1

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

Yacine Aït-Sahalia, Per A. Mykland, and Lan Zhang

[1] Bendheim Center for Finance, Princeton University, Princeton, NJ 08540 and NBER
yacine@princeton.edu
[2] Department of Statistics, The University of Chicago, Chicago, IL 60637
mykland@uchicago.edu
[3] Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213
lzhang@stat.cmu.edu

Note: Reprint of: Yacine Aït-Sahalia, Per A. Mykland and Lan Zhang, *How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise*, Review of Financial Studies, published in Volume 18 **2** (2005), pages 351–416, Oxford University Press (by permission of Oxford University Press).

Summary. In theory, the sum of squares of log returns sampled at high frequency estimates their variance. When market microstructure noise is present but unaccounted for, however, we show that the optimal sampling frequency is finite and derive its closed-form expression. But even with optimal sampling, using say five minute returns when transactions are recorded every second, a vast amount of data is discarded, in contradiction to basic statistical principles. We demonstrate that modelling the noise and using all the data is a better solution, even if one misspecifies the noise distribution. So the answer is: sample as often as possible.

Over the past few years, price data sampled at very high frequency have become increasingly available, in the form of the Olsen dataset of currency exchange rates or the TAQ database of NYSE stocks. If such data were not affected by market microstructure noise, the realized volatility of the process

* We are grateful for comments and suggestions from the editor, Maureen O'Hara, and two anonymous referees, as well as seminar participants at Berkeley, Harvard, NYU, MIT, Stanford, the Econometric Society and the Joint Statistical Meetings. Financial support from the NSF under grants SBR-0111140 (Aït-Sahalia), DMS-0204639 (Mykland and Zhang) and the NIH under grant RO1 AG023141-01 (Zhang) is also gratefully acknowledged.

(i.e., the average sum of squares of log-returns sampled at high frequency) would estimate the returns' variance, as is well known. In fact, sampling as often as possible would theoretically produce in the limit a perfect estimate of that variance.

We start by asking whether it remains optimal to sample the price process at very high frequency in the presence of market microstructure noise, consistently with the basic statistical principle that, ceteris paribus, more data is preferred to less. We first show that, if noise is present but unaccounted for, then the optimal sampling frequency is finite, and we derive a closed-form formula for it. 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 is the right balance to strike? 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) so that some time aggregation of the returns data is advisable.

By providing a quantitative answer to the question of how often one should sample, we hope to reduce the arbitrariness of the choices that have been made in the empirical literature using high frequency data: for example, using essentially the same Olsen exchange rate series, these somewhat ad hoc choices range from 5 minute intervals (e.g., [5], [8] and [19]) to as long as 30 minutes (e.g., [6]). When calibrating our analysis to the amount of microstructure noise that has been reported in the literature, we demonstrate how the optimal sampling interval should be determined: for instance, depending upon the amount of microstructure noise relative to the variance of the underlying returns, the optimal sampling frequency varies from 4 minutes to 3 hours, if 1 day's worth of data is used at a time. If a longer time period is used in the analysis, then the optimal sampling frequency can be considerably longer than these values.

But even if one determines the sampling frequency optimally, it remains the case that the empirical researcher is not making use of the full data at his/her disposal. For instance, suppose that we have available transaction records on a liquid stock, traded once every second. Over a typical 6.5 hour day, we therefore start with $23,400$ observations. If one decides to sample once every 5 minutes, then – whether or not this is the optimal sampling frequency – this amounts to retaining only 78 observations. Said differently, one is throwing away 299 out of every 300 transactions. From a statistical perspective, this is unlikely to be the optimal solution, even though it is undoubtedly better than computing a volatility estimate using noisy squared log-returns sampled every second. Somehow, an optimal solution should make use of all the data, and this is where our analysis goes next.

So, if 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. This will involve an estimator different from the simple sum of squared log-returns. Since we work within a fully *parametric* framework, *likelihood* is the key word. Hence we construct the likelihood function for the observed log-returns, which include microstructure noise. To do so, we must postulate a model for the noise term. We assume that the noise is Gaussian. In light of what we know from the sophisticated theoretical microstructure literature, this is likely to be overly simplistic and one may well be concerned about the effect(s) of this assumption. Could it do more harm than good? Surprisingly, we demonstrate that our likelihood correction, based on Gaussianity of the noise, works *even if* one misspecifies the assumed distribution of the noise term. Specifically, 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. This robustness result is, we think, 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.

In other words, the answer to the question we pose in our title is "as often as possible", provided one accounts for the presence of the noise when designing the estimator (and we suggest maximum likelihood as a means of doing so). If one is unwilling to account for the noise, then the answer is to rely on the finite optimal sampling frequency we start our analysis with, but we stress that while it is optimal if one insists upon using sums of squares of log-returns, this is not the best possible approach to estimate volatility given the complete high frequency dataset at hand.

In a companion paper ([43]), we study the corresponding *nonparametric* problem, where the volatility of the underlying price is a stochastic process, and nothing else is known about it, in particular no parametric structure. In that case, the object of interest is the integrated volatility of the process over a fixed time interval, such as a day, and we show how to estimate it using again all the data available (instead of sparse sampling at an arbitrarily lower frequency of, say, 5 minutes). Since the model is nonparametric, we no longer use a likelihood approach but instead propose a solution based on subsampling and averaging, which involves estimators constructed on two different time scales, and demonstrate that this again dominates sampling at a lower frequency, whether arbitrary or optimally determined.

This paper is organized as follows. We start by describing in Section 1.1 our reduced form setup and the underlying structural models that support it. We then review in Section 1.2 the base case where no noise is present, before analyzing in Section 1.3 the situation where the presence of the noise is ignored. In Section 1.4, we examine the concrete implications of this result for empirical work with high frequency data. Next, we show in Section 1.5 that accounting for the presence of the noise through the likelihood restores the optimality of high frequency sampling. Our robustness results are presented in Section 1.6 and interpreted in Section 1.7. We study the same questions when the observations are sampled at random time intervals, which are an essential feature of transaction-level data, in Section 1.8. We then turn to various extensions and relaxation of our assumptions in Section 1.9: we add a drift term, then serially correlated and cross-correlated noise respectively. Section 1.10 concludes. All proofs are in the Appendix.

1.1 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, $\mu(.,.)$ 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 usual, the restriction that σ is constant is without loss of generality since in the univariate case a one-to-one transformation can always reduce a known specification $\sigma(X_t)$ to that case. Also, as discussed in [4], the properties of parametric estimators in this model are quite different depending upon whether 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 [38]). So we simplify the analysis one step further by setting $\mu = 0$, which we do until Section 1.9.1, where we then show that adding a drift term does not alter our results. In Section 1.9.4, we discuss the situation where the instantaneous volatility σ is stochastic.

But for now,

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

Until Section 1.8, we treat the case where the 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 $\tau_0 = 0$, $\tau_1 = \Delta, \ldots, \tau_N = N\Delta = T$. In Section 1.8, we let the sampling intervals be themselves random variables, since this feature is an essential characteristic of high frequency transaction data.

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 [10]). So, where we depart from the inference setup previously studied in [4] 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 $U'_{\tau_i}s$ are i.i.d. 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., all summarized into the term U. That these phenomena are real are 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. The efficient price process X is typically modelled as a random walk, i.e., the discrete time equivalent of (1.2). Our specification coincides with that of [29], who discusses the theoretical market microstructure underpinnings of such a model and argues that the parameter a is a summary measure of market quality. Structural market microstructure models do generate (1.3). For instance, [39] proposes a model where U is due entirely to the bid-ask spread. [28] notes that in practice there are sources of noise other than just the bid-ask spread, and studies their effect on the Roll model and its estimators.

Indeed, a disturbance U can also be generated by adverse selection effects as in [20] and [21], 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. When asymmetric information is involved, the disturbance U would typically 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 1.9.2 and 1.9.3 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 modelled as a rounding off problem (see [14], [23] and [31]). The specification of the model in [27] combines both the rounding and bid-ask effects as the dual sources of the noise term U. Finally, structural models, such as that of [35], 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 1.1 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.

1.2 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 $Y_i = \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}}$, $i = 1, ..., N$. The observations $Y_i = \sigma \left(W_{\tau_{i+1}} - W_{\tau_i} \right)$ are then i.i.d. $N(0, \sigma^2 \Delta)$ so the likelihood function is

$$l(\sigma^2) = -N \ln(2\pi\sigma^2\Delta)/2 - (2\sigma^2\Delta)^{-1} Y'Y \,, \tag{1.4}$$

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

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^{N} Y_i^2 \tag{1.5}$$

which has the following exact small sample moments:

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

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

and the following asymptotic distribution

$$T^{1/2}\left(\hat{\sigma}^2 - \sigma^2\right) \underset{T \longrightarrow \infty}{\longrightarrow} N(0, \omega), \tag{1.6}$$

where

$$\omega = \mathrm{AVAR}(\hat{\sigma}^2) = \Delta E\left[-\ddot{l}(\sigma^2)\right]^{-1} = 2\sigma^4 \Delta. \tag{1.7}$$

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

1.3 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 (1.4) even though the true structure of the observed log-returns $Y_i's$ is given by an MA(1) process since

$$\begin{aligned} Y_i &= \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}} \\ &= X_{\tau_i} - X_{\tau_{i-1}} + U_{\tau_i} - U_{\tau_{i-1}} \\ &= \sigma\left(W_{\tau_i} - W_{\tau_{i-1}}\right) + U_{\tau_i} - U_{\tau_{i-1}} \\ &\equiv \varepsilon_i + \eta \varepsilon_{i-1}, \end{aligned} \tag{1.8}$$

where the $\varepsilon_i's$ are uncorrelated with mean zero and variance γ^2 (if the $U's$ are normally distributed, then the $\varepsilon_i's$ are i.i.d.). The relationship to the original parametrization (σ^2, a^2) is given by

$$\gamma^2(1 + \eta^2) = \mathrm{Var}[Y_i] = \sigma^2 \Delta + 2a^2 \tag{1.9}$$

$$\gamma^2 \eta = \mathrm{Cov}(Y_i, Y_{i-1}) = -a^2. \tag{1.10}$$

Equivalently, the inverse change of variable is given by

$$\gamma^2 = \frac{1}{2}\left\{2a^2 + \sigma^2\Delta + \sqrt{\sigma^2\Delta\left(4a^2 + \sigma^2\Delta\right)}\right\} \tag{1.11}$$

$$\eta = \frac{1}{2a^2}\left\{-2a^2 - \sigma^2\Delta + \sqrt{\sigma^2\Delta\left(4a^2 + \sigma^2\Delta\right)}\right\}. \tag{1.12}$$

Two important properties of the log-returns Y_i's emerge from the two equations (1.9)-(1.10). First, it is clear from (1.9) that microstructure noise leads to spurious variance in observed log-returns, $\sigma^2\Delta + 2a^2$ vs. $\sigma^2\Delta$. This is consistent with the predictions of theoretical microstructure models. For instance, [16] develop a model linking the arrival of information, the timing of trades, and the resulting price process. In their model, the transaction price will be a biased representation of the efficient price process, with a variance that is both overstated and heteroskedastic as a result of the fact that transactions (hence the recording of an observation on the process $\tilde{X}$) occur at intervals that are time-varying. While our specification is too simple to capture the rich joint dynamics of price and sampling times predicted by their model, heteroskedasticity of the observed variance will also arise in our case once we allow for time variation of the sampling intervals (see Section 1.8 below).

In our model, the proportion of the total return variance that is market microstructure-induced is

$$\pi = \frac{2a^2}{\sigma^2\Delta + 2a^2} \tag{1.13}$$

at observation interval Δ. As Δ gets smaller, π gets closer to 1, so that a larger proportion of the variance in the observed log-return is driven by market microstructure frictions, and correspondingly a lesser fraction reflects the volatility of the underlying price process X.

Second, (1.10) implies that $-1 < \eta < 0$, so that log-returns are (negatively) autocorrelated with first order autocorrelation $-a^2/(\sigma^2\Delta + 2a^2) = -\pi/2$. It has been noted that market microstructure noise has the potential to explain the empirical autocorrelation of returns. For instance, in the simple Roll model, $U_t = (s/2)Q_t$ where s is the bid/ask spread and Q_t, the order flow indicator, is a binomial variable that takes the values $+1$ and -1 with equal probability. Therefore $\mathrm{Var}[U_t] = a^2 = s^2/4$. Since $\mathrm{Cov}(Y_i, Y_{i-1}) = -a^2$, the bid/ask spread can be recovered in this model as $s = 2\sqrt{-\rho}$ where $\rho = \gamma^2\eta$ is the first order autocorrelation of returns. [18] proposed to adjust variance estimates to control for such autocorrelation and [28] studied the resulting estimators. In [41], U arises because of the strategic trading of institutional investors which is then put forward as an explanation for the observed serial correlation of returns. [33] show that infrequent trading has implications for the variance and autocorrelations of returns. Other empirical patterns in high frequency financial data have been documented: leptokurtosis, deterministic patterns and volatility clustering.

Our first result shows that the optimal sampling frequency is finite when noise is present but unaccounted for. The estimator $\hat{\sigma}^2$ obtained from maximizing the misspecified log-likelihood (1.4) is quadratic in the $Y_i's$: see (1.5). In order to obtain its exact (i.e., small sample) variance, we therefore need to calculate the fourth order cumulants of the $Y_i's$ since

$$\mathrm{Cov}\left(Y_i^2, Y_j^2\right) = 2\,\mathrm{Cov}\left(Y_i, Y_j\right)^2 + \mathrm{Cum}(Y_i, Y_i, Y_j, Y_j) \tag{1.14}$$

(see e.g., Section 2.3 of [36] for definitions and properties of the cumulants). We have:

Lemma 1. *The fourth cumulants of the log-returns are given by*

$$\mathrm{Cum}(Y_i, Y_j, Y_k, Y_l) = \begin{cases} 2\ \mathrm{Cum}_4\left[U\right] & \text{if } i = j = k = l\,, \\ (-1)^{s(i,j,k,l)}\,\mathrm{Cum}_4\left[U\right], & \text{if } \max(i,j,k,l) = \min(i,j,k,l) + 1, \\ 0 & \text{otherwise}\,, \end{cases} \tag{1.15}$$

where $s(i,j,k,l)$ denotes the number of indices among (i,j,k,l) that are equal to $\min(i,j,k,l)$ and U denotes a generic random variable with the common distribution of the $U_{\tau_i}'s$. Its fourth cumulant is denoted $\mathrm{Cum}_4\left[U\right]$.

Now U has mean zero, so in terms of its moments

$$\mathrm{Cum}_4\left[U\right] = E\left[U^4\right] - 3\left(E\left[U^2\right]\right)^2. \tag{1.16}$$

In the special case where U is normally distributed, $\mathrm{Cum}_4\left[U\right] = 0$ and as a result of (1.14) the fourth cumulants of the log-returns are all 0 (since W is normal, the log-returns are also normal in that case). If the distribution of U is binomial as in the simple bid/ask model described above, then $\mathrm{Cum}_4\left[U\right] = -s^4/8$; since in general s will be a tiny percentage of the asset price, say $s = 0.05\%$, the resulting $\mathrm{Cum}_4\left[U\right]$ will be very small.

We can now characterize the root mean squared error

$$RMSE\left[\hat{\sigma}^2\right] = \left(\left(E\left[\hat{\sigma}^2\right] - \sigma^2\right)^2 + \mathrm{Var}\left[\hat{\sigma}^2\right]\right)^{1/2}$$

of the estimator:

Theorem 1. *In small samples (finite T), the bias and variance of the estimator $\hat{\sigma}^2$ are given by*

$$E\left[\hat{\sigma}^2\right] - \sigma^2 = \frac{2a^2}{\Delta}\,, \tag{1.17}$$

$$\begin{aligned}\mathrm{Var}\left[\hat{\sigma}^2\right] = & \frac{2\left(\sigma^4\Delta^2 + 4\sigma^2\Delta a^2 + 6a^4 + 2\,\mathrm{Cum}_4\left[U\right]\right)}{T\Delta} - \\ & - \frac{2\left(2a^4 + \mathrm{Cum}_4\left[U\right]\right)}{T^2}.\end{aligned} \tag{1.18}$$

Its RMSE has a unique minimum in Δ which is reached at the optimal sampling interval

$$\begin{aligned}\Delta^* = & \left(\frac{2a^4T}{\sigma^4}\right)^{1/3}\left(\left(1 - \sqrt{1 - \frac{2\left(3a^4 + \mathrm{Cum}_4\left[U\right]\right)^3}{27\sigma^4a^8T^2}}\right)^{1/3}\right. \\ & \left. + \left(1 + \sqrt{1 - \frac{2\left(3a^4 + \mathrm{Cum}_4\left[U\right]\right)^3}{27\sigma^4a^8T^2}}\right)^{1/3}\right).\end{aligned} \tag{1.19}$$

As T grows, we have

$$\Delta^* = \frac{2^{2/3}a^{4/3}}{\sigma^{4/3}}T^{1/3} + O\left(\frac{1}{T^{1/3}}\right). \tag{1.20}$$

The trade-off between bias and variance made explicit in (1.17)-(1.19) is not unlike the situation in nonparametric estimation with Δ^{-1} playing the role of the bandwidth h. A lower h reduces the bias but increases the variance, and the optimal choice of h balances the two effects.

Note that these are exact small sample expressions, valid for all T. Asymptotically in T, $\mathrm{Var}\left[\hat{\sigma}^2\right] \to 0$, and hence the RMSE of the estimator is dominated by the bias term which is independent of T. And given the form of the bias (1.17), 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 1.2). The rate at which Δ^* should increase with T is given by (1.20). Also, in the limit where the noise disappears ($a \to 0$ and $\mathrm{Cum}_4\left[U\right] \to 0$), the optimal sampling interval Δ^* tends to 0.

How does a small departure from a normal distribution of the microstructure noise affect the optimal sampling frequency? The answer is that a small positive (resp. negative) departure of $\mathrm{Cum}\,4\left[U\right]$ starting from the normal value of 0 leads to an increase (resp. decrease) in Δ^*, since

$$\begin{aligned}\Delta^* = & \Delta^*_{\text{normal}} + \\ & + \frac{\left(\left(1 + \sqrt{1 - \frac{2a^4}{T^2\sigma^4}}\right)^{2/3} - \left(1 - \sqrt{1 - \frac{2a^4}{T^2\sigma^4}}\right)^{2/3}\right)}{3\;2^{1/3}a^{4/3}T^{1/3}\sqrt{1 - \frac{2a^4}{T^2\sigma^4}}\sigma^{8/3}}\mathrm{Cum}\,4\left[U\right] + \\ & + O\left(\mathrm{Cum}\,4\left[U\right]^2\right),\end{aligned} \tag{1.21}$$

where Δ^*_{normal} is the value of Δ^* corresponding to $\text{Cum}_4[U] = 0$. And of course the full formula (1.20) can be used to get the exact answer for any departure from normality instead of the comparative static one.

Another interesting asymptotic situation occurs if one attempts to use higher and higher frequency data ($\Delta \to 0$, say sampled every minute) over a fixed time period (T fixed, say a day). Since the expressions in Theorem 1 are exact small sample ones, they can in particular be specialized to analyze this situation. With $n = T/\Delta$, it follows from (1.17)-(1.19) that

$$E\left[\hat{\sigma}^2\right] = \frac{2na^2}{T} + o(n) = \frac{2nE\left[U^2\right]}{T} + o(n), \tag{1.22}$$

$$\text{Var}\left[\hat{\sigma}^2\right] = \frac{2n\left(6a^4 + 2\,\text{Cum}_4[U]\right)}{T^2} + o(n) = \frac{4nE\left[U^4\right]}{T^2} + o(n), \tag{1.23}$$

so $(T/2n)\hat{\sigma}^2$ becomes an estimator of $E\left[U^2\right] = a^2$ whose asymptotic variance is $E\left[U^4\right]$. Note in particular that $\hat{\sigma}^2$ estimates the variance of the noise, which is essentially unrelated to the object of interest σ^2. This type of asymptotics is relevant in the stochastic volatility case we analyze in our companion paper [43].

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., [2], [3] and [26]). 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., [6], [7] and [8]).

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 [2]), 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. Complementary to this are the results of [22] which show that the presence of even increasingly negligible noise is sufficient to adversely affect the identification of σ^2.

1.4 Concrete Implications for Empirical Work with High Frequency Data

The clear message of Theorem 1 for empirical researchers working with high frequency financial data is that it may be optimal to sample less frequently. As discussed in the Introduction, authors have reduced their sampling frequency below that of the actual record of observations in a somewhat ad hoc fashion, with typical choices 5 minutes and up. Our analysis provides not only a theoretical rationale for sampling less frequently, but also delivers a precise answer to the question of "how often one should sample?" For that purpose, we need to calibrate the parameters appearing in Theorem 1, namely σ, α, $\mathrm{Cum}_4[U]$, Δ and T. We assume in this calibration exercise that the noise is Gaussian, in which case $\mathrm{Cum}_4[U] = 0$.

1.4.1 Stocks

We use existing studies in empirical market microstructure to calibrate the parameters. One such study is [35], who estimated on the basis of a sample of 274 NYSE stocks that approximately 60% of the total variance of price changes is attributable to market microstructure effects (they report a range of values for π from 54% in the first half hour of trading to 65% in the last half hour, see their Table 4; they also decompose this total variance into components due to discreteness, asymmetric information, transaction costs and the interaction between these effects). Given that their sample contains an average of 15 transactions per hour (their Table 1), we have in our framework

$$\pi = 60\%, \ \Delta = 1/(15 \times 7 \times 252). \tag{1.24}$$

These values imply from (1.13) that $a = 0.16\%$ if we assume a realistic value of $\sigma = 30\%$ per year. (We do not use their reported volatility number since they apparently averaged the variance of price changes over the 274 stocks instead of the variance of the returns. Since different stocks have different price levels, the price variances across stocks are not directly comparable. This does not affect the estimated fraction π however, since the price level scaling factor cancels out between the numerator and the denominator).

The magnitude of the effect is bound to vary by type of security, market and time period. [29] estimates the value of a to be 0.33%. Some authors have reported even larger effects. Using a sample of NASDAQ stocks, [32] estimate that about 50% of the daily variance of returns in due to the bid-ask effect. With $\sigma = 40\%$ (NASDAQ stocks have higher volatility), the values

$$\pi = 50\%, \ \Delta = 1/252$$

yield the value $a = 1.8\%$. Also on NASDAQ, [12] estimate that 11% of the variance of weekly returns (see their Table 4, middle portfolio) is due to bid-ask effects. The values

$$\pi = 11\%, \ \Delta = 1/52$$

imply that $a = 1.4\%$.

In Table 1.1, we compute the value of the optimal sampling interval Δ^* implied by different combinations of sample length (T) and noise magnitude (a). The volatility of the efficient price process is held fixed at $\sigma = 30\%$ in Panel A, which is a realistic value for stocks. The numbers in the table show that the optimal sampling frequency can be substantially affected by even relatively small quantities of microstructure noise. For instance, using the value $a = 0.15\%$ calibrated from [35], we find an optimal sampling interval of 22 minutes if the sampling length is 1 day; longer sample lengths lead to higher optimal sampling intervals. With the higher value of $a = 0.3\%$, approximating the estimate from [29], the optimal sampling interval is 57 minutes. A lower value of the magnitude of the noise translates into a higher frequency: for instance, $\Delta^* = 5$ minutes if $a = 0.05\%$ and $T = 1$ day. Figure 1.2 displays the RMSE of the estimator as a function of Δ and T, using parameter values $\sigma = 30\%$ and $a = 0.15\%$. The figure illustrates the fact that deviations from the optimal choice of Δ lead to a substantial increase in the RMSE: for example, with $T = 1$ month, the RMSE more than doubles if, instead of the optimal $\Delta^* = 1$ hour, one uses $\Delta = 15$ minutes.

1.4.2 Currencies

Looking now at foreign exchange markets, empirical market microstructure studies have quantified the magnitude of the bid-ask spread. For example, [9] computes the average bid/ask spread s in the wholesale market for different currencies and reports values of $s = 0.05\%$ for the German mark, and 0.06% for the Japanese yen (see Panel B of his Table 2). We calculated the corresponding numbers for the 1996-2002 period to be 0.04% for the mark (followed by the euro) and 0.06% for the yen. Emerging market currencies have higher spreads: for instance, $s = 0.12\%$ for Korea and 0.10% for Brazil. During the same period, the volatility of the exchange rate was $\sigma = 10\%$ for the German mark, 12% for the Japanese yen, 17% for Brazil and 18% for Korea. In Panel B of Table 1.1, we compute Δ^* with $\sigma = 10\%$, a realistic value for the euro and yen. As we noted above, if the sole source of the noise were a bid/ask spread of size s, then a should be set to $s/2$. Therefore Panel B reports the values of Δ^* for values of a ranging from 0.02% to 0.1%. For example, the dollar/euro or dollar/yen exchange rates (calibrated to $\sigma = 10\%$, $a = 0.02\%$) should be sampled every $\Delta^* = 23$ minutes if the overall sample length is $T = 1$ day, and every 1.1 hours if $T = 1$ year.

Furthermore, using the bid/ask spread alone as a proxy for all microstructure frictions will lead, except in unusual circumstances, to an understatement of the parameter a, since variances are additive. Thus, since Δ^* is increasing in a, one should interpret the value of Δ^* read off 1.1 on the row corresponding to $a = s/2$ as a lower bound for the optimal sampling interval.

1.4.3 Monte Carlo Evidence

To validate empirically these results, we perform Monte Carlo simulations. We simulate $M = 10,000$ samples of length $T = 1$ year of the process X, add microstructure noise U to generate the observations $\tilde{X}$ and then the log returns Y. We sample the log-returns at various intervals Δ ranging from 5 minutes to 1 week and calculate the bias and variance of the estimator $\hat{\sigma}^2$ over the M simulated paths. We then compare the results to the theoretical values given in (1.17)-(1.19) of Theorem 1. The noise distribution is Gaussian, $\sigma = 30\%$ and $a = 0.15\%$ – the values we calibrated to stock returns data above. Table 1.2 shows that the theoretical values are in close agreement with the results of the Monte Carlo simulations.

The table also illustrates the magnitude of the bias inherent in sampling at too high a frequency. While the value of σ^2 used to generate the data is 0.09, the expected value of the estimator when sampling every 5 minutes is 0.18, so on average the estimated quadratic variation is twice as big as it should be in this case.

1.5 Incorporating Market Microstructure Noise Explicitly

So far we have stuck to the sum of squares of log-returns as our estimator of volatility. We then showed that, for this estimator, the optimal sampling frequency is finite. But this implies that one is discarding a large proportion of the high frequency sample (299 out of every 300 observations in the example described in the Introduction), in order to mitigate the bias induced by market microstructure noise. Next, we show that if we explicitly incorporate the $U's$ into the likelihood function, then we are back in the situation where the optimal sampling scheme consists in sampling as often as possible – i.e., using all the data available.

Specifying the likelihood function of the log-returns, while recognizing that they incorporate noise, requires that we take a stand on the distribution of the noise term. Suppose for now that the microstructure noise is normally distributed, an assumption whose effect we will investigate below in Section 1.6. Under this assumption, the likelihood function for the $Y's$ is given by

$$l(\eta, \gamma^2) = -\ln\det(V)/2 - N\ln(2\pi\gamma^2)/2 - (2\gamma^2)^{-1}Y'V^{-1}Y\,, \tag{1.25}$$

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

$$V = [v_{ij}]_{i,j=1,\dots,N} = \begin{pmatrix} 1+\eta^2 & \eta & 0 & \cdots & 0 \\ \eta & 1+\eta^2 & \eta & \ddots & \vdots \\ 0 & \eta & 1+\eta^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \eta \\ 0 & \cdots & 0 & \eta & 1+\eta^2 \end{pmatrix}. \tag{1.26}$$

Further,

$$\det(V) = \frac{1-\eta^{2N+2}}{1-\eta^2}, \tag{1.27}$$

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

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

(see [15]). The product $V\Omega$ differs from the identity matrix only on the first and last rows. The exact inverse is $V^{-1} = \left[v^{ij}\right]_{i,j=1,\dots,N}$ where

$$\begin{aligned} v^{ij} = {} & \left(1-\eta^2\right)^{-1}\left(1-\eta^{2N+2}\right)^{-1}\Big\{(-\eta)^{|i-j|} - (-\eta)^{i+j} - (-\eta)^{2N-i-j+2} - \\ & - (-\eta)^{2N+|i-j|+2} + (-\eta)^{2N+i-j+2} + (-\eta)^{2N-i+j+2}\Big\} \end{aligned} \tag{1.28}$$

(see [24] and [40]).

From the perspective of practical implementation, this estimator is nothing else than the MLE estimator of an MA(1) process with Gaussian errors: any existing computer routines for the MA(1) situation can therefore be applied (see e.g., Section 5.4 in [25]). In particular, the likelihood function can be expressed in a computationally efficient form by triangularizing the matrix V, yielding the equivalent expression:

$$l(\eta,\gamma^2) = -\frac{1}{2}\sum_{i=1}^{N}\ln(2\pi d_i) - \frac{1}{2}\sum_{i=1}^{N}\frac{\tilde{Y}_i^2}{d_i}, \tag{1.29}$$

where

$$d_i = \gamma^2\frac{1+\eta^2+\dots+\eta^{2i}}{1+\eta^2+\dots+\eta^{2(i-1)}},$$

and the $\tilde{Y}_i's$ are obtained recursively as $\tilde{Y}_1 = Y_1$ and for $i = 2,\dots,N$:

$$\tilde{Y}_i = Y_i - \frac{\eta\left(1+\eta^2+\dots+\eta^{2(i-2)}\right)}{1+\eta^2+\dots+\eta^{2(i-1)}}\tilde{Y}_{i-1}.$$

This latter form of the log-likelihood function involves only single sums as opposed to double sums if one were to compute $Y'V^{-1}Y$ by brute force using the expression of V^{-1} given above.

We now compute the distribution of the MLE estimators of σ^2 and a^2, which follows by the delta method from the classical result for the MA(1) estimators of γ and η :

Proposition 1. *When U is normally distributed, the MLE $(\hat{\sigma}^2, \hat{a}^2)$ is consistent and its asymptotic variance is given by*

$$\mathrm{AVAR}_{normal}(\hat{\sigma}^2, \hat{a}^2) = \begin{pmatrix} 4\sqrt{\sigma^6 \Delta \left(4a^2 + \sigma^2 \Delta\right)} + 2\sigma^4 \Delta & -\sigma^2 \Delta h(\Delta, \sigma^2, a^2) \\ \bullet & \frac{\Delta}{2}\left(2a^2 + \sigma^2 \Delta\right) h(\Delta, \sigma^2, a^2) \end{pmatrix}$$

with

$$h(\Delta, \sigma^2, a^2) \equiv 2a^2 + \sqrt{\sigma^2 \Delta \left(4a^2 + \sigma^2 \Delta\right)} + \sigma^2 \Delta \,. \tag{1.30}$$

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

$$\mathrm{AVAR}_{\mathrm{normal}}(\hat{\sigma}^2) = 8\sigma^3 a \Delta^{1/2} + 2\sigma^4 \Delta + o(\Delta)\,, \tag{1.31}$$

the loss of efficiency relative to the case where no market microstructure noise is present (and $\mathrm{AVAR}(\hat{\sigma}^2) = 2\sigma^4 \Delta$ as given in (1.7) if $a^2 = 0$ is not estimated, or $\mathrm{AVAR}(\hat{\sigma}^2) = 6\sigma^4 \Delta$ if $a^2 = 0$ is estimated) is at order $\Delta^{1/2}$. Figure 1.3 plots the asymptotic variances of $\hat{\sigma}^2$ as functions of Δ with and without noise (the parameter values are again $\sigma = 30\%$ and $a = 0.15\%$). Figure 1.4 reports histograms of the distributions of $\hat{\sigma}^2$ and $\hat{a}^2$ from 10,000 Monte Carlo simulations with the solid curve plotting the asymptotic distribution of the estimator from Proposition 1. The sample path is of length $T = 1$ year, the parameter values the same as above, and the process is sampled every 5 minutes – since we are now accounting explicitly for the presence of noise, there is no longer a reason to sample at lower frequencies. Indeed, the figure documents the absence of bias and the good agreement of the asymptotic distribution with the small sample one.

1.6 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 1.5, but mistakenly 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 i.i.d. 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 (1.25), using (1.9)-(1.10). This means that the scores $\dot{l}_{\sigma^2}$ and $\dot{l}_{a^2}$, or equivalently (C.1) and (C.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 $(Y_1, ..., Y_N)$, which is unchanged by the absence of normality, the moment functions are unbiased under the true distribution of the $U's$:

$$E_{\text{true}}[\dot{l}_{\eta}] = E_{\text{true}}[\dot{l}_{\gamma^2}] = 0\,, \tag{1.32}$$

and similarly for $\dot{l}_{\sigma^2}$ 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 therefore lies in the asymptotic variance matrix. By using the cumulants of the distribution of U, we express the asymptotic variance of these estimators in terms of deviations from normality. But as far as computing the actual estimator, nothing has changed relative to Section 1.5: we are still calculating the MLE for an MA(1) process with Gaussian errors and can apply exactly the same computational routine.

However, since the error distribution is potentially misspecified, one could expect the asymptotic distribution of the estimator to be altered. This turns out not be the case, as far as $\hat{\sigma}^2$ is concerned:

Theorem 2. *The estimators $(\hat{\sigma}^2, \hat{a}^2)$ obtained by maximizing the possibly misspecified log-likelihood (1.25) are consistent and their asymptotic variance is given by*

$$\mathrm{AVAR}_{true}(\hat{\sigma}^2, \hat{a}^2) = \mathrm{AVAR}_{normal}(\hat{\sigma}^2, \hat{a}^2) + \mathrm{Cum}_4\,[U] \begin{pmatrix} 0 & 0 \\ 0 & \Delta \end{pmatrix}, \tag{1.33}$$

where $\mathrm{AVAR}_{normal}(\hat{\sigma}^2, \hat{a}^2)$ is the asymptotic variance in the case where the distribution of U is normal, that is, the expression given in Proposition 1.

In other words, the asymptotic variance of $\hat{\sigma}^2$ is identical to its expression if the $U's$ had been normal. Therefore the correction we proposed for the presence of market microstructure noise relying on the assumption that the noise is Gaussian is robust to misspecification of the error distribution.

Documenting the presence of the correction term through simulations presents a challenge. At the parameter values calibrated to be realistic, the order of magnitude of a is a few basis points, say $a = 0.10\% = 10^{-3}$. But if U is of order 10^{-3}, $\mathrm{Cum}_4[U]$ which is of the same order as U^4, is of order 10^{-12}. In other words, with a typical noise distribution, the correction term in (1.33) will not be visible.

To nevertheless make it discernible, we use a distribution for U with the same calibrated standard deviation a as before, but a disproportionately large fourth cumulant. Such a distribution can be constructed by letting $U = \omega T_{\nu}$

where $\omega > 0$ is constant and T_ν is a Student t distribution with v degrees of freedom. T_ν has mean zero, finite variance as long as $v > 2$ and finite fourth moment (hence finite fourth cumulant) as long as $v > 4$. But as v approaches 4 from above, $E[T_\nu^4]$ tends to infinity. This allows us to produce an arbitrarily high value of $\mathrm{Cum}_4[U]$ while controlling for the magnitude of the variance. The specific expressions of a^2 and $\mathrm{Cum}_4[U]$ for this choice of U are given by

$$a^2 = \mathrm{Var}[U] = \frac{\omega^2 \nu}{\nu - 2}, \tag{1.34}$$

$$\mathrm{Cum}_4[U] = \frac{6\omega^4\nu^2}{(\nu-4)(\nu-2)^2}. \tag{1.35}$$

Thus we can select the two parameters (ω, ν) to produce desired values of $(a^2, \mathrm{Cum}_4[U])$. As before, we set $a = 0.15\%$. Then, given the form of the asymptotic variance matrix (1.33), we set $\mathrm{Cum}_4[U]$ so that $\mathrm{Cum}_4[U]\,\Delta = \mathrm{AVAR}_{\mathrm{normal}}(\hat{a}^2)/2$. This makes $\mathrm{AVAR}_{\mathrm{true}}(\hat{a}^2)$ by construction 50% larger than $\mathrm{AVAR}_{\mathrm{normal}}(\hat{a}^2)$. The resulting values of (ω, ν) from solving (1.34)-(1.35) are $\omega = 0.00115$ and $v = 4.854$. As above, we set the other parameters to $\sigma = 30\%$, $T = 1$ year, and $\Delta = 5$ minutes. Figure 1.5 reports histograms of the distributions of $\hat{\sigma}^2$ and $\hat{a}^2$ from 10,000 Monte Carlo simulations. The solid curve plots the asymptotic distribution of the estimator, given now by (1.33). There is again good adequacy between the asymptotic and small sample distributions. In particular, we note that as predicted by Theorem 2, the asymptotic variance of $\hat{\sigma}^2$ is unchanged relative to Figure 1.4 while that of $\hat{a}^2$ is 50% larger. The small sample distribution of $\hat{\sigma}^2$ appears unaffected by the non-Gaussianity of the noise; with a skewness of 0.07 and a kurtosis of 2.95, it is closely approximated by its asymptotic Gaussian limit. The small sample distribution of $\hat{a}^2$ does exhibit some kurtosis (4.83), although not large relative to that of the underlying noise distribution (the values of ω and ν imply a kurtosis for U of $3 + 6/(\nu - 4) = 10$). Similar simulations but with a longer time span of $T = 5$ years are even closer to the Gaussian asymptotic limit: the kurtosis of the small sample distribution of $\hat{a}^2$ goes down to 2.99.

1.7 Robustness to Misspecification of the Noise Distribution

Going back to the theoretical aspects, the above Theorem 2 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 (1.25), $-\ddot{l}(\hat{\sigma}^2, \hat{a}^2)$ denote the observed information matrix in the original parameters σ^2 and a^2. Then

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

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 $\mathrm{AVAR}_{\mathrm{normal}}(\hat{\sigma}^2, \hat{a}^2)$. Since this matrix coincides with $\mathrm{AVAR}_{\mathrm{true}}(\hat{\sigma}^2, \hat{a}^2)$ for all but the (a^2, a^2) term (see (1.33)), 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 $S \equiv E_{\mathrm{true}}[\dot{l}\dot{l}']/N$ and $D \equiv -E_{\mathrm{true}}[\ddot{l}]/N$ (differentiation refers to the original parameters (σ^2, a^2), not the transformed parameters (γ^2, η)) this identity says that

$$S - D = 0. \tag{1.36}$$

It implies that the asymptotic variance takes the form

$$\mathrm{AVAR} = \Delta(DS^{-1}D)^{-1} = \Delta D^{-1}. \tag{1.37}$$

It is clear that (1.37) remains valid if the second Bartlett identity holds only to first order, i.e.,

$$S - D = o(1) \tag{1.38}$$

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

However, in view of Theorem 2, equation (1.38) cannot be satisfied. In fact, we show in Appendix E that

$$S - D = \mathrm{Cum}_4[U]\, gg' + o(1), \tag{1.39}$$

where

$$g = \begin{pmatrix} g_{\sigma^2} \\ g_{a^2} \end{pmatrix} = \begin{pmatrix} \frac{\Delta^{1/2}}{\sigma(4a^2+\sigma^2\Delta)^{3/2}} \\ \frac{1}{2a^4}\left(1 - \frac{\Delta^{1/2}\sigma\left(6a^2+\sigma^2\Delta\right)}{(4a^2+\sigma^2\Delta)^{3/2}}\right) \end{pmatrix}. \tag{1.40}$$

From (1.40), we see that $g \neq 0$ whenever $\sigma^2 > 0$. This is consistent with the result in Theorem 2 that the true asymptotic variance matrix, $\mathrm{AVAR}_{\mathrm{true}}(\hat{\sigma}^2, \hat{a}^2)$, does not coincide with the one for Gaussian noise, $\mathrm{AVAR}_{\mathrm{normal}}(\hat{\sigma}^2, \hat{a}^2)$. 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 (1.39) 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 (1.39)

$$\begin{aligned} S_{\sigma^2\sigma^2} - D_{\sigma^2\sigma^2} &= N^{-1}E_{\text{true}}\left[\dot{l}_{\sigma^2\sigma^2}(\sigma^2,a^2)^2\right] + N^{-1}E_{\text{true}}\left[\ddot{l}_{\sigma^2\sigma^2}(\sigma^2,a^2)\right] \\ &= \text{Cum}_4\left[U\right]\left(g_{\sigma^2}\right)^2 + o(1) \end{aligned}$$

which is not $o(1)$ unless $\text{Cum}_4\left[U\right] = 0$.

To make the connection between Theorem 2 and the second Bartlett identity, one needs to go to the log profile likelihood

$$\lambda(\sigma^2) \quad \equiv \quad \sup_{a^2} \; l(\sigma^2, a^2)\,. \tag{1.41}$$

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

$$\ddot{\lambda}(\hat{\sigma}^2)^{-1} = \left[\ddot{l}(\hat{\sigma}^2, \hat{a}^2)^{-1}\right]_{\sigma^2\sigma^2}, \tag{1.42}$$

i.e., the first (upper left hand corner) component of the inverse observed information in the original problem. We recall the rationale for equation (1.42) in Appendix E, where we also show that $E_{\text{true}}[\dot{\lambda}] = o(N)$. In view of Theorem 2, $\ddot{\lambda}(\hat{\sigma}^2)$ 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 (1.38), that is

$$N^{-1}E_{\text{true}}[\dot{\lambda}(\sigma^2)^2] + N^{-1}E_{\text{true}}[\ddot{\lambda}(\sigma^2)] = o(1)\,. \tag{1.43}$$

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

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 [30] and [37], and the references therein, and for early econometrics examples see [34] and [42]). 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 ($\dot{l}_{\sigma^2}$ 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, page 326 of [37]). Thus only by going to the profile likelihood λ can one make the usual comparison to quasi-likelihood.

1.8 Randomly Spaced Sampling Intervals

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

[4], 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 i.i.d., independent of the W process. This assumption, while not completely realistic (see [17] for a discrete time analysis of the autoregressive dependence of the times between trades) allows us to make explicit calculations at the interface between the continuous and discrete time scales. We denote by N_T the number of observations recorded by time T. N_T is random if the $\Delta' s$ are. We also suppose that U_{τ_i} can be written U_i, where the U_i are i.i.d. and independent of the W process and the $\Delta_i' s$. Thus, the observation noise is the same at all observation times, whether random or nonrandom. If we define the Y_is as before, in the first two lines of (1.8), 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

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

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 $L\left(\Delta_{N_T}, ..., \Delta_1; \psi\right)$ 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

$$\sum_{n=1}^{N} \ln L\left(Y_N, ..., Y_1 | \Delta_N, ..., \Delta_1; \beta\right) + \sum_{n=1}^{N-1} \ln L\left(\Delta_N, ..., \Delta_1; \psi\right) , \tag{1.44}$$

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 $Y' s$, now given by

$$\Sigma = \begin{pmatrix} \sigma^2\Delta_1 + 2a^2 & -a^2 & 0 & \cdots & 0 \\ -a^2 & \sigma^2\Delta_2 + 2a^2 & -a^2 & \ddots & \vdots \\ 0 & -a^2 & \sigma^2\Delta_3 + 2a^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a^2 \\ 0 & \cdots & 0 & -a^2 & \sigma^2\Delta_n + 2a^2 \end{pmatrix} . \tag{1.45}$$

Note that in the equally spaced case, $\Sigma = \gamma^2 V$. But now Y no longer follows an MA(1) process in general. Furthermore, the time variation in $\Delta_i' s$ gives rise to heteroskedasticity as is clear from the diagonal elements of Σ. This is consistent with the predictions of the model of [16] where the variance of the transaction price process $\tilde{X}$ is heteroskedastic as a result of the influence of the sampling times. In their model, the sampling times are autocorrelated and correlated with the evolution of the price process, factors we have assumed away here. However, [4] show how to conduct likelihood inference in such a situation.

The log-likelihood function is given by

$$\begin{aligned}\ln L\left(Y_N, ..., Y_1 | \Delta_N, ..., \Delta_1; \beta\right) &\equiv l(\sigma^2, a^2) \\ &= -\ln\det(\Sigma)/2 - N\ln(2\pi)/2 - Y'\Sigma^{-1}Y/2\,.\end{aligned} \tag{1.46}$$

In order to calculate this log-likelihood function in a computationally efficient manner, it is desirable to avoid the "brute force" inversion of the $N \times N$ matrix Σ. We extend the method used in the MA(1) case (see (1.29)) as follows. By Theorem 5.3.1 in [13], and the development in the proof of their Theorem 5.4.3, we can decompose Σ in the form $\Sigma = LDL^T$, where L is a lower triangular matrix whose diagonals are all 1 and D is diagonal. To compute the relevant quantities, their Example 5.4.3 shows that if one writes $D = \text{diag}(g_1, ..., g_n)$ and

$$L = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \kappa_2 & 1 & 0 & \ddots & \vdots \\ 0 & \kappa_3 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \kappa_n & 1 \end{pmatrix}, \tag{1.47}$$

then the $g_k's$ and $\kappa_k's$ follow the recursion equation $g_1 = \sigma^2\Delta_1 + 2a^2$ and for $i = 2, ..., N$:

$$\kappa_i = -a^2/g_{i-1} \qquad \text{and} \qquad g_i = \sigma^2\Delta_i + 2a^2 + \kappa_i a^2. \tag{1.48}$$

Then, define $\tilde{Y} = L^{-1}Y$ so that $Y'\Sigma^{-1}Y = \tilde{Y}'D^{-1}\tilde{Y}$. From $Y = L\tilde{Y}$, it follows that $\tilde{Y}_1 = Y_1$ and, for $i = 2, ..., N$:

$$\tilde{Y}_i = Y_i - \kappa_i \tilde{Y}_{i-1}\,.$$

And $\det(\Sigma) = \det(D)$ since $\det(L) = 1$. Thus we have obtained a computationally simple form for (1.46) that generalizes the MA(1) form (1.29) to the case of non-identical sampling intervals:

$$l(\sigma^2, a^2) = -\frac{1}{2}\sum_{i=1}^{N}\ln\left(2\pi g_i\right) - \frac{1}{2}\sum_{i=1}^{N}\frac{\tilde{Y}_i^2}{g_i}\,. \tag{1.49}$$

We can now turn to statistical inference using this likelihood function. As usual, the asymptotic variance of $T^{1/2}(\hat{\sigma}^2 - \sigma^2, \hat{a}^2 - a^2)$ is of the form

$$\text{AVAR}(\hat{\sigma}^2, \hat{a}^2) = \lim_{T\ ->\ \infty} \begin{pmatrix} \frac{1}{T}E\left[-\ddot{l}_{\sigma^2\sigma^2}\right] & \frac{1}{T}E\left[-\ddot{l}_{\sigma^2a^2}\right] \\ \bullet & \frac{1}{T}E\left[-\ddot{l}_{a^2a^2}\right] \end{pmatrix}^{-1}. \tag{1.50}$$

To compute this quantity, suppose in the following that β_1 and β_2 can represent either σ^2 or a^2. We start with:

Lemma 2. *Fisher's Conditional Information is given by*

$$E\left[-\ddot{l}_{\beta_2\beta_1}\middle|\Delta\right] = -\frac{1}{2}\frac{\partial^2 \ln\det\Sigma}{\partial\beta_2\beta_1}. \tag{1.51}$$

To compute the asymptotic distribution of the MLE of (β_1, β_2), one would then need to compute the inverse of $E\left[-\ddot{l}_{\beta_2\beta_1}\right] = E_\Delta\left[E\left[-\ddot{l}_{\beta_2\beta_1}\middle|\Delta\right]\right]$ where E_Δ denotes expectation taken over the law of the sampling intervals. From (1.51), and since the order of E_Δ and $\partial^2/\partial\beta_2\beta_1$ can be interchanged, this requires the computation of

$$E_\Delta\left[\ln\det\Sigma\right] = E_\Delta\left[\ln\det D\right] = \sum_{i=1}^{N} E_\Delta\left[\ln\left(g_i\right)\right]$$

where from (1.48) the $g_i's$ are given by the continuous fraction

$$\begin{aligned}
g_1 &= \sigma^2\Delta_1 + 2a^2, \\
g_2 &= \sigma^2\Delta_2 + 2a^2 - \frac{a^4}{\sigma^2\Delta_1 + 2a^2}, \\
g_3 &= \sigma^2\Delta_3 + 2a^2 - \frac{a^4}{\sigma^2\Delta_2 + 2a^2 - \frac{a^4}{\sigma^2\Delta_1 + 2a^2}},
\end{aligned}$$

and in general

$$g_i = \sigma^2\Delta_i + 2a^2 - \cfrac{a^4}{\sigma^2\Delta_{i-1} + 2a^2 - \cfrac{a^4}{\ddots}}.$$

It therefore appears that computing the expected value of $\ln\left(g_i\right)$ over the law of $(\Delta_1, \Delta_2, ..., \Delta_i)$ will be impractical.

1.8.1 Expansion Around a Fixed Value of Δ

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

$$\Delta_i = \Delta_0\left(1 + \varepsilon\xi_i\right), \tag{1.52}$$

where ε and Δ_0 are nonrandom, the $\xi_i's$ are i.i.d. random variables with mean zero and finite distribution. We will Taylor-expand the expressions above around $\varepsilon = 0$, i.e., around the non-random sampling case we have just finished dealing with. Our expansion is one that is valid when the randomness of the sampling intervals remains small, i.e., when $\mathrm{Var}\left[\Delta_i\right]$ is small, or $o(1)$.

Then we have $\Delta_0 = E[\Delta] = O(1)$ and $\mathrm{Var}[\Delta_i] = \Delta_0^2 \varepsilon^2 \mathrm{Var}[\xi_i]$. The natural scaling is to make the distribution of ξ_i finite, i.e., $\mathrm{Var}[\xi_i] = O(1)$, so that $\varepsilon^2 = O(\mathrm{Var}[\Delta_i]) = o(1)$. But any other choice would have no impact on the result since $\mathrm{Var}[\Delta_i] = o(1)$ implies that the product $\varepsilon^2 \mathrm{Var}[\xi_i]$ is $o(1)$ and whenever we write reminder terms below they can be expressed as $O_p\left(\varepsilon^3 \xi^3\right)$ instead of just $O\left(\varepsilon^3\right)$. We keep the latter notation for clarity given that we set $\xi_i = O_p(1)$. Furthermore, for simplicity, we take the $\xi_i' s$ to be bounded.

We emphasize that the time increments or durations Δ_i do not tend to zero length as $\varepsilon \to 0$. It is only the variability of the Δ_i's that goes to zero.

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

$$\mathrm{AVAR}(\hat{\sigma}^2, \hat{a}^2) = A^{(0)} + \varepsilon^2 A^{(2)} + O(\varepsilon^3), \tag{1.53}$$

where

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

and

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

with

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

In connection with the preceding result, we underline that the quantity $\mathrm{AVAR}(\hat{\sigma}^2, \hat{a}^2)$ is a limit as $T \to \infty$, as in (1.50). The equation (1.53), therefore, is an expansion in ε *after* $T \to \infty$.

Note that $A^{(0)}$ is the asymptotic variance matrix already present in Proposition 1, except that it is evaluated at $\Delta_0 = E[\Delta]$. Note also that the second order correction term is proportional to $\mathrm{Var}[\xi]$, 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 1.

1.8.2 Randomly Spaced Sampling Intervals and Misspecified Microstructure Noise

Suppose now, as in Section 1.6, that the $U's$ are i.i.d., have mean zero and variance a^2, but are otherwise not necessarily Gaussian. We adopt the same approach as in Section 1.6, 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 4. *The asymptotic variance is given by*

$$\begin{aligned}\mathrm{AVAR}_{true}(\hat{\sigma}^2, \hat{a}^2) = {} & \left(A^{(0)} + \mathrm{Cum}_4\left[U\right] B^{(0)}\right) + \\ & + \varepsilon^2 \left(A^{(2)} + \mathrm{Cum}_4\left[U\right] B^{(2)}\right) + O(\varepsilon^3)\,,\end{aligned} \tag{1.54}$$

where $A^{(0)}$ and $A^{(2)}$ are given in the statement of Theorem 3 and

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

while

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

with

$$\begin{aligned} B^{(2)}_{\sigma^2\sigma^2} &= \frac{10\Delta_0^{3/2}\sigma^5}{(4a^2+\Delta_0\sigma^2)^{5/2}} + \frac{4\Delta_0^2\sigma^6\left(16a^4 + 11a^2\Delta_0\sigma^2 + 2\Delta_0^2\sigma^4\right)}{(2a^2+\Delta_0\sigma^2)^3(4a^2+\Delta_0\sigma^2)^2}\,, \\ B^{(2)}_{\sigma^2 a^2} &= \frac{-\Delta_0^2\sigma^4}{(2a^2+\Delta_0\sigma^2)^3(4a^2+\Delta_0\sigma^2)^{5/2}} \times \\ & \times \left(\sqrt{4a^2+\Delta_0\sigma^2}\left(32a^6 + 64a^4\Delta_0\sigma^2 + 35a^2\Delta_0^2\sigma^4 + 6\Delta_0^3\sigma^6\right) + \right. \\ & \left. + \Delta_0^{1/2}\sigma\left(116a^6 + 126a^4\Delta_0\sigma^2 + 47a^2\Delta_0^2\sigma^4 + 6\Delta_0^3\sigma^6\right)\right), \\ B^{(2)}_{a^2a^2} &= \frac{16a^8\Delta_0^{5/2}\sigma^3\left(13a^4 + 10a^2\Delta_0\sigma^2 + 2\Delta_0^2\sigma^4\right)}{(2a^2+\Delta_0\sigma^2)^3(4a^2+\Delta_0\sigma^2)^{5/2}\left(2a^2+\sigma^2\Delta - \sqrt{\sigma^2\Delta\left(4a^2+\sigma^2\Delta\right)}\right)^2}\,. \end{aligned}$$

The term $A^{(0)}$ is the base asymptotic variance of the estimator, already present with fixed sampling and Gaussian noise. The term $\mathrm{Cum}_4\left[U\right] B^{(0)}$ is the correction due to the misspecification of the error distribution. These two terms are identical to those present in Theorem 2. 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 3, and $\mathrm{Cum}_4\left[U\right] B^{(2)}$ is the further correction due to the sampling randomness. Both $A^{(2)}$ and $B^{(2)}$ are proportional to $\mathrm{Var}[\xi]$ and hence vanish in the absence of sampling randomness.

1.9 Extensions

In this section, we briefly sketch four extensions of our basic model. First, we show that the introduction of a drift term does not alter our conclusions. Then we examine the situation where market microstructure noise is serially correlated; there, we show that the insight of Theorem 1 remains valid, namely that the optimal sampling frequency is finite. Third, we then turn to the case where the noise is correlated with the efficient price signal. Fourth, we discuss what happens if volatility is stochastic.

In a nutshell, each one of these assumptions can be relaxed without affecting our main conclusion, namely that the presence of the noise gives rise to a finite optimal sampling frequency. The second part of our analysis, dealing with likelihood corrections for microstructure noise, will not necessarily carry through unchanged if the assumptions are relaxed (for instance, there is not even a known likelihood function if volatility is stochastic, and the likelihood must be modified if the assumed variance-covariance structure of the noise is modified).

1.9.1 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

$$X_t = \mu t + \sigma W_t \,. \tag{1.55}$$

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

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

where the covariance matrix is given in (1.45), and where $\Delta = (\Delta_1, ..., \Delta_N)'$. 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,

$$E[\ddot{l}_{\mu\beta}] = 0 \,. \tag{1.56}$$

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 $\mathrm{Cov}(\dot{l}, \dot{l}) = TE[\Delta]S + o(T)$. 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 $\text{AVAR}_{\sigma^2,a^2}$. Since D is block diagonal by (1.56),

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

it follows that

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

Hence

$$\text{AVAR}(\hat{\sigma}^2, \hat{a}^2) = E[\Delta]D_{\sigma^2,a^2}^{-1}S_{\sigma^2,a^2}D_{\sigma^2,a^2}^{-1}. \tag{1.57}$$

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.

1.9.2 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 i.i.d. with mean 0 and variance a^2, the market microstructure noise follows

$$dU_t = -bU_t dt + c dZ_t \tag{1.58}$$

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 $\frac{c^2}{2b}\left(1 - e^{-2b\Delta}\right)$. 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 be less 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 (although [11] have argued that serial correlation in the transaction type can be a component of the bid-ask spread, and extended the model of [39] to allow for it). On the other hand, the model (1.58) 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

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

where the $w_i's$ are i.i.d. $N(0, \sigma^2 \Delta)$, the $u_i's$ are independent of the $w_i's$, so we have $\mathrm{Var}\left[Y_i\right] = \sigma^2 \Delta + E\left[u_i^2\right]$, and they are Gaussian with mean zero and variance

$$E\left[u_i^2\right] = E\left[\left(U_{\tau_i} - U_{\tau_{i-1}}\right)^2\right] = \frac{c^2\left(1 - e^{-b\Delta}\right)}{b} = c^2 \Delta + o(\Delta)\,, \tag{1.59}$$

instead of $2a^2$.

In addition, the $u_i's$ are now serially correlated at all lags since

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

for $i \geq k$. The first order correlation of the log-returns is now

$$\mathrm{Cov}\left(Y_i, Y_{i-1}\right) = -\frac{c^2\left(1 - e^{-b\Delta}\right)^2}{2b} = -\frac{c^2 b}{2}\Delta^2 + o\left(\Delta^2\right)\,,$$

instead of η.

The result analogous to Theorem 1 is as follows. If one ignores the presence of this type of serially correlated noise when estimating σ^2, then:

Theorem 5. *In small samples (finite T), the RMSE of the estimator $\hat{\sigma}^2$ is given by*

$$\begin{aligned} RMSE\left[\hat{\sigma}^2\right] &= \left(\frac{c^4\left(1 - e^{-b\Delta}\right)^2}{b^2\Delta^2} + \frac{c^4\left(1 - e^{-b\Delta}\right)^2\left(\frac{T}{\Delta}e^{-2b\Delta} - 1 + e^{-2Tb}\right)}{T^2 b^2\left(1 + e^{-b\Delta}\right)^2} + \right. \\ &\quad \left. + \frac{2}{T\Delta}\left(\sigma^2\Delta + \frac{c^2\left(1 - e^{-b\Delta}\right)}{b}\right)^2\right)^{1/2} \qquad (1.60) \\ &= c^2 - \frac{bc^2}{2}\Delta + \frac{\left(\sigma^2 + c^2\right)^2 \Delta}{c^2 T} + O(\Delta^2) + O\left(\frac{1}{T^2}\right), \end{aligned}$$

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

One would expect this type of noise to be not nearly as bad as i.i.d. noise for the purpose of inferring σ^2 from high frequency data. Indeed, 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 as much noise as previously ($O(\Delta)$ vs. $O(1)$) and the squared bias β^2 of the estimator $\hat{\sigma}^2$ no longer diverges to infinity as $\Delta \to 0$: it has the finite limit c^4. Nevertheless, β^2 first decreases as Δ increases from 0, since

$$\beta^2 = \left(E\left[\hat{\sigma}^2\right] - \sigma^2\right)^2 = \frac{c^4\left(1 - e^{b\Delta}\right)^2}{b^2\Delta^2 e^{2b\Delta}} ,$$

and $\partial b_2/\partial\Delta \to -bc^4 < 0$ as $\Delta \to 0$. For large enough T, this is sufficient to generate a finite optimal sampling frequency.

To calibrate the parameter values b and c, we refer to the same empirical microstructure studies we mentioned in Section 1.4. We now have $\pi = E\left[u_i^2\right] / (\sigma^2\Delta + E\left[u_i^2\right])$ as the proportion of total variance that is microstructure-induced; we match it to the numbers in (1.24) from [35]. In their Table 5, they report the first order correlation of price changes (hence returns) to be approximately $\rho = -0.2$ at their frequency of observation. Here $\rho = \mathrm{Cov}\left(Y_i, Y_{i-1}\right) / \mathrm{Var}\left[Y_i\right]$. If we match $\pi = 0.6$ and $\rho = -0.2$, with $\sigma = 30\%$ as before, we obtain (after rounding) $c = 0.5$ and $b = 3 \times 10^4$. Figure 1.6 displays the resulting RMSE of the estimator as a function of Δ and T. The overall picture is comparable to Figure 1.2.

As for the rest of the analysis of the paper, dealing with likelihood corrections for microstructure noise, the covariance matrix of the log-returns, $\gamma^2 V$ in (1.26), should be replaced by the matrix whose diagonal elements are

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

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

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

Having modified the matrix $\gamma^2 V$, the artificial "normal" distribution that assumes i.i.d. $U's$ that are $N(0, \alpha^2)$ would no longer use the correct second moment structure of the data. Thus we cannot relate a priori the asymptotic variance of the estimator of the estimator $\hat{\sigma}^2$ to that of the i.i.d. Normal case, as we did in Theorem 2.

1.9.3 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 [20]. 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 (1.26) with

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

where δ_{ij} is the Kronecker symbol.

The small sample properties of the misspecified MLE for σ^2 analogous to those computed in the independent case, including its RMSE, can be obtained from

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

Specific expressions for all these quantities depend upon the assumptions of the particular structural model under consideration: for instance, in the [20] 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 $\sigma\left(W_{\tau_i} - W_{\tau_{i-1}}\right)$ is also uncorrelated with the transaction noises $U_{\tau_{i+1}}$ and $U_{\tau_{i-2}}$. With these in hand, the analysis of the RMSE and its minimum can then proceed as above. As for the likelihood corrections for microstructure noise, 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 and the likelihood must be modified accordingly.

1.9.4 Stochastic Volatility

One important departure from our basic model is the case where volatility is stochastic. The observed log-returns are still generated by equation (1.3). Now, however, the constant volatility assumption (1.2) is replaced by

$$dX_t = \sigma_t dW_t \,. \tag{1.61}$$

The object of interest in much of the literature on high frequency volatility estimation (see e.g., [8] and [6]) is then the integral

$$\int_0^T \sigma_t^2 dt \tag{1.62}$$

over a fixed time period $[0, T]$, or possibly several such time periods. The estimation is based on observations $0 = t_0 < t_1 < \ldots < t_n = T$, and asymptotic results are obtained when $\max \Delta t_i \to 0$. The usual estimator for (1.62) is the "realized variance"

$$\sum_{i=1}^{n} (\tilde{X}_{t_{i+1}} - \tilde{X}_{t_i})^2. \tag{1.63}$$

In the context of stochastic volatility, ignoring market microstructure noise leads to an even more dangerous situation than when σ is constant and $T \to \infty$. We show in the companion paper [43] that, after suitable scaling, the realized variance is a consistent and asymptotically normal estimator – but of the quantity $2a^2$. This quantity has, in general, nothing to do with the object of interest (1.62). Said differently, market microstructure noise totally swamps the variance of the price signal at the level of the realized variance. To obtain a finite optimal sampling interval, one needs that $a^2 \to 0$ as $n \to \infty$, that is the amount of noise must disappear asymptotically. For further developments on this topic, we refer to [43].

1.10 Conclusions

We showed that the presence of market microstructure noise makes it optimal to sample less often than would otherwise be the case in the absence of noise, and we determined accordingly the optimal sampling frequency in closed-form.

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 remains the case 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, but the estimator has the same asymptotic variance as if the noise distribution had been correctly specified. This robustness result is, we think, 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. Hence, the answer to the question we pose in our title is "as often as possible," provided one accounts for the presence of the noise when designing the estimator.

Appendix A – Proof of Lemma 1

To calculate the fourth cumulant $\mathrm{Cum}(Y_i, Y_j, Y_k, Y_l)$, recall from (1.8) that the observed log-returns are

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

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

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

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,

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

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

$$\begin{aligned}
&\mathrm{Cum}(U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_i} - U_{\tau_{i-1}}) \\
&= \mathrm{Cum}_4(U_{\tau_i} - U_{\tau_{i-1}}) \\
&= \mathrm{Cum}_4(U_{\tau_i}) + \mathrm{Cum}_4(-U_{\tau_{i-1}}) \\
&= 2 \;\mathrm{Cum}_4\left[U\right] ,
\end{aligned}$$

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.

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

$$\mathrm{Cum}(U_{\tau_m} - U_{\tau_{m-1}}, U_{\tau_{m+1}} - U_{\tau_m}, U_{\tau_{m+1}} - U_{\tau_m}, U_{\tau_{m+1}} - U_{\tau_m}) = -\,\mathrm{Cum}_4(U_{\tau_m}) .$$

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

$$\mathrm{Cum}(U_{\tau_m} - U_{\tau_{m-1}}, U_{\tau_m} - U_{\tau_{m-1}}, U_{\tau_{m+1}} - U_{\tau_m}, U_{\tau_{m+1}} - U_{\tau_m}) = \mathrm{Cum}_4(U_{\tau_m}) .$$

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

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

by independence of the $U's$.

Appendix B –Proof of Theorem 1

Given The estimator (1.5) has the following expected value

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

The estimator's variance is

$$\operatorname{Var}\left[\hat{\sigma}^2\right] = \frac{1}{T^2}\operatorname{Var}\left[\sum_{i=1}^{N} Y_i^2\right] = \frac{1}{T^2}\sum_{i,j=1}^{N}\operatorname{Cov}\left(Y_i^2, Y_j^2\right).$$

Applying Lemma 1 in the special case where the first two indices and the last two respectively are identical yields

$$\operatorname{Cum}(Y_i, Y_i, Y_j, Y_j) = \begin{cases} 2\ \operatorname{Cum}_4[U] & \text{if } j = i\,, \\ \operatorname{Cum}_4[U] & \text{if } j = i+1 \text{ or } j = i-1\,, \\ 0 \text{ otherwise}. \end{cases} \tag{B.1}$$

In the middle case, i.e., whenever $j = i+1$ or $j = i-1$, the number s of indices that are equal to the minimum index is always 2. Combining (B.1) with (1.14), we have

$$\begin{aligned}
\operatorname{Var}\left[\hat{\sigma}^2\right] &= \frac{1}{T^2}\sum_{i=1}^{N}\operatorname{Cov}\left(Y_i^2, Y_i^2\right) + \frac{1}{T^2}\sum_{i=1}^{N-1}\operatorname{Cov}\left(Y_i^2, Y_{i+1}^2\right) + \frac{1}{T^2}\sum_{i=2}^{N}\operatorname{Cov}\left(Y_i^2, Y_{i-1}^2\right) \\
&= \frac{1}{T^2}\sum_{i=1}^{N}\left\{2\operatorname{Cov}(Y_i, Y_i)^2 + 2\ \operatorname{Cum}_4[U]\right\} + \\
&+ \frac{1}{T^2}\sum_{i=1}^{N-1}\left\{2\operatorname{Cov}(Y_i, Y_{i+1})^2 + \operatorname{Cum}_4[U]\right\} + \\
&+ \frac{1}{T^2}\sum_{i=2}^{N}\left\{2\operatorname{Cov}(Y_i, Y_{i-1})^2 + \operatorname{Cum}_4[U]\right\} \\
&= \frac{2N}{T^2}\left\{\operatorname{Var}[Y_i]^2 + \operatorname{Cum}_4[U]\right\} + \frac{2(N-1)}{T^2}\left\{2\operatorname{Cov}(Y_i, Y_{i-1})^2 + \operatorname{Cum}_4[U]\right\},
\end{aligned}$$

with $\operatorname{Var}[Y_i]$ and $\operatorname{Cov}(Y_i, Y_{i-1}) = \operatorname{Cov}(Y_i, Y_{i+1})$ given in (1.9)-(1.10), so that

$$\begin{aligned}
\operatorname{Var}\left[\hat{\sigma}^2\right] &= \frac{2N}{T^2}\left\{\left(\sigma^2\Delta + 2a^2\right)^2 + \operatorname{Cum}_4[U]\right\} + \frac{2(N-1)}{T^2}\left\{2a^4 + \operatorname{Cum}_4[U]\right\}, \\
&= \frac{2\left(\sigma^4\Delta^2 + 4\sigma^2\Delta a^2 + 6a^4 + 2\operatorname{Cum}_4[U]\right)}{T\Delta} - \frac{2\left(2a^4 + \operatorname{Cum}_4[U]\right)}{T^2},
\end{aligned}$$

since $N = T/\Delta$. The expression for the RMSE follows from those for the expected value and variance given in (1.17) and (1.19):

$$\begin{aligned}
RMSE\left[\hat{\sigma}^2\right] = \Bigg(&\frac{4a^4}{\Delta^2} + \frac{2\left(\sigma^4\Delta^2 + 4\sigma^2\Delta a^2 + 6a^4 + 2\operatorname{Cum}_4[U]\right)}{T\Delta} - \\
&- \frac{2\left(2a^4 + \operatorname{Cum}_4[U]\right)}{T^2}\Bigg)^{1/2}.
\end{aligned} \tag{B.2}$$

The optimal value Δ^* of the sampling interval given in (1.20) is obtained by minimizing $RMSE\left[\hat{\sigma}^2\right]$ over Δ. The first order condition that arises from setting $\partial\,\mathrm{RMSE}\left[\hat{\sigma}^2\right]/\partial\Delta$ to 0 is the cubic equation in Δ :

$$\Delta^3 - \frac{2\left(3a^4 + \mathrm{Cum}_4\left[U\right]\right)}{\sigma^4}\Delta - \frac{4a^4 T}{\sigma^4} = 0\,. \tag{B.3}$$

We now show that (B.3) has a unique positive root, and that it corresponds to a minimum of $\mathrm{RMSE}\left[\hat{\sigma}^2\right]$. We are therefore looking for a real positive root in $\Delta = z$ to the cubic equation

$$z^3 + pz - q = 0\,, \tag{B.4}$$

where $q > 0$ and $p < 0$ since from (1.16):

$$3a^4 + \mathrm{Cum}_4\left[U\right] = 3a^4 + E\left[U^4\right] - 3E\left[U^2\right]^2 = E\left[U^4\right] > 0\,.$$

Using Vièta's change of variable from z to w given by $z = w - p/(3w)$ reduces, after multiplication by w^3, the cubic to the quadratic equation

$$y^2 - qy - \frac{p^3}{27} = 0 \tag{B.5}$$

in the variable $y \equiv w^3$.

Define the discriminant

$$D = \left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2.$$

The two roots of (B.5)

$$y_1 = \frac{q}{2} + D^{1/2} \quad \text{and} \quad y_2 = \frac{q}{2} - D^{1/2}$$

are real if $D \geq 0$ (and distinct if $D > 0$) and complex conjugates if $D < 0$. Then the three roots of (B.4) are

$$\begin{aligned}
z_1 &= y_1^{1/3} + y_2^{1/3}\,, \\
z_2 &= -\frac{1}{2}\left(y_1^{1/3} + y_2^{1/3}\right) + \mathrm{i}\frac{3^{1/2}}{2}\left(y_1^{1/3} - y_2^{1/3}\right), \\
z_3 &= -\frac{1}{2}\left(y_1^{1/3} + y_2^{1/3}\right) - \mathrm{i}\frac{3^{1/2}}{2}\left(y_1^{1/3} - y_2^{1/3}\right),
\end{aligned} \tag{B.6}$$

(see e.g., Section 3.8.2 in [1]). If $D > 0$, the two roots in y are both real and positive because $p < 0$ and $q > 0$ imply

$$y_1 > y_2 > 0$$

and hence of the three roots given in (B.6), z_1 is real and positive and z_2 and z_3 are complex conjugates. If $D = 0$, then $y_1 = y_2 = q/2 > 0$ and the three roots are real (two of which are identical) and given by

$$\begin{aligned}
z_1 &= y_1^{1/3} + y_2^{1/3} = 2^{2/3} q^{1/3}\,, \\
z_2 &= z_3 = -\frac{1}{2}\left(y_1^{1/3} + y_2^{1/3}\right) = -\frac{1}{2} z_1\,.
\end{aligned}$$

Of these, $z_1 > 0$ and $z_2 = z_3 < 0$. If $D < 0$, the three roots are distinct and real because

$$y_1 = \frac{q}{2} + \mathrm{i}(-D)^{1/2} \equiv re^{\mathrm{i}\theta}, \quad y_2 = \frac{q}{2} - \mathrm{i}(-D)^{1/2} \equiv re^{-\mathrm{i}\theta},$$

so

$$y_1^{1/3} = r^{1/3}e^{\mathrm{i}\theta/3}, \quad y_2^{1/3} = r^{1/3}e^{-\mathrm{i}\theta/3},$$

and therefore

$$y_1^{1/3} + y_2^{1/3} = 2r^{1/3}\cos(\theta/3), \quad y_1^{1/3} - y_2^{1/3} = 2\mathrm{i}r^{1/3}\sin(\theta/3),$$

so that

$$\begin{aligned} z_1 &= 2r^{1/3}\cos(\theta/3) \\ z_2 &= -r^{1/3}\cos(\theta/3) + 3^{1/2}r^{1/3}\sin(\theta/3) \\ z_3 &= -r^{1/3}\cos(\theta/3) - 3^{1/2}r^{1/3}\sin(\theta/3) . \end{aligned}$$

Only z_1 is positive because $q > 0$ and $(-D)^{1/2} > 0$ imply that $0 < \theta < \pi/2$. Therefore $\cos(\theta/3) > 0$, so $z_1 > 0$; $\sin(\theta/3) > 0$, so $z_3 < 0$; and

$$\cos(\theta/3) > \cos(\pi/6) = \frac{3^{1/2}}{2} = 3^{1/2}\sin(\pi/6) > 3^{1/2}\sin(\theta/3),$$

so $z_2 < 0$.

Thus the equation (B.4) has exactly one root that is positive, and it is given by z_1 in (B.6). Since RMSE $[\hat{\sigma}^2]$ is of the form

$$\begin{aligned} &\mathrm{RMSE}\left[\hat{\sigma}^2\right] \\ &= \left(\frac{2T\Delta^3\sigma^4 - 2\Delta^2\left(2a^4 - 4a^2T\sigma^2 + \mathrm{Cum}_4[U]\right) + 2\Delta\left(6a^4T + 2T\,\mathrm{Cum}_4[U]\right) + 4a^4T^2}{T^2\Delta^2}\right)^{1/2} \\ &= \left(\frac{a_3\Delta^3 + a_2\Delta^2 + a_1\Delta + a_0}{T^2\Delta^2}\right)^{1/2}, \end{aligned}$$

with $a_3 > 0$, it tends to $+\infty$ when Δ tends to $+\infty$. Therefore that single positive root corresponds to a minimum of RMSE $[\hat{\sigma}^2]$ which is reached at

$$\begin{aligned} \Delta^* &= y_1^{1/3} + y_2^{1/3} \\ &= \left(\frac{q}{2} + D^{1/2}\right)^{1/3} + \left(\frac{q}{2} - D^{1/2}\right)^{1/3}. \end{aligned}$$

Replacing q and p by their values in the expression above yields (1.20). As shown above, if the expression inside the square root in formula (1.20) is negative, the resulting Δ^* is still a positive real number.

Appendix C –Proof of Proposition 1

The result follows from an application of the delta method to the known properties of the MLE estimator of an MA(1) process (Section 5.4 in [25]), as follows. Because we re-use these calculations below in the proof of Theorem 2 (whose result cannot be inferred from known MA(1) properties), we recall some of the expressions of the score vector of the MA(1) likelihood. The partial derivatives of the log-likelihood function (1.25) have the form

$$\dot{l}_\eta = -\frac{1}{2}\frac{\partial \ln \det(V)}{\partial \eta} - \frac{1}{2\gamma^2}Y'\frac{\partial V^{-1}}{\partial \eta}Y\,, \tag{C.1}$$

and

$$\dot{l}_{\gamma^2} = -\frac{N}{2\gamma^2} + \frac{1}{2\gamma^4}Y'V^{-1}Y\,. \tag{C.2}$$

so that the MLE for γ^2 is

$$\hat{\gamma}^2 = \frac{1}{N}Y'V^{-1}Y\,. \tag{C.3}$$

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

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

thus as $N \to \infty$

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

Similarly, it follows from (C.2) that

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

Turning now to Fisher's information, we have

$$E\left[-\ddot{l}_{\gamma^2\gamma^2}\right] = -\frac{N}{2\gamma^4} + \frac{1}{\gamma^6}E\left[Y'V^{-1}Y\right] = \frac{N}{2\gamma^4}\,, \tag{C.4}$$

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

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

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

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

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

and evaluate both terms. For the first term in (C.6), we have from (1.27):

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

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

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

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

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

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

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

In light of that and (C.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 [15] 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 (1.9)-(1.10):

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

Hence

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

where

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

Appendix D – Proof of Theorem 2

We have that

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

where "true" denotes the true distribution of the $Y's$, not the incorrectly specified normal distribution, and Cum denotes the cumulants given in Lemma 1. The last transition is because

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

since Y has mean zero (see e.g., Section 2.3 of [36]). 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 (D.1), note that $\mathrm{Cum}_{\text{normal}}(Y_i, Y_j, Y_k, Y_l) = 0$, 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

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

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

$$
E_{\text{normal}}\left[\dot{l}_\eta \dot{l}_{\gamma^2}\right] = -E_{\text{normal}}\left[\ddot{l}_{\eta\gamma^2}\right] = -E_{\text{true}}\left[\ddot{l}_{\eta\gamma^2}\right] , \qquad (D.2)
$$

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 (D.1) that

$$\begin{aligned} E_{\text{true}}\left[\dot{l}_{\eta}\dot{l}_{\gamma^2}\right] &= E_{\text{normal}}\left[\dot{l}_{\eta}\dot{l}_{\gamma^2}\right] - \frac{1}{4\gamma^6}\sum_{i,j,k,l=1}^{N}\frac{\partial v^{ij}}{\partial\eta}v^{kl}\,\text{Cum}_{\text{true}}(Y_i,Y_j,Y_k,Y_l) \\ &= -E_{\text{true}}\left[\ddot{l}_{\eta\gamma^2}\right] - \frac{1}{4\gamma^6}\sum_{i,j,k,l=1}^{N}\frac{\partial v^{ij}}{\partial\eta}v^{kl}\,\text{Cum}_{\text{true}}(Y_i,Y_j,Y_k,Y_l)\,. \end{aligned} \tag{D.3}$$

It follows similarly that

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

and

$$\begin{aligned} E_{\text{true}}\left[\left(\dot{l}_{\gamma^2}\right)^2\right] &= \text{Var}_{\text{true}}(\dot{l}_{\gamma^2}) \\ &= -E_{\text{true}}\left[\ddot{l}_{\gamma^2\gamma^2}\right] + \frac{1}{4\gamma^8}\sum_{i,j,k,l=1}^{N}v^{ij}v^{kl}\,\text{Cum}_{\text{true}}(Y_i,Y_j,Y_k,Y_l)\,. \end{aligned} \tag{D.5}$$

We now need to evaluate the sums that appear on the right hand sides of (D.3), (D.4) and (D.5). Consider two generic symmetric $N\times N$ matrices $\left[\nu^{i,j}\right]$ and $\left[\omega^{i,j}\right]$. We are interested in expressions of the form

$$\begin{aligned} \sum_{i,j,k,l:M=m+1}(-1)^s\nu^{i,j}\omega^{k,l} &= \sum_{h=1}^{N-1}\sum_{i,j,k,l:m=h,M=h+1}(-1)^s\nu^{i,j}\omega^{k,l} \\ &= \sum_{h=1}^{N-1}\sum_{r=1}^{3}\sum_{i,j,k,l:m=h,M=h+1,s=r}(-1)^r\nu^{i,j}\omega^{k,l} \\ &= \sum_{h=1}^{N-1}\Big\{-2\nu^{h,h+1}\omega^{h+1,h+1} - 2\nu^{h+1,h+1}\omega^{h,h+1} + \\ &\quad + \nu^{h,h}\omega^{h+1,h+1} + \nu^{h+1,h+1}\omega^{h,h} + 4\nu^{h,h+1}\omega^{h,h+1} - \\ &\quad -2\nu^{h+1,h}\omega^{h,h} - 2\nu^{h,h}\omega^{h+1,h}\Big\}\,. \end{aligned} \tag{D.6}$$

It follows that if we set

$$\Upsilon(\nu,\omega) = \sum_{i,j,k,l=1}^{N}\nu^{i,j}\omega^{k,l}\,\text{Cum}_{\text{true}}(Y_i,Y_j,Y_k,Y_l)\,, \tag{D.7}$$

then $\Upsilon(\nu,\omega) = \text{Cum}_4\left[U\right]\,\Psi(\nu,\omega)$ where

$$\begin{aligned} \psi(\nu,\omega) &= 2\sum_{h=1}^{N}\nu^{h,h}\omega^{h,h} + \sum_{h=1}^{N-1}\Big\{-2\nu^{h,h+1}\omega^{h+1,h+1} - 2\nu^{h+1,h+1}\omega^{h,h+1} + \\ &\quad + \nu^{h,h}\omega^{h+1,h+1} + \nu^{h+1,h+1}\omega^{h,h} + 4\nu^{h,h+1}\omega^{h,h+1} - \\ &\quad -2\nu^{h+1,h}\omega^{h,h} - 2\nu^{h,h}\omega^{h+1,h}\Big\}\,. \end{aligned} \tag{D.8}$$

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 (D.8) simplifies to:

$$\psi(\nu,\omega) = 2\sum_{h=1}^{N} \nu^{h,h}\omega^{h,h} + \sum_{h=1}^{N-1}\Big\{-4\nu^{h,h+1}\omega^{h+1,h+1} - 4\nu^{h+1,h+1}\omega^{h,h+1} + \\ +2\nu^{h,h}\omega^{h+1,h+1} + 4\nu^{h,h+1}\omega^{h,h+1}\Big\}.$$

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 (1.28), and consequently for $\partial v^{i,j}/\partial\eta$.

If we wish to compute the sums in equations (D.3), (D.4), and (D.5), 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 (1.28), we obtain:

$$\begin{aligned}\psi(v,v) &= \frac{2}{(1+\eta^2)(1-\eta)^3(1-\eta^{2(1+N)})^2} \\ &\quad \Big\{-(1+\eta)\left(1-\eta^{2N}\right)\left(1+2\eta^2+2\eta^{2(1+N)}+\eta^{2(2+N)}\right) + \\ &\quad + N(1-\eta)\left(1+\eta^2\right)\left(2+\eta^{2N}+\eta^{2(1+N)}+6\eta^{1+2N}+2\eta^{2+4N}\right)\Big\} \\ &= \frac{4N}{(1-\eta)^2} + o(N), \qquad \text{(D.9)}\end{aligned}$$

$$\begin{aligned}\psi\left(\frac{\partial v}{\partial\eta}, v\right) &= \frac{2\left(O(1)+2N(1-\eta)\left(1+\eta^2\right)\eta\left(1+\eta^2+O(\eta^{2N})\right)+N^2O(\eta^{2N})\right)}{\eta(1-\eta)^4(1+\eta^2)^2(1-\eta^{2(1+N)})^3} \\ &= \frac{4N}{(1-\eta)^3} + o(N), \qquad \text{(D.10)}\end{aligned}$$

$$\begin{aligned}&\psi\left(\frac{\partial v}{\partial\eta}, \frac{\partial v}{\partial\eta}\right) \\ &= \frac{4\left(O(1)+3N\left(1-\eta^4\right)\eta^2\left(\left(1+\eta^2\right)^2+O(\eta^{2N})\right)+N^2O(\eta^{2N})+N^3O(\eta^{2N})\right)}{3\eta^2(1+\eta)(1+\eta^2)^3(1-\eta)^5(1-\eta^{2(1+N)})^4} \\ &= \frac{4N}{(1-\eta)^4} + o(N). \qquad \text{(D.11)}\end{aligned}$$

The asymptotic variance of the estimator $(\hat{\gamma}^2, \hat{\eta})$ obtained by maximizing the (incorrectly-specified) log-likelihood (1.25) that assumes Gaussianity of the $U's$ is given by

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

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

$$\begin{aligned}D = D' &= -\frac{1}{N}E_{\text{true}}\left[\ddot{l}\right] = -\frac{1}{N}E_{\text{normal}}\left[\ddot{l}\right] = \frac{1}{N}E_{\text{normal}}\left[\dot{l}\dot{l}'\right] \\ &= \begin{pmatrix} \frac{1}{2\gamma^4} & -\frac{\eta}{N\gamma^2(1-\eta^2)}+o\left(\frac{1}{N}\right) \\ \bullet & \frac{1}{(1-\eta^2)}+o(1)\end{pmatrix}, \qquad \text{(D.12)}\end{aligned}$$

and, in light of (D.3), (D.4), and (D.5),

$$S = \frac{1}{N} E_{\text{true}} \left[\dot{l} \dot{l}' \right] = -\frac{1}{N} E_{\text{true}} \left[\ddot{l} \right] + \text{Cum}_4 [U] \;\; \Psi = D + \text{Cum}_4 [U] \;\; \Psi , \tag{D.13}$$

where

$$\begin{aligned} \Psi &= \frac{1}{4N} \begin{pmatrix} \frac{1}{\gamma^8} \psi (v, v) & -\frac{1}{\gamma^6} \psi \left(\frac{\partial v}{\partial \eta}, v \right) \\ \bullet & \frac{1}{\gamma^4} \psi \left(\frac{\partial v}{\partial \eta}, \frac{\partial v}{\partial \eta} \right) \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\gamma^8 (1-\eta)^2} + o(1) & \frac{-1}{\gamma^6 (1-\eta)^3} + o(1) \\ \bullet & \frac{1}{\gamma^4 (1-\eta)^4} + o(1) \end{pmatrix} , \end{aligned} \tag{D.14}$$

from the expressions just computed. It follows that

$$\begin{aligned} \text{AVAR}_{\text{true}}(\hat{\gamma}^2, \hat{\eta}) &= \Delta \;\; \left(D \left(D + \text{Cum}_4 [U] \;\; \Psi \right)^{-1} D \right)^{-1} \\ &= \Delta \;\; \left(D \left(Id + \text{Cum}_4 [U] \;\; D^{-1} \Psi \right)^{-1} \right)^{-1} \\ &= \Delta \;\; \left(Id + \text{Cum}_4 [U] \;\; D^{-1} \Psi \right) D^{-1} \\ &= \Delta \;\; \left(Id + \text{Cum}_4 [U] \;\; D^{-1} \Psi \right) D^{-1} \\ &= \text{AVAR}_{\text{normal}}(\hat{\gamma}^2, \hat{\eta}) + \Delta \;\; \text{Cum}_4 [U] \;\; D^{-1} \Psi D^{-1}, \end{aligned}$$

where Id denotes the identity matrix and

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

so that

$$\text{AVAR}_{\text{true}}(\hat{\gamma}^2, \hat{\eta}) = \Delta \begin{pmatrix} 2\gamma^4 & 0 \\ 0 & (1 - \eta^2) \end{pmatrix} + \Delta \, \text{Cum}_4 [U] \begin{pmatrix} \frac{4}{(1-\eta)^2} & \frac{-2(1+\eta)}{\gamma^2 (1-\eta)^2} \\ \bullet & \frac{(1+\eta)^2}{\gamma^4 (1-\eta)^2} \end{pmatrix} .$$

By applying the delta method to change the parametrization, we now recover the asymptotic variance of the estimates of the original parameters:

$$\begin{aligned} &\text{AVAR}_{\text{true}}(\hat{\sigma}^2, \hat{a}^2) \\ &= \nabla f(\gamma^2, \eta) . \, \text{AVAR}_{\text{true}}(\hat{\gamma}^2, \hat{\eta}) . \nabla f(\gamma^2, \eta)' \\ &= \begin{pmatrix} 4\sqrt{\sigma^6 \Delta \left(4a^2 + \sigma^2 \Delta \right)} + 2\sigma^4 \Delta & -\sigma^2 \Delta h(\Delta, \sigma^2, a^2) \\ \bullet & \frac{\Delta}{2} \left(2a^2 + \sigma^2 \Delta \right) h(\Delta, \sigma^2, a^2) + \Delta \, \text{Cum}_4 [U] \end{pmatrix} . \end{aligned}$$

Appendix E –Derivations for Section 1.7

To see (1.39), let "orig" (E.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 (C.10), namely

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

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

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

where $H[\dot{l}_{\text{transf}}]$ is a 2×2 matrix whose terms are linear in $\dot{l}_{\text{transf}}$ 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

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

with

$$D_{\text{transf}} = N^{-1}E_{\text{true}}[-\ddot{l}_{\text{transf}}] ,$$

given in (D.12). Similarly,

$$\dot{l}_{\text{orig}}\dot{l}'_{\text{orig}} = \nabla f_{\text{inv}}(\sigma^2, \alpha^2)'.\dot{l}_{\text{transf}}\dot{l}'_{\text{transf}}.\nabla f_{\text{inv}}(\sigma^2, \alpha^2) ,$$

and so

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

with the second equality following from the expression for S_{transf} given in (D.13).

To complete the calculation, note from (D.14) that

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

where

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

Thus

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

where

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

which is the result (1.40). Inserting (E.4) into (E.3) yields the result (1.39).

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

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

so that

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

The profile score then follows

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

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

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

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

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

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

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

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

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

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

Further differentiating (E.7), one obtains

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

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

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

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 (1.42) is valid.

Alternatively, we can see that the profile likelihood λ satisfies the Bartlett identity to first order, i.e., (1.43). Note that by (E.8),

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

so that

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

by invoking (1.39).

Continuing the calculation,

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

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

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

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

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

and (1.43) follows from combining (E.10) with (E.12).

Appendix F –Proof of Lemma 2

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

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

and, since Σ is linear in the parameters σ^2 and a^2 (see (1.45)) we have

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

so that

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

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

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

with the second equality following from (1.46). Then, for any nonrandom Q, we have

$$E\left[Y'QY\right] = Tr\left[QE\left[YY'\right]\right] = Tr\left[Q\Sigma\right] . \tag{F.5}$$

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

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

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

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

again because of (F.2).

In light of (1.46), the expected information (conditional on the $\Delta's$) is given by

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

Then,

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

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

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

Appendix G – Proof of Theorem 3

In light of (1.45) and (1.52),

$$\Sigma = \Sigma_0 + \varepsilon\sigma^2\Xi, \qquad \text{(G.1)}$$

from which it follows that

$$\begin{aligned}
\Sigma^{-1} &= \left(\Sigma_0\left(Id + \varepsilon\sigma^2\Sigma_0^{-1}\Xi\right)\right)^{-1} \\
&= \left(Id + \varepsilon\sigma^2\Sigma_0^{-1}\Xi\right)^{-1}\Sigma_0^{-1} \\
&= \Sigma_0^{-1} - \varepsilon\sigma^2\Sigma_0^{-1}\Xi\Sigma_0^{-1} + \varepsilon^2\sigma^4\left(\Sigma_0^{-1}\Xi\right)^2\Sigma_0^{-1} + O(\varepsilon^3), \qquad \text{(G.2)}
\end{aligned}$$

since

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

Also,

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

Therefore, recalling (F.6), we have

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

We now consider the behavior as $N \to \infty$ of the terms up to order ε^2. The remainder term is handled similarly.

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

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

Similarly, the coefficient of order ε^2 is

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

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

$$E\left[\Xi \Sigma_0^{-1} \Xi\right] = \Delta_0^2 \operatorname{Var}[\xi] \; diag\left[\Sigma_0^{-1}\right], \tag{G.4}$$

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

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

To calculate $E\left[\ddot{l}_{\beta_2 \beta_1}\right]$, in light of (1.51), we need to differentiate $E[\partial \ln \det \Sigma / \partial \beta_1]$ with respect to β_2. Indeed

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

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 $\Delta_i' s$, which is independent of the β parameters (i.e., σ^2 and a^2). Therefore, differentiating (G.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 (G.5), we therefore have

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

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 (1.9)-(1.10) and (1.26), except that Δ is replaced by Δ_0 in these expressions. To compute $\phi^{(0)}$, we start with

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

with $\partial\gamma^2/\partial\beta_1$ and $\partial\eta/\partial\beta_1$ to be computed from (1.11)-(1.12). 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

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

Therefore the first term in (G.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 (G.7) with respect to β_2. Indeed, given the form of (G.8), we have that

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

since the remainder term in (G.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

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

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 (G.9), we have

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

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

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

so that

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

We have

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

Next, we compute separately

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

and

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

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 (G.11),

$$
\alpha=\sigma^4 Tr\left[\Sigma_0^{-1}\operatorname{diag}\left[\Sigma_0^{-1}\right]\Sigma_0^{-1}\frac{\partial\Sigma_0}{\partial\beta_1}\right]-\sigma^2\frac{\partial\sigma^2}{\partial\beta_1}Tr\left[\Sigma_0^{-1}\operatorname{diag}\left[\Sigma_0^{-1}\right]\right], \tag{G.12}
$$

the first trace becomes

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

so that we have

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

where the calculation of $Tr\left[V^{-1}diag\left[V^{-1}\right]\right]$ 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

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

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

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

Putting it all together, we have obtained

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

where

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

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

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

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

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 (G.15), are given in the statement of the Theorem.

Appendix H – Proof of Theorem 4

It follows as in (D.3), (D.4) and (D.5) that

$$
\begin{aligned}
&E_{\text{true}}\left[\dot{l}_{\beta_1}\dot{l}_{\beta_2}|\Delta\right] \\
&= \mathrm{Cov}_{\text{true}}(\dot{l}_{\beta_1},\dot{l}_{\beta_2}|\Delta) \\
&= \mathrm{Cov}_{\text{true}}\left(-\frac{1}{2}\sum_{i,j=1}^{N} Y_iY_j\left(\frac{\partial\Sigma^{-1}}{\partial\beta_1}\right)_{ij}, -\frac{1}{2}\sum_{k,l=1}^{N} Y_kY_l\left(\frac{\partial\Sigma^{-1}}{\partial\beta_2}\right)_{kl}\right) \\
&= \frac{1}{4}\sum_{i,j,k,l=1}^{N}\left(\frac{\partial\Sigma^{-1}}{\partial\beta_1}\right)_{ij}\left(\frac{\partial\Sigma^{-1}}{\partial\beta_2}\right)_{kl}\mathrm{Cov}_{\text{true}}(Y_iY_j,Y_kY_l|\Delta) \\
&= -E_{\text{true}}\left[\ddot{l}_{\beta_1\beta_2}|\Delta\right] + \frac{1}{4}\sum_{i,j,k,l=1}^{N}\left(\frac{\partial\Sigma^{-1}}{\partial\beta_1}\right)_{ij}\left(\frac{\partial\Sigma^{-1}}{\partial\beta_2}\right)_{kl}\mathrm{Cum}_{\text{true}}(Y_i,Y_j,Y_k,Y_l|\Delta) \\
&= -E_{\text{true}}\left[\ddot{l}_{\beta_1\beta_2}|\Delta\right] + \frac{1}{4}\,\mathrm{Cum}_4\,[U]\,\psi\left(\frac{\partial\Sigma^{-1}}{\partial\beta_1},\frac{\partial\Sigma^{-1}}{\partial\beta_2}\right), \qquad \text{(H.1)}
\end{aligned}
$$

since $\mathrm{Cum}_{\text{true}}(Y_i,Y_j,Y_k,Y_l|\Delta) = 2, \pm 1$, or 0, $\times\,\mathrm{Cum}_{\text{true}}(U)$, as in (1.15), and with ψ defined in (D.8). Taking now unconditional expectations, we have

$$
\begin{aligned}
E_{\text{true}}\left[\dot{l}_{\beta_1}\dot{l}_{\beta_2}\right] &= \mathrm{Cov}_{\text{true}}(\dot{l}_{\beta_1},\dot{l}_{\beta_2}) \\
&= E\left[\mathrm{Cov}_{\text{true}}(\dot{l}_{\beta_1},\dot{l}_{\beta_2}|\Delta)\right] + \mathrm{Cov}_{\text{true}}(E_{\text{true}}[\dot{l}_{\beta_1}|\Delta], E_{\text{true}}[\dot{l}_{\beta_2}|\Delta]) \\
&= E\left[\mathrm{Cov}_{\text{true}}(\dot{l}_{\beta_1},\dot{l}_{\beta_2}|\Delta)\right] \\
&= -E_{\text{true}}\left[\ddot{l}_{\beta_1\beta_2}\right] + \frac{1}{4}\,\mathrm{Cum}_4\,[U]\,E\left[\psi\left(\frac{\partial\Sigma^{-1}}{\partial\beta_1},\frac{\partial\Sigma^{-1}}{\partial\beta_2}\right)\right], \qquad \text{(H.2)}
\end{aligned}
$$

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 (G.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 (G.2), we have for $i = 1, 2$

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

and therefore by bilinearity of ψ we have

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

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 (H.3),

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

corresponds to the equally spaced, misspecified noise distribution, situation studied in Section 1.6.

The second term, linear in ε, is zero since

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

with the first equality following from the bilinearity of ψ, the second from the fact that the unconditional expectation over the $\Delta_i' s$ 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 $E\left[\Xi\right]=0$.

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

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

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

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

and

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

from which ψ in (H.4) can be evaluated through the sum given in (D.8).

Summing these terms, we obtain

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

where

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

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

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

is obtained by first expressing

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

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

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

we recall our definition of $\psi(\nu,\omega)$ given in (D.8) whose unconditional expected value (over the $\Delta_i's$, 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

$$\begin{aligned}\tilde{\psi}(\lambda) \equiv 2\sum_{h=1}^{N}\lambda^{h,h,h,h} + \sum_{h=1}^{N-1}\Big\{&-2\lambda^{h,h+1,h+1,h+1} - 2\lambda^{h+1,h+1,h,h+1} +\\ &+\lambda^{h,h,h+1,h+1} + \lambda^{h+1,h+1,h,h} + 4\lambda^{h,h+1,h,h+1} -\\ &-2\lambda^{h+1,h,h,h} - 2\lambda^{h,h,h+1,h}\Big\},\end{aligned} \tag{H.5}$$

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 (D.8), 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

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

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

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

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

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

Thus we have

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

and

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

Summing these terms, we obtain

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

where

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

Putting it all together, we have

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

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

$$
\mathrm{AVAR}_{\mathrm{true}}(\hat{\sigma}^2, \hat{a}^2) = E[\Delta]\ \left(D'S^{-1}D\right)^{-1}, \tag{H.8}
$$

where

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

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

$$
S = \frac{1}{N}E_{\mathrm{true}}\left[\dot{l}\dot{l}'\right] = -\frac{1}{N}E_{\mathrm{true}}\left[\ddot{l}\right] + \mathrm{Cum}_4[U]\ \Psi = D + \mathrm{Cum}_4[U]\ \Psi\,,
$$

where

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

Since, from (H.3), we have

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

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.,

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

and

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

with

$$\frac{1}{4N}\alpha_1[2] = \operatorname{Var}[\xi]\begin{pmatrix} A & B \\ \bullet & C \end{pmatrix} + o(1),$$

with

$$\begin{aligned} A &= \frac{2\Delta_0^{3/2}\left(-4a^2+\Delta_0\sigma^2\right)}{\sigma\left(4a^2+\Delta_0\sigma^2\right)^{9/2}}\,, \\ B &= \frac{\Delta_0\left(\left(-4a^2+\Delta_0\sigma^2\right)\left(4a^2+\Delta_0\sigma^2\right)^{3/2} - \Delta_0^{1/2}\sigma\left(-40a^4+2a^2\Delta_0\sigma^2+\Delta_0^2\sigma^4\right)\right)}{2a^4\left(4a^2+\Delta_0\sigma^2\right)^{9/2}}\,, \\ C &= \frac{8\Delta_0\sigma^2\left(\left(4a^2+\Delta_0\sigma^2\right)^{3/2} - \Delta_0^{1/2}\sigma\left(6a^2+\Delta_0\sigma^2\right)\right)}{a^4\left(4a^2+\Delta_0\sigma^2\right)^{9/2}}\,, \end{aligned}$$

$$\frac{1}{4N}\alpha_2 = \operatorname{Var}[\xi] \\ \begin{pmatrix} \frac{\Delta_0^{3/2}\left(40a^8-12a^4\Delta_0{}^2\sigma^4+\Delta_0{}^4\sigma^8\right)}{2\sigma\left(2a^2+\Delta_0\sigma^2\right)^3\left(4a^2+\Delta_0\sigma^2\right)^{9/2}} & \frac{\Delta_0^{3/2}\sigma\left(-44a^6-18a^4\Delta_0\sigma^2+7a^2\Delta_0{}^2\sigma^6+3\Delta_0{}^3\sigma^6\right)}{\left(2a^2+\Delta_0\sigma^2\right)^3\left(4a^2+\Delta_0\sigma^2\right)^{9/2}} \\ \bullet & \frac{2\Delta_0^{3/2}\sigma^3\left(50a^{22}+42a^2\Delta_0\sigma^2+9\Delta_0{}^2\sigma^4\right)}{\left(2a^2+\Delta_0\sigma^2\right)^3\left(4a^2+\Delta_0\sigma^2\right)^{9/2}} \end{pmatrix} + \\ + o(1)\,.$$

It follows from (H.8) that

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

where

$$\mathrm{AVAR}_{\mathrm{normal}}(\hat{\sigma}^2, \hat{a}^2) = \Delta_0 D^{-1} = A^{(0)} + \varepsilon^2 A^{(2)} + O(\varepsilon^3)$$

is the result given in Theorem 3, namely (1.53).

The correction term due to the misspecification of the error distribution is determined by $\mathrm{Cum}_4[U]$ times

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

where the matrices are given in the text. The asymptotic variance is then given by

$$
\mathrm{AVAR_{true}}(\hat{\sigma}^2, \hat{a}^2) = \left(A^{(0)} + \mathrm{Cum}_4\left[U\right] B^{(0)}\right) + \varepsilon^2 \left(A^{(2)} + \mathrm{Cum}_4\left[U\right] B^{(2)}\right) + O(\varepsilon^3),
$$

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

Appendix I – Proof of Theorem 5

From

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

it follows that the estimator (1.5) has the following expected value

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

The estimator's variance is

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

Since the $Y_i's$ are normal with mean zero,

$$\operatorname{Var}\left[Y_i^2\right] = 2\operatorname{Var}[Y_i]^2 = 2E\left[Y_i^2\right]^2$$

and for $i > j$

$$\operatorname{Cov}\left(Y_i^2, Y_j^2\right) = 2\ \operatorname{Cov}\left(Y_i, Y_j\right)^2 = 2\ E\left[u_i u_j\right]^2,$$

since

$$\operatorname{Cov}\left(Y_i, Y_j\right) = E\left[Y_i Y_j\right] = E\left[\left(w_i + u_i\right)\left(w_j + u_j\right)\right] = E\left[u_i u_j\right].$$

Now we have

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

so that

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

and consequently

$$\begin{aligned}
\operatorname{Var}\left[\hat{\sigma}^2\right] = \frac{1}{T^2}\Bigg\{&\frac{c^4\left(1 - e^{-b\Delta}\right)^2\left(Ne^{-2b\Delta} - 1 + e^{-2Nb\Delta}\right)}{b^2\left(1 + e^{-b\Delta}\right)^2} + \\
&+ 2N\left(\sigma^2\Delta + \frac{c^2\left(1 - e^{-b\Delta}\right)}{b}\right)^2\Bigg\},
\end{aligned} \tag{I.2}$$

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

Tables

Value of a	T = 1 day	T = 1 year	T = 5 years
Panel A: $\sigma = 30\%$		**Stocks**	
0.01%	1 mn	4 mn	6 mn
0.05%	5 mn	31 mn	53 mn
0.1%	12 mn	1.3 hr	2.2 hr
0.15%	22 mn	2.2 hr	3.8 hr
0.2%	32 mn	3.3 hr	5.6 hr
0.3%	57 mn	5.6 hr	1.5 day
0.4%	1.4 hr	1.3 day	2.2 days
0.5%	2 hr	1.7 day	2.9 days
0.6%	2.6 hr	2.2 days	3.7 days
0.7%	3.3 hr	2.7 days	4.6 days
0.8%	4.1 hr	3.2 days	1.1 week
0.9%	4.9 hr	3.8 days	1.3 week
1.0%	5.9 hr	4.3 days	1.5 week
Panel B: $\sigma = 10\%$		**Currencies**	
0.005%	4 mn	23 mn	39 mn
0.01%	9 mn	58 mn	1.6 hr
0.02%	23 mn	2.4 hr	4.1 hr
0.05%	1.3 hr	8.2 hr	14.0 hr
0.10%	3.5 hr	20.7 hr	1.5 day

Table 1.1. Optimal Sampling Frequency

This table reports the optimal sampling frequency Δ^* given in equation (1.20) for different values of the standard deviation of the noise term a and the length of the sample T. Throughout the table, the noise is assumed to be normally distributed (hence $\mathrm{Cum}_4[U] = 0$ in formula (1.20)). In Panel A, the standard deviation of the efficient price process is $\sigma = 30\%$ per year, and at $\sigma = 10\%$ per year in Panel B. In both panels, 1 year = 252 days, but in Panel A, 1 day = 6.5 hours (both the NYSE and NASDAQ are open for 6.5 hours from 9:30 to 16:00 EST), while in Panel B, 1 day = 24 hours as is the case for major currencies. A value of $a = 0.05\%$ means that each transaction is subject to Gaussian noise with mean 0 and standard deviation equal to 0.05% of the efficient price. If the sole source of the noise were a bid/ask spread of size s, then a should be set to $s/2$. For example, a bid/ask spread of 10 cents on a $10 stock would correspond to $a = 0.05\%$. For the dollar/euro exchange rate, a bid/ask spread of $s = 0.04\%$ translates into $a = 0.02\%$. For the bid/ask model, which is based on binomial instead of Gaussian noise, $\mathrm{Cum}_4[U] = -s^4/8$, but this quantity is negligible given the tiny size of s.

Sampling Interval	Theoretical Mean	Sample Mean	Theoretical Stand. Dev.	Sample Stand. Dev.
5 minutes	0.185256	0.185254	0.00192	0.00191
15 minutes	0.121752	0.121749	0.00208	0.00209
30 minutes	0.10588	0.10589	0.00253	0.00254
1 hour	0.097938	0.097943	0.00330	0.00331
2 hours	0.09397	0.09401	0.00448	0.00440
1 day	0.09113	0.09115	0.00812	0.00811
1 week	0.0902	0.0907	0.0177	0.0176

Table 1.2. Monte Carlo Simulations: Bias and Variance when Market Microstructure Noise is Ignored

This table reports the results of $M = 10,000$ Monte Carlo simulations of the estimator $\hat{\sigma}^2$, with market microstructure noise present but ignored. The column "theoretical mean" reports the expected value of the estimator, as given in (1.17) and similarly for the column "theoretical standard deviation" (the variance is given in (1.19)). The "sample" columns report the corresponding moments computed over the M simulated paths. The parameter values used to generate the simulated data are $\sigma^2 = 0.3^2 = 0.09$ and $a^2 = (0.15\%)^2$ and the length of each sample is $T = 1$ year.

Figures

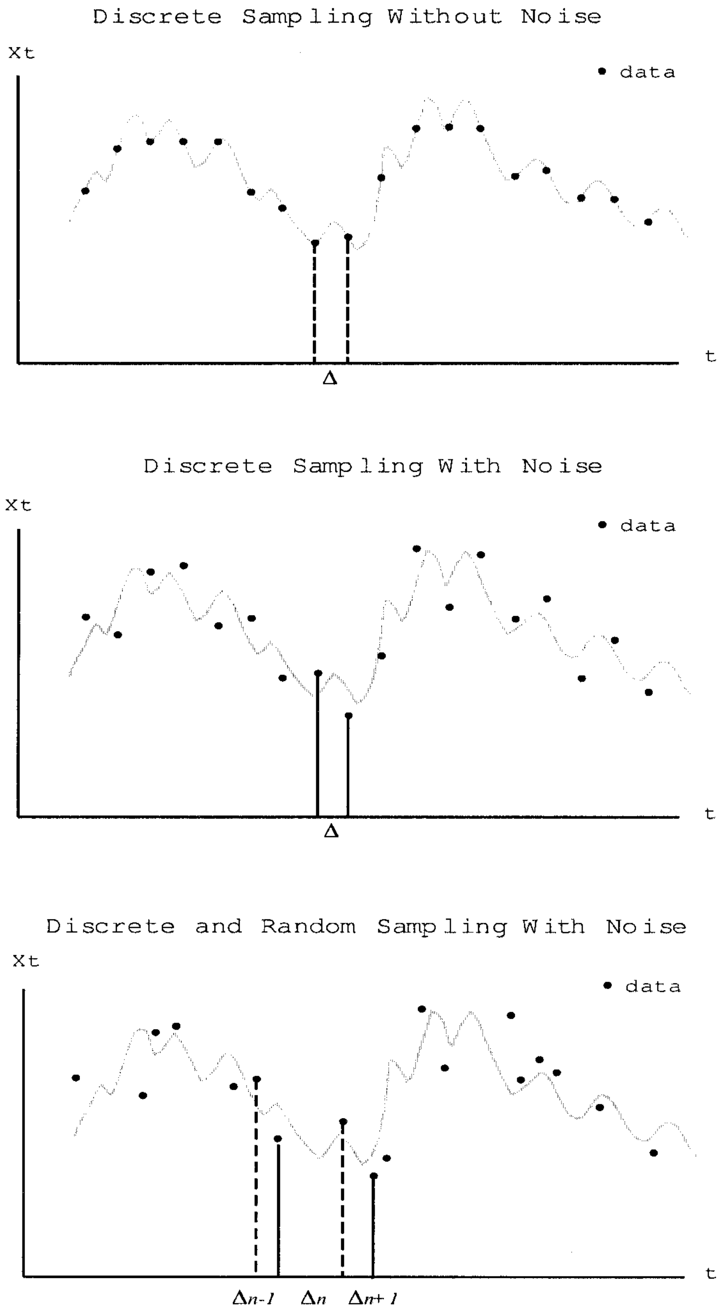

Fig. 1.1. Various discrete sampling modes – no noise (Section 1.2), with noise (Sections 1.3-1.7) and randomly spaced with noise (Section 1.8)

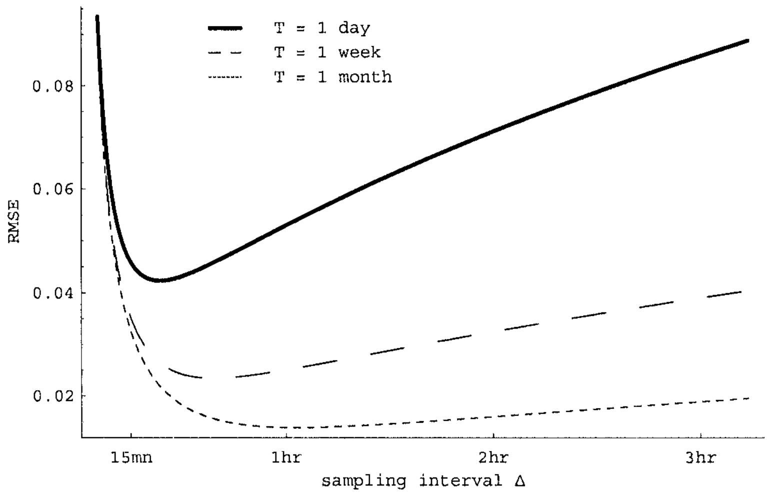

Fig. 1.2. RMSE of the estimator $\hat{\sigma}^2$ when the presence of the noise is ignored

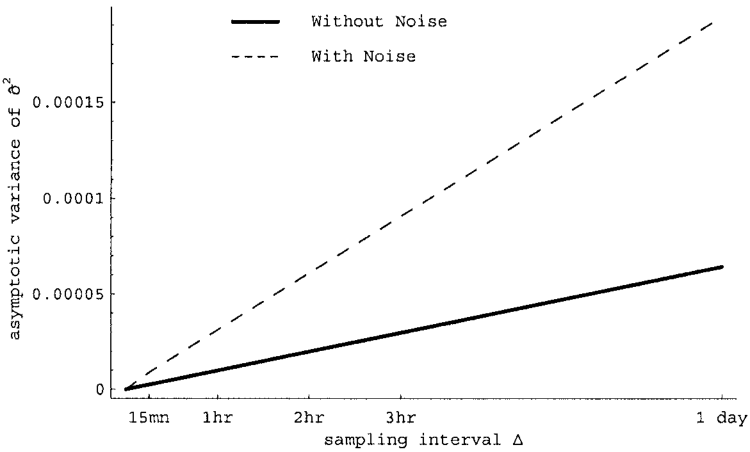

Fig. 1.3. Comparison of the asymptotic variances of the MLE $\hat{\sigma}^2$ without and with noise taken into account

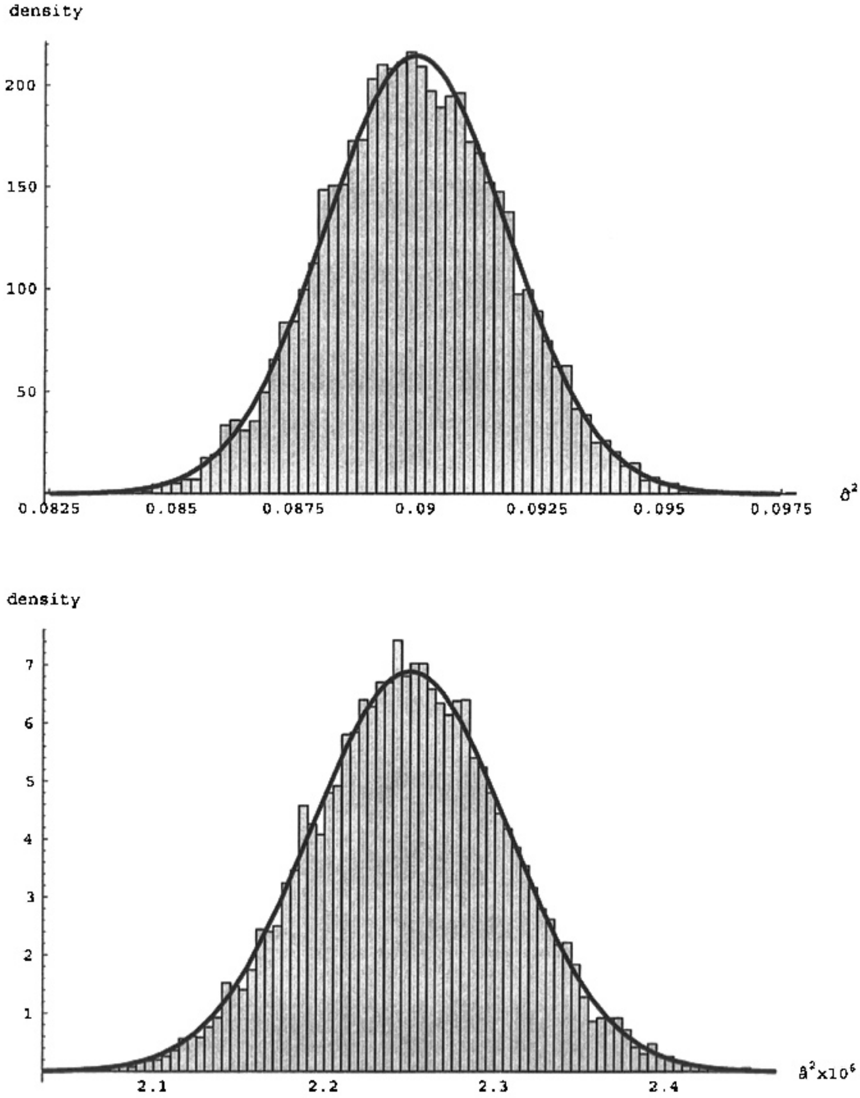

Fig. 1.4. Asymptotic and Monte Carlo distributions of the MLE $(\hat{\sigma}^2, \hat{a}^2)$ with Gaussian microstructure noise

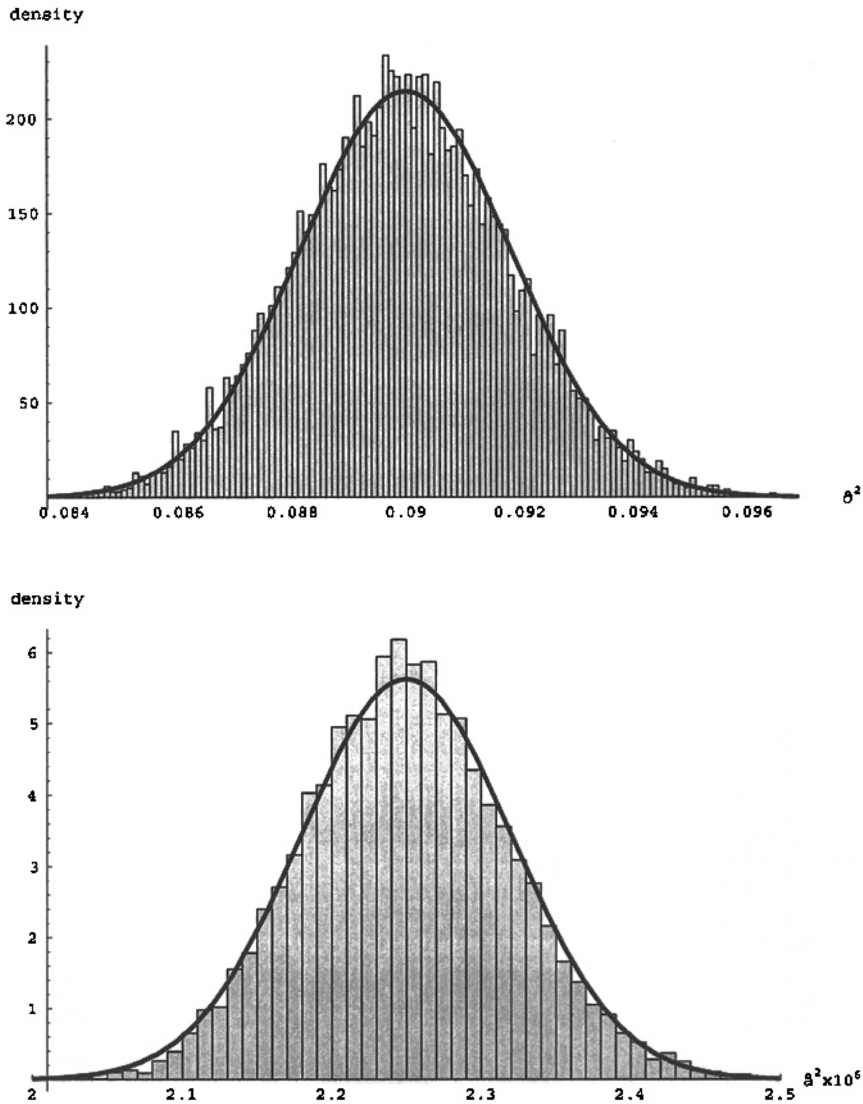

Fig. 1.5. Asymptotic and Monte Carlo distributions of the QMLE $(\hat{\sigma}^2, \hat{a}^2)$ with misspecified microstructure noise

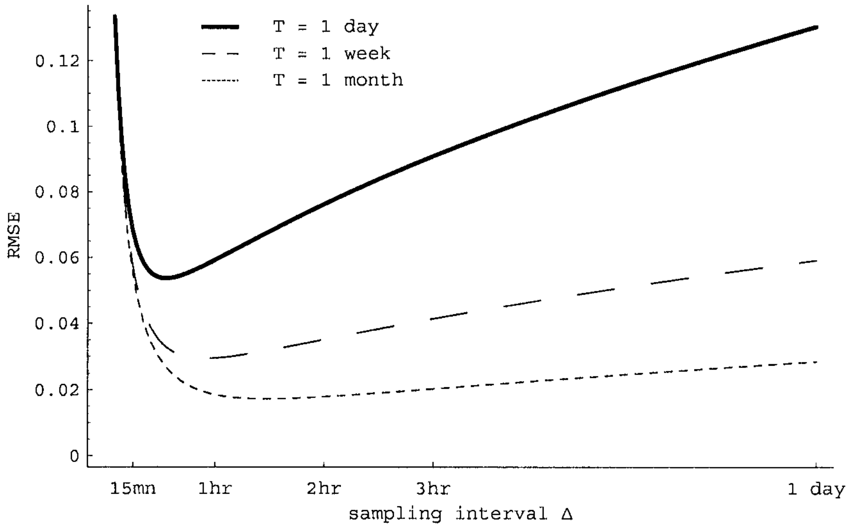

Fig. 1.6. RMSE of the estimator $\hat{\sigma}^2$ when the presence of serially correlated noise is ignored

References

[1] Abramowitz, M., and I. A. Stegun, 1972, *Handbook of Mathematical Functions.* Dover, New York, NY.

[2] Aït-Sahalia, Y., 1996, "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527–560.

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

[4] Aït-Sahalia, Y., and P. A. Mykland, 2003, "The Effects of Random and Discrete Sampling When Estimating Continuous-Time Diffusions," *Econometrica*, 71, 483–549.

[5] Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2001, "The Distribution of Exchange Rate Realized Volatility," *Journal of the American Statistical Association*, 96, 42–55.

[6] Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2003, "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

[7] Bandi, F. M., and P. C. B. Phillips, 2003, "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica*, 71, 241–283.

[8] Barndorff-Nielsen, O. E., and N. Shephard, 2002, "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, B*, 64, 253–280.

[9] Bessembinder, H., 1994, "Bid-Ask Spreads in the Interbank Foreign Exchange Markets," *Journal of Financial Economics*, 35, 317–348.

[10] Black, F., 1986, "Noise," *Journal of Finance*, 41, 529–543.

[11] Choi, J. Y., D. Salandro, and K. Shastri, 1988, "On the Estimation of Bid-Ask Spreads: Theory and Evidence," *The Journal of Financial and Quantitative Analysis*, 23, 219–230.

[12] Conrad, J., G. Kaul, and M. Nimalendran, 1991, "Components of Short-Horizon Individual Security Returns," *Journal of Financial Economics*, 29, 365–384.

[13] Dahlquist, G., and A. Björck, 1974, *Numerical Methods.* Prentice-Hall Series in Automatic Computation, New York.

[14] 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.

[15] Durbin, J., 1959, "Efficient Estimation of Parameters in Moving-Average Models," *Biometrika*, 46, 306–316.

[16] Easley, D., and M. O'Hara, 1992, "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47, 577–605.

[17] Engle, R. F., and J. R. Russell, 1998, "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

[18] French, K., and R. Roll, 1986, "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics*, 17, 5–26.

[19] Gençay, R., G. Ballocchi, M. Dacorogna, R. Olsen, and O. Pictet, 2002, "Real-Time Trading Models and the Statistical Properties of Foreign Exchange Rates," *International Economic Review*, 43, 463–491.

[20] Glosten, L. R., 1987, "Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices," *Journal of Finance*, 42, 1293–1307.

[21] Glosten, L. R., and L. E. Harris, 1988, "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economics*, 21, 123–142.
[22] Gloter, A., and J. Jacod, 2000, "Diffusions with Measurement Errors: I - Local Asymptotic Normality and II - Optimal Estimators," working paper, Université de Paris VI.
[23] Gottlieb, G., and A. Kalay, 1985, "Implications of the Discreteness of Observed Stock Prices," *Journal of Finance*, 40, 135–153.
[24] Haddad, J., 1995, "On the Closed Form of the Likelihood Function of the First Order Moving Average Model," *Biometrika*, 82, 232–234.
[25] Hamilton, J. D., 1995, *Time Series Analysis*. Princeton University Press, Princeton.
[26] Hansen, L. P., and J. A. Scheinkman, 1995, "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767–804.
[27] Harris, L., 1990a, "Estimation of Stock Price Variances and Serial Covariances from Discrete Observations," *Journal of Financial and Quantitative Analysis*, 25, 291–306.
[28] Harris, L., 1990b, "Statistical Properties of the Roll Serial Covariance Bid/Ask Spread Estimator," *Journal of Finance*, 45, 579–590.
[29] Hasbrouck, J., 1993, "Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement," *Review of Financial Studies*, 6, 191–212.
[30] Heyde, C. C., 1997, *Quasi-Likelihood and Its Application*. Springer-Verlag, New York.
[31] Jacod, J., 1996, "La Variation Quadratique du Brownien en Présence d'Erreurs d'Arrondi," *Astérisque*, 236, 155–162.
[32] Kaul, G., and M. Nimalendran, 1990, "Price Reversals: Bid-Ask Errors or Market Overreaction," *Journal of Financial Economics*, 28, 67–93.
[33] Lo, A. W., and A. C. MacKinlay, 1990, "An Econometric Analysis of Nonsynchronous Trading," *Journal of Econometrics*, 45, 181–211.
[34] 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.
[35] Madhavan, A., M. Richardson, and M. Roomans, 1997, "Why Do Security Prices Change?," *Review of Financial Studies*, 10, 1035–1064.
[36] McCullagh, P., 1987, *Tensor Methods in Statistics*. Chapman and Hall, London, U.K.
[37] McCullagh, P., and J. Nelder, 1989, *Generalized Linear Models*, second edn., Chapman and Hall, London, U.K.
[38] Merton, R. C., 1980, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.
[39] Roll, R., 1984, "A Simple Model of the Implicit Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39, 1127–1139.
[40] Shaman, P., 1969, "On the Inverse of the Covariance Matrix of a First Order Moving Average," *Biometrika*, 56, 595–600.
[41] Sias, R. W., and L. T. Starks, 1997, "Return Autocorrelation and Institutional Investors," *Journal of Financial Economics*, 46, 103–131.
[42] White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

[43] Zhang, L., P. A. Mykland, and Y. Aït-Sahalia, 2002, "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *forthcoming in Journal of the American Statistical Association.*