{2} \Delta_0^2 \text{Var}[\xi] \frac{\partial \alpha}{\partial \beta_2}.$$

We have

$$\begin{aligned} \alpha &= \sigma^4 Tr \left[\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}] \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] - \sigma^2 \frac{\partial \sigma^2}{\partial \beta_1} Tr [\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}]] \\ &= \sigma^4 \gamma^{-6} Tr \left[V^{-1} \text{diag} [V^{-1}] V^{-1} \frac{\partial (\gamma^2 V)}{\partial \beta_1} \right] - \sigma^2 \frac{\partial \sigma^2}{\partial \beta_1} \gamma^{-4} Tr [V^{-1} \text{diag} [V^{-1}]] \\ &= \sigma^4 \gamma^{-4} Tr \left[V^{-1} \text{diag} [V^{-1}] V^{-1} \frac{\partial V}{\partial \eta} \frac{\partial \eta}{\partial \beta_1} \right] \\ &\quad + \sigma^4 \gamma^{-6} \frac{\partial \gamma^2}{\partial \beta_1} Tr [V^{-1} \text{diag} [V^{-1}]] - \sigma^2 \frac{\partial \sigma^2}{\partial \beta_1} \gamma^{-4} Tr [V^{-1} \text{diag} [V^{-1}]] \\ &= \sigma^4 \gamma^{-4} \frac{\partial \eta}{\partial \beta_1} Tr \left[V^{-1} \text{diag} [V^{-1}] V^{-1} \frac{\partial V}{\partial \eta} \right] + \left(\sigma^4 \gamma^{-6} \frac{\partial \gamma^2}{\partial \beta_1} - \sigma^2 \frac{\partial \sigma^2}{\partial \beta_1} \gamma^{-4} \right) Tr [V^{-1} \text{diag} [V^{-1}]]. \end{aligned}$$

Next, we compute separately

$$\begin{aligned} Tr \left[V^{-1} \text{diag} [V^{-1}] V^{-1} \frac{\partial V}{\partial \eta} \right] &= Tr \left[\text{diag} [V^{-1}] V^{-1} \frac{\partial V}{\partial \eta} V^{-1} \right] \\ &= -Tr \left[\text{diag} [V^{-1}] \frac{\partial V^{-1}}{\partial \eta} \right] \\ &= -\sum_{i=1}^N v^{i,i} \frac{\partial v^{i,i}}{\partial \eta} \\ &= \frac{O(1) - 2N\eta (1 + \eta^2 - \eta^4 - \eta^6 + O(\eta^{2N})) + N^2 O(\eta^{2N})}{(1 + \eta^2)^2 (1 - \eta^2)^4 (1 - \eta^{2(1+N)})^3} \\ &= \frac{-2N\eta}{(1 - \eta^2)^3} + o(N) \end{aligned}$$

and

$$\begin{aligned} Tr [V^{-1} \text{diag} [V^{-1}]] &= \sum_{i=1}^N (v^{i,i})^2 \\ &= \frac{O(1) + N (1 - \eta^4 + O(\eta^{2N}))}{(1 + \eta^2) (1 - \eta^2)^3 (1 - \eta^{2(1+N)})^2} \\ &= \frac{N}{(1 - \eta^2)^2} + o(N) \end{aligned}$$

Therefore

$$\alpha = \sigma^4 \gamma^{-4} \frac{\partial \eta}{\partial \beta_1} \left(\frac{-2N\eta}{(1 - \eta^2)^3} \right) + \left(\sigma^4 \gamma^{-6} \frac{\partial \gamma^2}{\partial \beta_1} - \sigma^2 \frac{\partial \sigma^2}{\partial \beta_1} \gamma^{-4} \right) \left(\frac{N}{(1 - \eta^2)^2} \right) + o(N),$$

which can be differentiated with respect to β_2 to produce $\partial \alpha / \partial \beta_2$. As above, differentiation of the remainder term $o(N)$ still produces a $o(N)$ term because of the structure of the terms there (they are again of the form $p(N)\eta^{q(N)}$.)

Note that an alternative expression for α can be obtained as follows. Going back to the definition (F.11),

$$\alpha = \sigma^4 \text{Tr} \left[\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}] \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] - \sigma^2 \frac{\partial \sigma^2}{\partial \beta_1} \text{Tr} [\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}]] \quad (\text{F.12})$$

the first trace becomes

$$\begin{aligned} \text{Tr} \left[\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}] \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \right] &= \text{Tr} \left[\text{diag} [\Sigma_0^{-1}] \Sigma_0^{-1} \frac{\partial \Sigma_0}{\partial \beta_1} \Sigma_0^{-1} \right] \\ &= -\text{Tr} \left[\text{diag} [\Sigma_0^{-1}] \frac{\partial \Sigma_0^{-1}}{\partial \beta_1} \right] \\ &= -\sum_{i=1}^N (\Sigma_0^{-1})_{ii} \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1} \right)_{ii} \\ &= -\frac{1}{2} \frac{\partial}{\partial \beta_1} \sum_{i=1}^N (\Sigma_0^{-1})_{ii}^2 \\ &= -\frac{1}{2} \frac{\partial}{\partial \beta_1} \text{Tr} [\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}]]. \end{aligned}$$

so that we have

$$\begin{aligned} \alpha &= -\sigma^4 \frac{1}{2} \frac{\partial}{\partial \beta_1} \text{Tr} [\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}]] - \sigma^2 \frac{\partial \sigma^2}{\partial \beta_1} \text{Tr} [\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}]] \\ &= -\sigma^4 \frac{1}{2} \frac{\partial}{\partial \beta_1} \text{Tr} [\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}]] - \frac{1}{2} \left(\frac{\partial \sigma^4}{\partial \beta_1} \right) \text{Tr} [\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}]] \\ &= -\frac{1}{2} \frac{\partial}{\partial \beta_1} \left(\sigma^4 \text{Tr} [\Sigma_0^{-1} \text{diag} [\Sigma_0^{-1}]] \right) \\ &= -\frac{1}{2} \frac{\partial}{\partial \beta_1} \left(\sigma^4 \gamma^{-4} \text{Tr} [V^{-1} \text{diag} [V^{-1}]] \right) \\ &= -\frac{1}{2} \frac{\partial}{\partial \beta_1} \left(\sigma^4 \gamma^{-4} \left(\frac{N}{(1-\eta^2)^2} + o(N) \right) \right) \\ &= -\frac{N}{2} \frac{\partial}{\partial \beta_1} \left(\frac{\sigma^4 \gamma^{-4}}{(1-\eta^2)^2} \right) + o(N), \end{aligned}$$

where the calculation of $\text{Tr} [V^{-1} \text{diag} [V^{-1}]]$ is as before, and where the $o(N)$ term is a sum of terms of the form $p(N)\eta^{q(N)}$ as discussed above. From this one can interchange differentiation and the $o(N)$ term, yielding the final equality above.

Therefore

$$\begin{aligned} \frac{\partial \alpha}{\partial \beta_2} &= -\frac{1}{2} \frac{\partial^2}{\partial \beta_1 \partial \beta_2} \left(\sigma^4 \gamma^{-4} \left(\frac{N}{(1-\eta^2)^2} + o(N) \right) \right) \\ &= -\frac{N}{2} \frac{\partial^2}{\partial \beta_1 \partial \beta_2} \left(\frac{\sigma^4 \gamma^{-4}}{(1-\eta^2)^2} \right) + o(N). \end{aligned} \quad (\text{F.13})$$

Writing the result in matrix form and calculating the partial derivatives, we obtain

$$\phi^{(2)} = -\frac{1}{2} \Delta_0^2 \text{Var}[\xi] \frac{\partial \alpha}{\partial \beta_2} = \frac{N \Delta_0^2 \text{Var}[\xi]}{(4a^2 + \sigma^2 \Delta_0)^3} \begin{pmatrix} -2a^2 & -\frac{(8a^2 - 2\sigma^2 \Delta_0)}{2\Delta_0} \\ \bullet & -\frac{8\sigma^2}{\Delta_0} \end{pmatrix} + o(N) \quad (\text{F.14})$$

Putting it all together, we have obtained

$$\begin{aligned}\frac{1}{N}E\left[-\ddot{l}_{\beta_2\beta_1}\right] &= \frac{1}{N}\left(\phi^{(0)} + \epsilon^2\phi^{(2)} + O(\epsilon^3)\right) \\ &\equiv F^{(0)} + \epsilon^2F^{(2)} + O(\epsilon^3) + o(1)\end{aligned}\tag{F.15}$$

where

$$F^{(0)} = \begin{pmatrix} \frac{\Delta_0^{1/2}(2a^2 + \sigma^2\Delta_0)}{2\sigma^3(4a^2 + \sigma^2\Delta_0)^{3/2}} & \frac{\Delta_0^{1/2}}{\sigma(4a^2 + \sigma^2\Delta_0)^{3/2}} \\ \bullet & \frac{1}{2a^4}\left(1 - \frac{\Delta_0^{1/2}\sigma(6a^2 + \sigma^2\Delta_0)}{(4a^2 + \sigma^2\Delta_0)^{3/2}}\right) \end{pmatrix}\tag{F.16}$$

$$F^{(2)} \equiv \frac{\Delta_0^2 Var[\xi]}{(4a^2 + \sigma^2\Delta_0)^3} \begin{pmatrix} -2a^2 & -\frac{(8a^2 - 2\sigma^2\Delta_0)}{2\Delta_0} \\ \bullet & -\frac{8\sigma^2}{\Delta_0} \end{pmatrix}.\tag{F.17}$$

The asymptotic variance of the maximum-likelihood estimators $AVAR(\hat{\sigma}^2, \hat{a}^2)$ is therefore given by

$$\begin{aligned}AVAR(\hat{\sigma}^2, \hat{a}^2) &= E[\Delta] \left(F^{(0)} + \epsilon^2F^{(2)} + O(\epsilon^3)\right)^{-1} \\ &= \Delta_0 \left(F^{(0)} \left(Id + \epsilon^2 \left[F^{(0)}\right]^{-1} F^{(2)} + O(\epsilon^3)\right)\right)^{-1} \\ &= \Delta_0 \left(Id + \epsilon^2 \left[F^{(0)}\right]^{-1} F^{(2)} + O(\epsilon^3)\right)^{-1} \left[F^{(0)}\right]^{-1} \\ &= \Delta_0 \left(Id - \epsilon^2 \left[F^{(0)}\right]^{-1} F^{(2)} + O(\epsilon^3)\right) \left[F^{(0)}\right]^{-1} \\ &= \Delta_0 \left[F^{(0)}\right]^{-1} - \epsilon^2\Delta_0 \left[F^{(0)}\right]^{-1} F^{(2)} \left[F^{(0)}\right]^{-1} + O(\epsilon^3) \\ &\equiv A^{(0)} + \epsilon^2A^{(2)} + O(\epsilon^3)\end{aligned}$$

where the final results for $A^{(0)} = \Delta_0 \left[F^{(0)}\right]^{-1}$ and $A^{(2)} = -\Delta_0 \left[F^{(0)}\right]^{-1} F^{(2)} \left[F^{(0)}\right]^{-1}$, obtained by replacing $F^{(0)}$ and $F^{(2)}$ by their expressions in (F.15), are given in the statement of the Theorem.

Appendix G: Proof of Theorem 3

It follows as in (5.4), (5.5) and (5.6) that

$$\begin{aligned}E_{\text{true}}\left[\dot{l}_{\beta_1}\dot{l}_{\beta_2}|\Delta\right] &= Cov_{\text{true}}(\dot{l}_{\beta_1}, \dot{l}_{\beta_2}|\Delta) \\ &= Cov_{\text{true}}\left(-\frac{1}{2}\sum_{i,j=1}^N Y_i Y_j \left(\frac{\partial\Sigma^{-1}}{\partial\beta_1}\right)_{ij}, -\frac{1}{2}\sum_{k,l=1}^N Y_k Y_l \left(\frac{\partial\Sigma^{-1}}{\partial\beta_2}\right)_{kl}\right) \\ &= \frac{1}{4}\sum_{i,j,k,l=1}^N \left(\frac{\partial\Sigma^{-1}}{\partial\beta_1}\right)_{ij} \left(\frac{\partial\Sigma^{-1}}{\partial\beta_2}\right)_{kl} Cov_{\text{true}}(Y_i Y_j, Y_k Y_l|\Delta) \\ &= -E_{\text{true}}\left[\ddot{l}_{\beta_1\beta_2}|\Delta\right] + \frac{1}{4}\sum_{i,j,k,l=1}^N \left(\frac{\partial\Sigma^{-1}}{\partial\beta_1}\right)_{ij} \left(\frac{\partial\Sigma^{-1}}{\partial\beta_2}\right)_{kl} Cum_{\text{true}}(Y_i, Y_j, Y_k, Y_l|\Delta) \\ &= -E_{\text{true}}\left[\ddot{l}_{\beta_1\beta_2}|\Delta\right] + \frac{1}{4}Cum_4(U)\psi\left(\frac{\partial\Sigma^{-1}}{\partial\beta_1}, \frac{\partial\Sigma^{-1}}{\partial\beta_2}\right)\end{aligned}\tag{G.1}$$

since $Cum_{\text{true}}(Y_i, Y_j, Y_k, Y_l | \Delta) = 2, \pm 1, \text{ or } 0, \times Cum_{\text{true}}(U)$, as in (5.8), and with ψ defined in (C.3). Taking now unconditional expectations, we have

$$\begin{aligned}
E_{\text{true}} \left[\dot{l}_{\beta_1} \dot{l}_{\beta_2} \right] &= Cov_{\text{true}}(\dot{l}_{\beta_1}, \dot{l}_{\beta_2}) \\
&= E \left[Cov_{\text{true}}(\dot{l}_{\beta_1}, \dot{l}_{\beta_2} | \Delta) \right] + Cov_{\text{true}}(E_{\text{true}}[\dot{l}_{\beta_1} | \Delta], E_{\text{true}}[\dot{l}_{\beta_2} | \Delta]) \\
&= E \left[Cov_{\text{true}}(\dot{l}_{\beta_1}, \dot{l}_{\beta_2} | \Delta) \right] \\
&= -E_{\text{true}} \left[\ddot{l}_{\beta_1 \beta_2} \right] + \frac{1}{4} Cum_4(U) E \left[\psi \left(\frac{\partial \Sigma^{-1}}{\partial \beta_1}, \frac{\partial \Sigma^{-1}}{\partial \beta_2} \right) \right].
\end{aligned} \tag{G.2}$$

with the first and third equalities following from the fact that $E_{\text{true}}[\dot{l}_{\beta_i} | \Delta] = 0$.

Since

$$E_{\text{true}} \left[\ddot{l}_{\beta_1 \beta_2} | \Delta \right] = E_{\text{normal}} \left[\ddot{l}_{\beta_1 \beta_2} | \Delta \right]$$

and consequently

$$E_{\text{true}} \left[\ddot{l}_{\beta_1 \beta_2} \right] = E_{\text{normal}} \left[\ddot{l}_{\beta_1 \beta_2} \right]$$

have been found in the previous subsection (see (F.15)), what we need to do to obtain $E_{\text{true}} \left[\dot{l}_{\beta_1} \dot{l}_{\beta_2} \right]$ is to calculate

$$E \left[\psi \left(\frac{\partial \Sigma^{-1}}{\partial \beta_1}, \frac{\partial \Sigma^{-1}}{\partial \beta_2} \right) \right].$$

With Σ^{-1} given by (F.2), we have for $i = 1, 2$

$$\frac{\partial \Sigma^{-1}}{\partial \beta_i} = \frac{\partial \Sigma_0^{-1}}{\partial \beta_i} - \epsilon \frac{\partial}{\partial \beta_i} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) + \epsilon^2 \frac{\partial}{\partial \beta_i} (\sigma^4 (\Sigma_0^{-1} \Xi)^2 \Sigma_0^{-1}) + O(\epsilon^3)$$

and therefore by bilinearity of ψ we have

$$\begin{aligned}
E \left[\psi \left(\frac{\partial \Sigma^{-1}}{\partial \beta_1}, \frac{\partial \Sigma^{-1}}{\partial \beta_2} \right) \right] &= \psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial \Sigma_0^{-1}}{\partial \beta_2} \right) - \epsilon E \left[\psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial}{\partial \beta_2} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right) \right] [2] \\
&\quad + \epsilon^2 E \left[\psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial}{\partial \beta_2} (\sigma^4 (\Sigma_0^{-1} \Xi)^2 \Sigma_0^{-1}) \right) \right] [2] \\
&\quad + \epsilon^2 E \left[\psi \left(\frac{\partial}{\partial \beta_1} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}), \frac{\partial}{\partial \beta_2} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right) \right] \\
&\quad + O(\epsilon^3),
\end{aligned} \tag{G.3}$$

where the “[2]” refers to the sum over the two terms where β_1 and β_2 are permuted.

The first (and leading) term in (G.3),

$$\begin{aligned}
\psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial \Sigma_0^{-1}}{\partial \beta_2} \right) &= \psi \left(\frac{\partial (\gamma^{-2} V^{-1})}{\partial \beta_1}, \frac{\partial (\gamma^{-2} V^{-1})}{\partial \beta_2} \right) \\
&= \psi \left(\frac{\partial \gamma^{-2}}{\partial \beta_1} V^{-1} + \gamma^{-2} \frac{\partial V^{-1}}{\partial \beta_1}, \frac{\partial \gamma^{-2}}{\partial \beta_2} V^{-1} + \gamma^{-2} \frac{\partial V^{-1}}{\partial \beta_2} \right) \\
&= \psi \left(\frac{\partial \gamma^{-2}}{\partial \beta_1} V^{-1} + \gamma^{-2} \frac{\partial V^{-1}}{\partial \eta} \frac{\partial \eta}{\partial \beta_1}, \frac{\partial \gamma^{-2}}{\partial \beta_1} V^{-1} + \gamma^{-2} \frac{\partial V^{-1}}{\partial \eta} \frac{\partial \eta}{\partial \beta_1} \right)
\end{aligned}$$

corresponds to the equally spaced, misspecified noise distribution, situation studied in Section 5.

The second term, linear in ϵ , is zero since

$$\begin{aligned} E \left[\psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial}{\partial \beta_2} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right) \right] &= \psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, E \left[\frac{\partial}{\partial \beta_2} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right] \right) \\ &= \psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial}{\partial \beta_2} (\sigma^2 \Sigma_0^{-1} E[\Xi] \Sigma_0^{-1}) \right) \\ &= 0 \end{aligned}$$

with the first equality following from the bilinearity of ψ , the second from the fact that the unconditional expectation over the Δ_i 's does not depend on the β parameters, so expectation and differentiation with respect to β_2 can be interchanged, and the third equality from the fact that $E[\Xi] = 0$.

To calculate the third term in (G.3), the first of two that are quadratic in ϵ , note that

$$\begin{aligned} \alpha_1 &\equiv E \left[\psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial}{\partial \beta_2} (\sigma^4 \Sigma_0^{-1} \Xi \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right) \right] \\ &= \psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial}{\partial \beta_2} (\sigma^4 \Sigma_0^{-1} E[\Xi \Sigma_0^{-1} \Xi] \Sigma_0^{-1}) \right) \\ &= \Delta_0^2 \text{Var}[\xi] \psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial}{\partial \beta_2} (\sigma^4 \Sigma_0^{-1} \text{diag}(\Sigma_0^{-1}) \Sigma_0^{-1}) \right) \\ &= \Delta_0^2 \text{Var}[\xi] \psi \left(\frac{\partial (\gamma^{-2} V^{-1})}{\partial \beta_1}, \frac{\partial}{\partial \beta_2} (\sigma^4 \gamma^{-6} V^{-1} \text{diag}(V^{-1}) V^{-1}) \right), \end{aligned} \quad (\text{G.4})$$

with the second equality obtained by replacing $E[\Xi \Sigma_0^{-1} \Xi]$ with its value given in (F.4), and the third by recalling that $\Sigma_0 = \gamma^2 V$. The elements (i, j) of the two arguments of ψ in (G.4) are

$$\nu^{i,j} = \frac{\partial (\gamma^{-2} v^{i,j})}{\partial \beta_1} = \frac{\partial \gamma^{-2}}{\partial \beta_1} v^{i,j} + \gamma^{-2} \frac{\partial v^{i,j}}{\partial \eta} \frac{\partial \eta}{\partial \beta_1}$$

and

$$\begin{aligned} \omega^{k,l} &= \frac{\partial}{\partial \beta_2} \left(\sigma^4 \gamma^{-6} \sum_{m=1}^N v^{k,m} v^{m,m} v^{m,l} \right) \\ &= \frac{\partial (\sigma^4 \gamma^{-6})}{\partial \beta_2} \sum_{m=1}^N v^{k,m} v^{m,m} v^{m,l} + \sigma^4 \gamma^{-6} \frac{\partial}{\partial \eta} \left(\sum_{m=1}^N v^{k,m} v^{m,m} v^{m,l} \right) \frac{\partial \eta}{\partial \beta_2} \end{aligned}$$

from which ψ in (G.4) can be evaluated through the sum given in (C.3).

Summing these terms, we obtain

$$\begin{aligned} &\psi \left(\frac{\partial (\gamma^{-2} V^{-1})}{\partial \beta_1}, \frac{\partial}{\partial \beta_2} (\sigma^4 \gamma^{-6} V^{-1} \text{diag}(V^{-1}) V^{-1}) \right) \\ &= \frac{4N (C_{1V} (1 - \eta) + C_{2V} C_{3V}) (C_{1W} (1 - \eta^2) + 2C_{2W} C_{3W} (1 + 3\eta))}{(1 - \eta)^7 (1 + \eta)^3} + o(N) \end{aligned}$$

where

$$\begin{aligned} C_{1V} &= \frac{\partial \gamma^{-2}}{\partial \beta_1}, \quad C_{2V} = \gamma^{-2}, \quad C_{3V} = \frac{\partial \eta}{\partial \beta_1} \\ C_{1W} &= \frac{\partial (\sigma^4 \gamma^{-6})}{\partial \beta_2}, \quad C_{2W} = \sigma^4 \gamma^{-6}, \quad C_{3W} = \frac{\partial \eta}{\partial \beta_2} \end{aligned}$$

The fourth and last term in (G.3), also quadratic in ϵ ,

$$\alpha_2 \equiv E \left[\psi \left(\frac{\partial}{\partial \beta_1} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}), \frac{\partial}{\partial \beta_2} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right) \right]$$

is obtained by first expressing

$$\psi \left(\frac{\partial}{\partial \beta_1} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}), \frac{\partial}{\partial \beta_2} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right)$$

in its sum form and then taking expectations term by term. Letting now

$$\nu^{i,j} = \left(\frac{\partial}{\partial \beta_1} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right)_{ij}, \quad \omega^{k,l} = \left(\frac{\partial}{\partial \beta_2} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right)_{kl}$$

we recall our definition of $\psi(\nu, \omega)$ given in (C.3) whose unconditional expected value (over the Δ'_i 's, i.e., over Ξ) we now need to evaluate in order to obtain α_2 .

We are thus led to consider four-index tensors λ^{ijkl} and to define

$$\begin{aligned} \tilde{\psi}(\lambda) \equiv & 2 \sum_{h=1}^N \lambda^{h,h,h,h} + \sum_{h=1}^{N-1} \left\{ -2\lambda^{h,h+1,h+1,h+1} - 2\lambda^{h+1,h+1,h,h+1} \right. \\ & + \lambda^{h,h,h+1,h+1} + \lambda^{h+1,h+1,h,h} + 4\lambda^{h,h+1,h,h+1} \\ & \left. - 2\lambda^{h+1,h,h,h} - 2\lambda^{h,h,h+1,h} \right\} \end{aligned} \quad (\text{G.5})$$

where λ^{ijkl} is symmetric in the two first and the two last indices, respectively, i.e., $\lambda^{ijkl} = \lambda^{jikl}$ and $\lambda^{ijkl} = \lambda^{ijlk}$. In terms of our definition of ψ in (C.3), it should be noted that $\psi(\nu, \omega) = \tilde{\psi}(\lambda)$ when one takes $\lambda^{ijkl} = \nu^{i,j} \omega^{k,l}$. The expression we seek is therefore

$$\alpha_2 = E \left[\psi \left(\frac{\partial}{\partial \beta_1} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}), \frac{\partial}{\partial \beta_2} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right) \right] = \tilde{\psi}(\lambda) \quad (\text{G.6})$$

where λ^{ijkl} is taken to be the following expected value

$$\begin{aligned} \lambda^{ijkl} & \equiv E [\nu^{i,j} \omega^{k,l}] \\ & = E \left[\left(\frac{\partial}{\partial \beta_1} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right)_{ij} \left(\frac{\partial}{\partial \beta_2} (\sigma^2 \Sigma_0^{-1} \Xi \Sigma_0^{-1}) \right)_{kl} \right] \\ & = E \left[\sum_{r,s,t,u=1}^N \frac{\partial}{\partial \beta_1} (\sigma^2 (\Sigma_0^{-1})_{ir} \Xi_{rs} (\Sigma_0^{-1})_{sj}) \frac{\partial}{\partial \beta_2} (\sigma^2 (\Sigma_0^{-1})_{kt} \Xi_{tu} (\Sigma_0^{-1})_{ul}) \right] \\ & = \sum_{r,s,t,u=1}^N \frac{\partial}{\partial \beta_1} (\sigma^2 (\Sigma_0^{-1})_{ir} (\Sigma_0^{-1})_{sj}) \frac{\partial}{\partial \beta_2} (\sigma^2 (\Sigma_0^{-1})_{kt} (\Sigma_0^{-1})_{ul}) E [\Xi_{rs} \Xi_{tu}] \\ & = \Delta_0^2 \text{Var}[\xi] \sum_{r=1}^N \frac{\partial}{\partial \beta_1} (\sigma^2 (\Sigma_0^{-1})_{ir} (\Sigma_0^{-1})_{rj}) \frac{\partial}{\partial \beta_2} (\sigma^2 (\Sigma_0^{-1})_{kr} (\Sigma_0^{-1})_{rl}) \end{aligned}$$

with the third equality following from the interchangeability of unconditional expectations and differentiation with respect to β , and the fourth from the fact that $E [\Xi_{rs} \Xi_{tu}] \neq 0$ only when $r = s = t = u$, and

$$E [\Xi_{rr} \Xi_{rr}] = \Delta_0^2 \text{Var}[\xi].$$

Thus we have

$$\begin{aligned}
\lambda^{ijkl} &= \Delta_0^2 Var[\xi] \sum_{r=1}^N \frac{\partial}{\partial \beta_1} (\sigma^2 \gamma^{-4} (V^{-1})_{ir} (V^{-1})_{rj}) \frac{\partial}{\partial \beta_2} (\sigma^2 \gamma^{-4} (V^{-1})_{kr} (V^{-1})_{rl}) \\
&= \Delta_0^2 Var[\xi] \sum_{r=1}^N \frac{\partial}{\partial \beta_1} (\sigma^2 \gamma^{-4} v^{i,r} v^{r,j}) \frac{\partial}{\partial \beta_2} (\sigma^2 \gamma^{-4} v^{k,r} v^{r,l}).
\end{aligned} \tag{G.7}$$

and

$$\begin{aligned}
\frac{\partial}{\partial \beta_1} (\sigma^2 \gamma^{-4} v^{i,r} v^{r,j}) \frac{\partial}{\partial \beta_2} (\sigma^2 \gamma^{-4} v^{k,r} v^{r,l}) &= \left(\frac{\partial (\sigma^2 \gamma^{-4})}{\partial \beta_1} v^{i,r} v^{r,j} + \sigma^2 \gamma^{-4} \frac{\partial (v^{i,r} v^{r,j})}{\partial \eta} \frac{\partial \eta}{\partial \beta_1} \right) \\
&\quad \times \left(\frac{\partial (\sigma^2 \gamma^{-4})}{\partial \beta_2} v^{k,r} v^{r,l} + \sigma^2 \gamma^{-4} \frac{\partial (v^{k,r} v^{r,l})}{\partial \eta} \frac{\partial \eta}{\partial \beta_2} \right).
\end{aligned}$$

Summing these terms, we obtain

$$\begin{aligned}
\tilde{\psi}(\lambda) &= \frac{\Delta_0^2 Var[\xi] \ 2N}{(1-\eta)^7 (1+\eta)^3 (1+\eta^2)^3} (C_{1\lambda} (1-\eta^4) (2C_{5\lambda} C_{6\lambda} (1+\eta+\eta^2+2\eta^3) + C_{4\lambda} (1-\eta^4)) \\
&\quad + 2C_{2\lambda} C_{3\lambda} (2C_{5\lambda} C_{6\lambda} (1+2\eta+4\eta^2+6\eta^3+5\eta^4+4\eta^5+4\eta^6) \\
&\quad + C_{4\lambda} (1+\eta+\eta^2+2\eta^3-\eta^4-\eta^5-\eta^6-2\eta^7))) \\
&\quad + o(N)
\end{aligned}$$

where

$$\begin{aligned}
C_{1\lambda} &= \frac{\partial (\sigma^2 \gamma^{-4})}{\partial \beta_1}, \quad C_{2\lambda} = \sigma^2 \gamma^{-4}, \quad C_{3\lambda} = \frac{\partial \eta}{\partial \beta_1} \\
C_{4\lambda} &= \frac{\partial (\sigma^2 \gamma^{-4})}{\partial \beta_2}, \quad C_{5\lambda} = \sigma^2 \gamma^{-4}, \quad C_{6\lambda} = \frac{\partial \eta}{\partial \beta_2}
\end{aligned}$$

Putting it all together, we have

$$E \left[\psi \left(\frac{\partial \Sigma^{-1}}{\partial \beta_1}, \frac{\partial \Sigma^{-1}}{\partial \beta_2} \right) \right] = \psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial \Sigma_0^{-1}}{\partial \beta_2} \right) + \epsilon^2 (\alpha_1[2] + \alpha_2) + O(\epsilon^3)$$

Finally, the asymptotic variance of the estimator $(\hat{\sigma}^2, \hat{a}^2)$ is given by

$$AVAR_{\text{true}}(\hat{\sigma}^2, \hat{a}^2) = E[\Delta] (D' S^{-1} D)^{-1} \tag{G.8}$$

where

$$\begin{aligned}
D &= D' = -\frac{1}{N} E_{\text{true}} [\dot{l}] = -\frac{1}{N} E_{\text{normal}} [\dot{l}] = \frac{1}{N} E_{\text{normal}} [\dot{l}l'] \\
&\equiv F^{(0)} + \epsilon^2 F^{(2)} + O(\epsilon^3)
\end{aligned}$$

is given by the expression in the correctly specified case (F.15), with $F^{(0)}$ and $F^{(2)}$ given in (F.16) and (F.17) respectively. Also, in light of (G.1), we have

$$S = \frac{1}{N} E_{\text{true}} [\ddot{l}l'] = -\frac{1}{N} E_{\text{true}} [\ddot{l}] + Cum_4(U) \Psi = D + Cum_4(U) \Psi$$

where

$$\begin{aligned}
\Psi &= \frac{1}{4N} \left(\begin{array}{c} E \left[\psi \left(\frac{\partial \Sigma^{-1}}{\partial \sigma^2}, \frac{\partial \Sigma^{-1}}{\partial \sigma^2} \right) \right] \\ \bullet \\ E \left[\psi \left(\frac{\partial \Sigma^{-1}}{\partial a^2}, \frac{\partial \Sigma^{-1}}{\partial a^2} \right) \right] \end{array} \right) \\
&\equiv \Psi^{(0)} + \epsilon^2 \Psi^{(2)} + O(\epsilon^3).
\end{aligned}$$

Since, from (G.3), we have

$$E \left[\psi \left(\frac{\partial \Sigma^{-1}}{\partial \beta_1}, \frac{\partial \Sigma^{-1}}{\partial \beta_2} \right) \right] = \psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial \Sigma_0^{-1}}{\partial \beta_2} \right) + \epsilon^2 \alpha_1[2] + \epsilon^2 \alpha_2 + O(\epsilon^3),$$

it follows that $\Psi^{(0)}$ is the matrix with entries $\frac{1}{4N} \psi \left(\frac{\partial \Sigma_0^{-1}}{\partial \beta_1}, \frac{\partial \Sigma_0^{-1}}{\partial \beta_2} \right)$, i.e.,

$$\Psi^{(0)} = \begin{pmatrix} \frac{\Delta_0}{\sigma^2(4a^2 + \Delta_0 \sigma^2)^3} & \frac{\Delta_0^{1/2}}{2a^4} \left(\frac{1}{\sigma(4a^2 + \Delta_0 \sigma^2)^{3/2}} - \frac{\Delta_0^{1/2}(6a^2 + \Delta_0 \sigma^2)}{(4a^2 + \Delta_0 \sigma^2)^3} \right) \\ \bullet & \frac{1}{2a^8} \left(1 - \frac{\Delta_0^{1/2} \sigma(6a^2 + \Delta_0 \sigma^2)}{(4a^2 + \Delta_0 \sigma^2)^{3/2}} - \frac{2a^4(16a^2 + 3\Delta_0 \sigma^2)}{(4a^2 + \Delta_0 \sigma^2)^3} \right) \end{pmatrix} + o(1),$$

and

$$\Psi^{(2)} = \frac{1}{4N} (\alpha_1[2] + \alpha_2).$$

with

$$\frac{1}{4N} \alpha_1[2] = \text{Var}[\xi] \begin{pmatrix} \frac{2\Delta_0^{3/2}(-4a^2 + \Delta_0 \sigma^2)}{\sigma(4a^2 + \Delta_0 \sigma^2)^{9/2}} & \frac{\Delta_0 \left((-4a^2 + \Delta_0 \sigma^2)(4a^2 + \Delta_0 \sigma^2)^{3/2} - \Delta_0^{1/2} \sigma(-40a^4 + 2a^2 \Delta_0 \sigma^2 + \Delta_0^2 \sigma^4) \right)}{\frac{2a^4(4a^2 + \Delta_0 \sigma^2)^{9/2}}{8\Delta_0 \sigma^2 \left((4a^2 + \Delta_0 \sigma^2)^{3/2} - \Delta_0^{1/2} \sigma(6a^2 + \Delta_0 \sigma^2) \right)}} \\ \bullet & \frac{a^4(4a^2 + \Delta_0 \sigma^2)^{9/2}}{a^4(4a^2 + \Delta_0 \sigma^2)^{9/2}} \end{pmatrix} + o(1),$$

$$\frac{1}{4N} \alpha_2 = \text{Var}[\xi] \begin{pmatrix} \frac{\Delta_0^{3/2}(40a^8 - 12a^4 \Delta_0^2 \sigma^4 + \Delta_0^4 \sigma^8)}{2\sigma(2a^2 + \Delta_0 \sigma^2)^3(4a^2 + \Delta_0 \sigma^2)^{9/2}} & \frac{\Delta_0^{3/2} \sigma(-44a^6 - 18a^4 \Delta_0 \sigma^2 + 7a^2 \Delta_0^2 \sigma^6 + 3\Delta_0^3 \sigma^6)}{\frac{(2a^2 + \Delta_0 \sigma^2)^3(4a^2 + \Delta_0 \sigma^2)^{9/2}}{2\Delta_0^{3/2} \sigma^3(50a^2 + 42a^2 \Delta_0 \sigma^2 + 9\Delta_0^2 \sigma^4)}} \\ \bullet & \frac{2\Delta_0^{3/2} \sigma^3(50a^2 + 42a^2 \Delta_0 \sigma^2 + 9\Delta_0^2 \sigma^4)}{(2a^2 + \Delta_0 \sigma^2)^3(4a^2 + \Delta_0 \sigma^2)^{9/2}} \end{pmatrix} + o(1).$$

It follows from (G.8) that

$$\begin{aligned} AVAR_{\text{true}}(\hat{\sigma}^2, \hat{a}^2) &= E[\Delta] \left(D(D + \text{Cum}_4(U) \Psi)^{-1} D \right)^{-1} \\ &= \Delta_0 \left(D(Id + \text{Cum}_4(U) D^{-1} \Psi)^{-1} \right)^{-1} \\ &= \Delta_0 (Id + \text{Cum}_4(U) D^{-1} \Psi) D^{-1} \\ &= \Delta_0 (Id + \text{Cum}_4(U) D^{-1} \Psi) D^{-1} \\ &= AVAR_{\text{normal}}(\hat{\sigma}^2, \hat{a}^2) + \Delta_0 \text{Cum}_4(U) D^{-1} \Psi D^{-1} \end{aligned}$$

where

$$AVAR_{\text{normal}}(\hat{\sigma}^2, \hat{a}^2) = \Delta_0 D^{-1} = A^{(0)} + \epsilon^2 A^{(2)} + O(\epsilon^3)$$

is the result given in Theorem 2, namely (7.6).

The correction term due to the misspecification of the error distribution is determined by $Cum_4(U)$ times

$$\begin{aligned}
\Delta_0 D^{-1} \Psi D^{-1} &= \Delta_0 \left(F^{(0)} + \epsilon^2 F^{(2)} + O(\epsilon^3) \right)^{-1} \left(\Psi^{(0)} + \epsilon^2 \Psi^{(2)} + O(\epsilon^3) \right) \\
&\quad \times \left(F^{(0)} + \epsilon^2 F^{(2)} + O(\epsilon^3) \right)^{-1} \\
&= \Delta_0 \left(Id - \epsilon^2 \left[F^{(0)} \right]^{-1} F^{(2)} + O(\epsilon^3) \right) \left[F^{(0)} \right]^{-1} \left(\Psi^{(0)} + \epsilon^2 \Psi^{(2)} + O(\epsilon^3) \right) \\
&\quad \times \left(Id - \epsilon^2 \left[F^{(0)} \right]^{-1} F^{(2)} + O(\epsilon^3) \right) \left[F^{(0)} \right]^{-1} \\
&= \Delta_0 \left[F^{(0)} \right]^{-1} \Psi^{(0)} \left[F^{(0)} \right]^{-1} + \\
&\quad + \epsilon^2 \Delta_0 \left(\left[F^{(0)} \right]^{-1} \Psi^{(2)} \left[F^{(0)} \right]^{-1} - \left[F^{(0)} \right]^{-1} F^{(2)} \left[F^{(0)} \right]^{-1} \Psi^{(0)} \left[F^{(0)} \right]^{-1} \right. \\
&\quad \quad \left. - \left[F^{(0)} \right]^{-1} \Psi^{(0)} \left[F^{(0)} \right]^{-1} F^{(2)} \left[F^{(0)} \right]^{-1} \right) \\
&\quad + O(\epsilon^3) \\
&\equiv B^{(0)} + \epsilon^2 B^{(2)} + O(\epsilon^3).
\end{aligned}$$

where the matrices The asymptotic variance is then given by

$$AVAR_{\text{true}}(\hat{\sigma}^2, \hat{a}^2) = \left(A^{(0)} + Cum_4(U)B^{(0)} \right) + \epsilon^2 \left(A^{(2)} + Cum_4(U)B^{(2)} \right) + O(\epsilon^3).$$

with the terms $A^{(0)}$, $A^{(2)}$, $B^{(0)}$ and $B^{(2)}$ given in the statement of the Theorem.

Appendix H: Proof of Proposition 3

From

$$E[Y_i^2] = E[w_i^2] + E[u_i^2] = \sigma^2 \Delta + \frac{c^2(1 - e^{-b\Delta})}{b}$$

it follows that the estimator (2.2) has the following expected value

$$\begin{aligned}
E[\hat{\sigma}^2] &= \frac{1}{T} \sum_{i=1}^N E[Y_i^2] \\
&= \frac{N}{T} \left(\sigma^2 \Delta + \frac{c^2(1 - e^{-b\Delta})}{b} \right) \\
&= \sigma^2 + \frac{c^2(1 - e^{-b\Delta})}{b\Delta} \\
&= (\sigma^2 + c^2) - \frac{bc^2}{2} \Delta + O(\Delta^2). \tag{H.1}
\end{aligned}$$

The estimator's variance is

$$\begin{aligned}
Var[\hat{\sigma}^2] &= \frac{1}{T^2} Var \left[\sum_{i=1}^N Y_i^2 \right] \\
&= \frac{1}{T^2} \sum_{i=1}^N Var[Y_i^2] + \frac{2}{T^2} \sum_{i=1}^N \sum_{j=1}^{i-1} cov(Y_i^2, Y_j^2)
\end{aligned}$$

Since the Y_i 's are normal with mean zero,

$$\text{Var} [Y_i^2] = 2\text{Var}[Y_i]^2 = 2E [Y_i^2]^2$$

and for $i > j$

$$\text{cov} (Y_i^2, Y_j^2) = 2 \text{cov} (Y_i, Y_j)^2 = 2 E [u_i u_j]^2$$

since

$$\text{cov} (Y_i, Y_j) = E [Y_i Y_j] = E [(w_i + u_i) (w_j + u_j)] = E [u_i u_j].$$

Now we have

$$\begin{aligned} E [u_i u_j] &= E [(U_{\tau_i} - U_{\tau_{i-1}}) (U_{\tau_j} - U_{\tau_{j-1}})] \\ &= E [U_{\tau_i} U_{\tau_j}] - E [U_{\tau_i} U_{\tau_{j-1}}] - E [U_{\tau_{i-1}} U_{\tau_j}] + E [U_{\tau_{i-1}} U_{\tau_{j-1}}] \\ &= -\frac{c^2 (1 - e^{-b\Delta})^2 e^{-b\Delta(i-j-1)}}{2b} \end{aligned}$$

so that

$$\begin{aligned} \text{cov} (Y_i^2, Y_j^2) &= 2 \left(-\frac{c^2 (1 - e^{-b\Delta})^2 e^{-b\Delta(i-j-1)}}{2b} \right)^2 \\ &= \frac{c^4 e^{-2b\Delta(i-j-1)} (1 - e^{-b\Delta})^4}{2b^2} \end{aligned}$$

and consequently

$$\text{Var} [\hat{\sigma}^2] = \frac{1}{T^2} \left\{ \frac{c^4 (1 - e^{-b\Delta})^2 (N e^{-2b\Delta} - 1 + e^{-2Nb\Delta})}{b^2 (1 + e^{-b\Delta})^2} + 2N \left(\sigma^2 \Delta + \frac{c^2 (1 - e^{-b\Delta})}{b} \right)^2 \right\} \quad (\text{H.2})$$

with $N = T/\Delta$. The RMSE expression follows from (H.1) and (H.2). As in Proposition 1, these are exact small sample expressions, valid for all (T, Δ) .

References

- Aït-Sahalia, Y., 1996, “Nonparametric Pricing of Interest Rate Derivative Securities,” *Econometrica*, 64, 527–560.
- , 2002, “Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach,” *Econometrica*, 70, 223–262.
- Aït-Sahalia, Y., and P. A. Mykland, 2003, “The Effects of Random and Discrete Sampling When Estimating Continuous-Time Diffusions,” *Econometrica*, 71, 483–549.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2003, “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71, 579–625.
- Bandi, F. M., and P. C. B. Phillips, 2003, “Fully Nonparametric Estimation of Scalar Diffusion Models,” *Econometrica*, 71, 241–283.
- Barndorff-Nielsen, O. E., and N. Shephard, 2002, “Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models,” *Journal of Royal Statistical Society, B*, 64, 253–280.
- Black, F., 1986, “Noise,” *Journal of Finance*, 41, 529–543.
- Delattre, S., and J. Jacod, 1997, “A Central Limit Theorem for Normalized Functions of the Increments of a Diffusion Process, in the Presence of Round-Off Errors,” *Bernoulli*, 3, 1–28.
- Durbin, J., 1959, “Efficient Estimation of Parameters in Moving-Average Models,” *Biometrika*, 46, 306–316.
- Glosten, L. R., 1987, “Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices,” *Journal of Finance*, 42, 1293–1307.
- Gloter, A., and J. Jacod, 2000, “Diffusions with Measurement Errors: I - Local Asymptotic Normality and II - Optimal Estimators,” Working paper, Université de Paris VI.
- Haddad, J., 1995, “On the Closed Form of the Likelihood Function of the First Order Moving Average Model,” *Biometrika*, 82, 232–234.
- Hansen, L. P., and J. A. Scheinkman, 1995, “Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes,” *Econometrica*, 63, 767–804.
- Hasbrouck, J., 1993, “Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement,” *Review of Financial Studies*, 6, 191–212.
- Heyde, C. C., 1997, *Quasi-Likelihood and Its Application*. Springer-Verlag, New York.
- Jacod, J., 1996, “La Variation Quadratique du Brownien en Présence d’Erreurs d’Arrondi,” *Astérisque*, 236, 155–162.
- Macurdy, T. E., 1982, “The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis,” *Journal of Econometrics*, 18, 83–114.
- Madhavan, A., M. Richardson, and M. Roomans, 1997, “Why Do Security Prices Change?,” *Review of Financial Studies*, 10, 1035–1064.

- McCullagh, P., 1987, *Tensor Methods in Statistics*. Chapman and Hall, London, U.K.
- McCullagh, P., and J. Nelder, 1989, *Generalized Linear Models*. Chapman and Hall, London, U.K., second edn.
- Merton, R. C., 1980, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.
- Roll, R., 1984, "A Simple Model of the Implicit Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39, 1127–1139.
- Shaman, P., 1969, "On the Inverse of the Covariance Matrix of a First Order Moving Average," *Biometrika*, 56, 595–600.
- White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

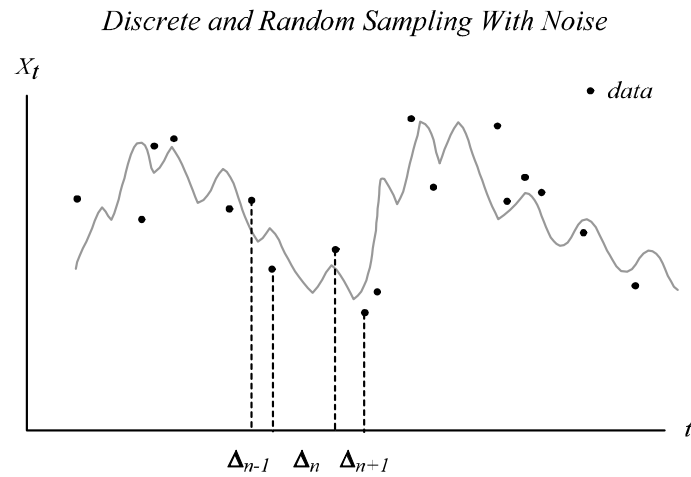
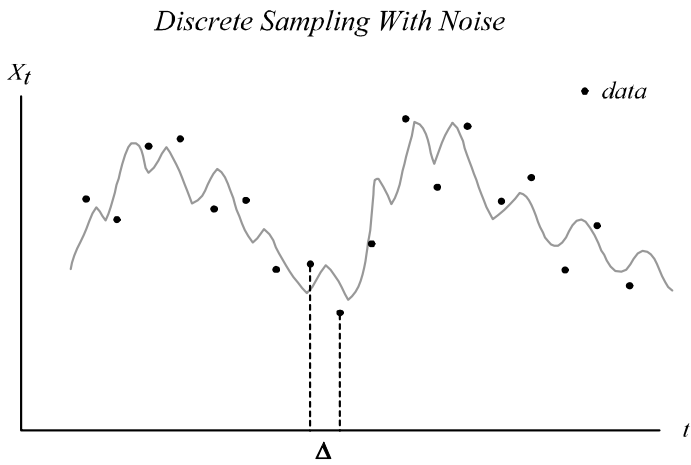
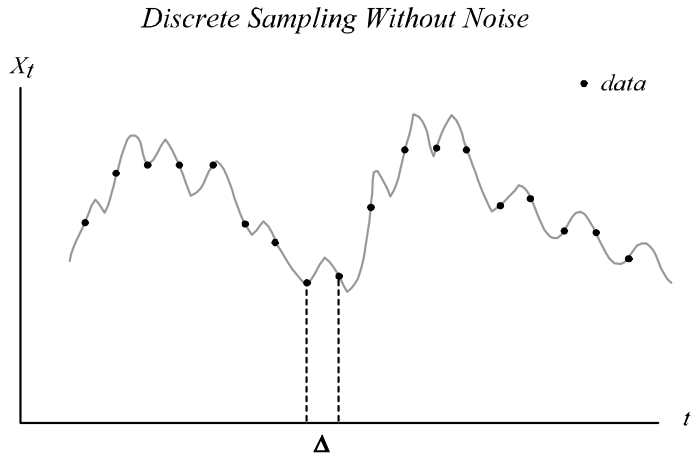


Figure 1
Various Discrete Sampling Modes – No Noise (Section 2), With Noise (Sections 3-6)
and Randomly Spaced with Noise (Section 7).

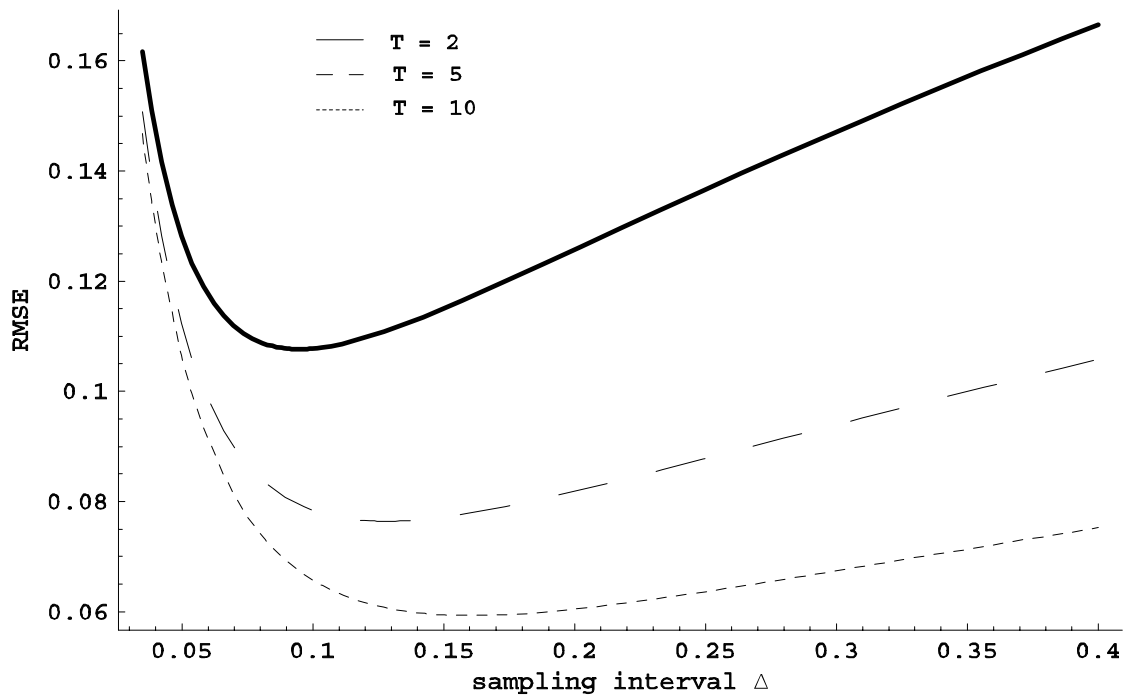


Figure 2
RMSE of the Estimator $\hat{\sigma}^2$ When the Presence of the Noise is Ignored

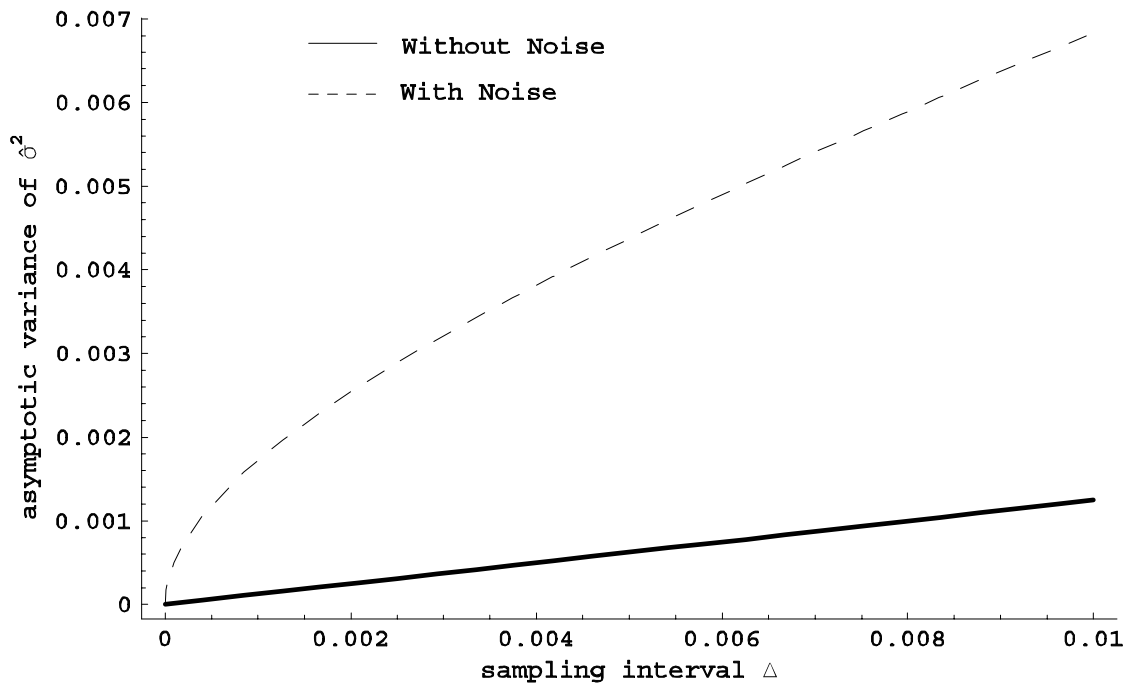


Figure 3
Comparison of the Asymptotic Variances of the MLE $\hat{\sigma}^2$ Without and With Noise
Taken into Account

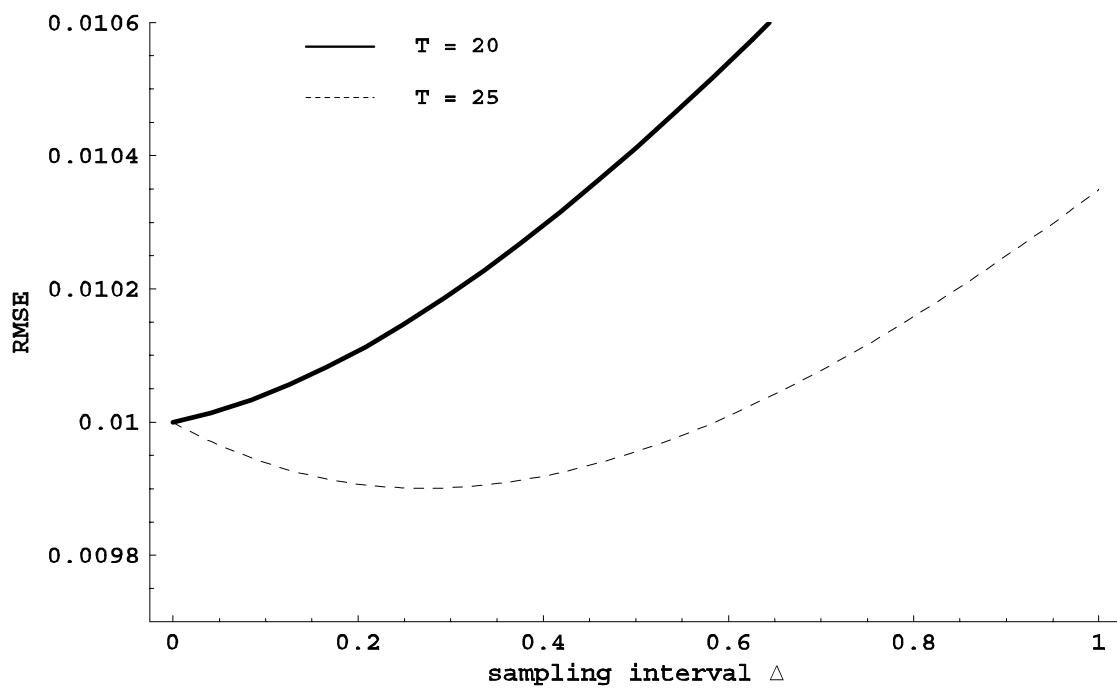


Figure 4
RMSE of the Estimator $\hat{\sigma}^2$ When the Presence of Serially Correlated Noise is Ignored