The Effect of Microstructure Noise on the Realized Variance

1 The Efficient Price

We use the Gordon growth model[1] to get an idea of the efficient or correct stock price valuation. Let E denote expected earnings for next year, g the expected growth rate of earnings, and ρ the appropriate discount rate relative to the risk of the stock. Then we have the valuation expression for the correct price P as

$$P = \sum_{j=0}^{\infty} \frac{(1+g)^j E}{(1+\rho)^{j+1}}.$$

or

$$P = \frac{E}{\rho - g}.$$

After using the formula for the geometric sum. Evidently $\rho > g$ as to be expected for the stock to be worth a finite amount of money. As time t passes over seconds or minutes, the values of E, ρ, g get revised continuously, so we put the t subscript on the variables,

$$P_t = \frac{E_t}{\rho_t - g_t}.$$

No financial market could be designed to keep the traded price exactly equal to P_t continuously. We introduce a small multiplicative observation error,

$$P_t^{\text{Observed}} = (\text{Measurement Error}_t) \times P_t$$

Then, to be consistent with previous lectures, we take $X_t = \log(P_t)$, so that

$$\log(P_t^{\text{Observed}}) = X_t + \text{Noise}_t,$$

which makes the noise additive in logs and the analysis tractable.

2 Sampling and Observation

We consider the efficient X evolving in continuous time as

$$dX_t = \sqrt{c_t} dW_t + dJ_t.$$

[1] All modern asset valuation models are variants of the Gordon model.

For a while we restrict $t \in [0,1]$ but everything below holds day-by-day if $t \in [0,T]$. As before,

$$\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$$

We allow for the possibility of measurement error (noise) by way of

$$Y_i^n = X_{i\Delta_n} + \chi_i$$

where χ_i reflects the noise, taken to have mean zero, variance σ_χ^2, and to be serially uncorrelated. If $\sigma_\chi^2 = 0$ then we are back in the case studied before, while if $\sigma_\chi^2 > 0$ we are in the noisy case. Keep in mind that the Δ_i^n notation automatically adjusts the gap between observations on X when n varies.

Now we consider sums of squares of the noisy data:

$$\sum_{i=1}^{n}(Y_i^n - Y_{i-1}^n)^2.$$

Write this out as

$$\begin{aligned}\sum_{i=1}^{n}(Y_i^n - Y_{i-1}^n)^2 &= \sum_{i=1}^{n}(\Delta_i^n X)^2 + \chi_i^2 + \chi_{i-1}^2 + \text{cross products}\\ &\approx \sum_{i=1}^{n}(\Delta_i^n X)^2 + \chi_i^2 + \chi_{i-1}^2\end{aligned}$$

where the cross product terms are like $\Delta_i^n X \times \chi_i, \ldots$, with completely random signs $(^{\pm})$ and magnitudes, so the cross product terms will largely cancel out in the sum. The text Ait-Sahalia and Jacod, 2014, p. 216 states the same thing with slightly more complicated notation. The first term in the above acts like

$$\sum_{i=1}^{n}(\Delta_i^n X)^2 \approx IV = \int_0^1 c_s ds,$$

regardless of n so long as n is reasonably large; remember that for sums of squared differences of X, it does not matter much if we use 10-second, 5-minute, 6-minute, or 10-minute sampling, we should get approximately the same number. On other hand by the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n}\chi_i^2 + \chi_{i-1}^2 \approx 2\sigma_\chi^2 \quad \Rightarrow \quad \sum_{i=1}^{n}\chi_i^2 + \chi_{i-1}^2 \approx 2n\sigma_\chi^2$$

where on the right-hand side the contribution of the noise grows directly with n. Considering consecutively using 10-minute ($n = 38$), 5-minute ($n = 77$), 1-minute ($n = 385$), 1-second ($n = 23,100$), so

$$\begin{aligned}\sum_{i=1}^{n}(Y_i^n - Y_{i-1}^n)^2 &\approx IV + 38(2\sigma_\chi^2)\\ &\approx IV + 77(2\sigma_\chi^2)\\ &\approx IV + 385(2\sigma_\chi^2)\\ &\approx IV + 23100(2\sigma_\chi^2)\\ &\approx IV + n(2\sigma_\chi^2)\end{aligned}$$

The noise term explodes with n; this explosion actually happens in the data. The figure in the text Ait-Sahalia and Jacod, 2014, p. 217 reveals this sort of behavior of the sum of squared observed returns. This explosion with n is the basic reason we do not regularly drill down to the very finest time interval recorded by the exchange.

3 Volatility Signature Plot

A key to understanding the effects of the noise is the volatility signature plot, which is a graphical representation of certain anomalies related to price variation at very high frequencies. Under the model conditions, so long as n is large enough that we think the asymptotic approximations are accurate, then the realized variance should be independent of the sampling frequency. That is

$$RV_{\text{5-min}} \approx RV_{\text{3-min}} \approx RV_{\text{1-min}} \approx RV_{\text{30-second}} \approx RV_{\text{10-second}} \approx RV_{\text{1-second}}$$

The reason is that each RV estimates the same thing, IV, and so each RV should be about the same.

Another way to view things is to remember that

$$IV = \int_0^1 c_s \, ds,$$

and no matter how we slice and dice IV the sum of the pieces will add back up to IV. Consider $n \geq 1$, Δ_n, and the intervals $I_1 = [0, \Delta_n = 1/n]$, $I_2 = (\Delta_n, 2\Delta_n]$, $\dots, I_n = ((n-1)\Delta_n, 1]$, then always, i.e., for all $n \geq 1$:

$$IV = \int_{I_1} c_s \, ds + \int_{I_2} c_s \, ds + \cdots + \int_{I_n} c_s \, ds.$$

By definition the realized variance is

$$\begin{aligned} RV &= (\Delta_1^n X)^2 + (\Delta_2^n X)^2 + \cdots + (\Delta_n^n X)^2 \\ &= \sum_{i=1}^{n} (\Delta_i^n X)^2 \end{aligned}$$

We showed (sketched the proof) early in the course that sum above approximates IV for large n.

What happens, however, if you drill down to ultra-high frequency data? Using data based on 5-second sampling we can compute the volatility signature plot for 5-sec, 10-sec, 15-sec, etc. In project 2 you did just that. At sampling frequencies higher than 1-minute, the effects of the noise are quite evident. Raw data sampled more frequently without noise corrections are unusable for regular economic purposes. However, ultra-high frequency data are of intense interest to economists who study the economics of the trading frictions that lead to the noise; study of these trading frictions are not a topic for this course but are important in other contexts.