Microstructure Noise

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 \(\rho\) 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 worth a finite amount of money. As time \(t\) passes over seconds or minutes, the values of \(E, \rho, g\) 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 \(X_t = 100 \log(P_t)\), so that

\[
100 \log(P_{t,\text{Observed}}) = X_t + \text{Noise}_t,
\]

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

2 Sampling and Observation

For simplicity we consider only days without jumps where the efficient \(X\) evolves in continuous time as

\[
dX_t = \sqrt{c_t} dW_t. \tag{1}
\]

\(^1\)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,

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

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

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

where $\chi_i$ reflects the noise, taken to have mean zero, variance $\sigma^2_\chi$, and be serially uncorrelated. If $\sigma^2_\chi = 0$ then we are back in the case studied before, while if $\sigma^2_\chi > 0$ we are in the noisy case. Keep in mind that the $\Delta_t^n$ notation automatically widens 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

$$\sum_{i=1}^n (Y_i^n - Y_{i-1}^n)^2 = \sum_{i=1}^n (\Delta_t^n X)^2 + \chi_i^2 + \chi_{i-1}^2 + \text{cross products}$$

$$\approx \sum_{i=1}^n (\Delta_t^n X)^2 + \chi_i^2 + \chi_{i-1}^2$$

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

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

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

$$\sum_{i=1}^n \chi_i^2 + \chi_{i-1}^2 \approx 2n\sigma^2_\chi$$

while 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

$$\sum_{i=1}^n (Y_i^n - Y_{i-1}^n)^2 \approx IV + 38(2\sigma^2_\chi)$$

$$\approx IV + 77(2\sigma^2_\chi)$$

$$\approx IV + 385(2\sigma^2_\chi)$$

$$\approx IV + 23100(2\sigma^2_\chi)$$

2
While \( \sigma^2_\chi \) should be fairly small, the noise term explodes with \( n \); this explosion actually happens in the data. The figure in (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 A glimpse at \( \sigma^2_\chi \) for the Noise

Now we work through some details. Suppose we choose consecutively lower sampling frequencies \( n_1, n_2, \ldots, n_J \); for example, if we think of sampling at 1, 2, 3, 4, 5 minutes, then \( n_1 = 385, n_5 = 77 \), as special cases of \( n_j = \lfloor 385/j \rfloor \). Define

\[
RV_j = \sum_{i=1}^{n_j} (Y_i^{n_j} - Y_{i-1}^{n_j})^2
\]

Then from the above,

\[
RV_j \approx IV + 2\sigma^2_\chi n_j
\]

so at different scales we get different noisy measurements of the same \( IV \), which we can difference to generate different estimates of \( \sigma^2_\chi \). The different estimates could be averaged to derive a common estimate of \( \sigma^2_\chi \). This is left as an exercise.

### 4 Induced Autocorrelation

Again we consider for simplicity only days without noise. The efficient \( X \) evolving in continuous time as

\[
dX_t = \sqrt{c_t}dW_t
\]

While we restrict \( t \in [0,1] \) but everything below holds day-by-day if \( t \in [0,T] \). As before,

\[
\Delta^n_i X = X_{i\Delta_n} - X_{(i-1)\Delta_n}
\]

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

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

where \( \chi_i \) reflects the noise, taken to have mean zero, variance \( \sigma^2_\chi \), and be serially uncorrelated. If \( \sigma^2_\chi = 0 \) then we are back in the case studied before, while if \( \sigma^2_\chi > 0 \) we are in the noisy case. Keep in mind that the \( \Delta^n_i \) notation automatically widens the gap between observations on \( X \) when \( n \) varies. For now, assume \( X \) is continuous so \( \text{Jumps} = 0 \).

The noise will create an artificial negative correlation (false predictability) in the price moves:

\[
\Delta^n_i Y = \Delta^n_i X + \chi_i - \chi_{i-1}
\]

\[
\Delta^n_{i+1} Y = \Delta^n_{i+1} X + \chi_{i+1} - \chi_i
\]
Suppose $c_t = c$, a constant. Then

\[
\begin{align*}
\text{Var}(\Delta^n_i Y) &= \Delta_n c + 2\sigma^2_X \\
\text{Var}(\Delta^n_{i+1} Y) &= \Delta_n c + 2\sigma^2_X \\
\text{Cov}(\Delta^n_i Y, \Delta^n_{i+1} Y) &= -\sigma^2_X \\
\text{Corr}(\Delta^n_i Y, \Delta^n_{i+1} Y) &= \rho
\end{align*}
\]

\[\rho = \frac{-\sigma^2_X}{\Delta_n c + 2\sigma^2_X}\]

For example, with 1-min data on BAC (Bank of America), we have that on average

\[\rho = -0.115\]

which is fairly high.

5 Pre-Averaging

We can attenuate the serial correlation by averaging the returns over short intervals of length $k_n^2$. Think of pre-averaging one-minute returns:

\[r_{i}^{\text{pre-ave}} = k_n \sum_{j=0}^{k_n-1} w_j \Delta^n_{(i-1)k_n+j} Y, \quad i = 1, 2, \ldots, n, \quad n = \lfloor 365/k_n \rfloor.\]

Forming 5-minute returns is a type of pre-averaging where $w_0 = w_1 = \cdots = w_4 = 1$:

\[r_{1}^{\text{pre-ave}} = \Delta^n_1 Y + \cdots + \Delta^n_5 Y,\]
\[r_{2}^{\text{pre-ave}} = \Delta^n_6 Y + \cdots + \Delta^n_{10} Y,\]
\[\vdots\]
\[r_{77}^{\text{pre-ave}} = \Delta^n_{381} Y + \cdots + \Delta^n_{385} Y\]
\[r_{1}^{\text{pre-ave}} = \Delta^n_1 X + \cdots + \Delta^n_5 X + \chi_5 - \chi_0,\]
\[r_{2}^{\text{pre-ave}} = \Delta^n_6 X + \cdots + \Delta^n_{10} X + \chi_{10} - \chi_5\]

As for the artificial correlation, observe that if $c_t = c$ a constant, then

\[
\begin{align*}
\text{Var}(r_1^{\text{pre-ave}}) &= 5\Delta_n c + 2\sigma^2_X \\
\text{Var}(r_2^{\text{pre-ave}}) &= 5\Delta_n c + 2\sigma^2_X \\
\text{Cov}(r_1^{\text{pre-ave}}, r_2^{\text{pre-ave}}) &= -\sigma^2_X \\
\text{Corr}(r_1^{\text{pre-ave}}, r_2^{\text{pre-ave}}) &= \rho
\end{align*}
\]

\[\rho = \frac{-\sigma^2_X}{5\Delta_n c + 2\sigma^2_X}\]

\[\text{Note that } k_n \text{ here defines a different window that used for local variance estimation.}\]

4
We can see this with 5-min data on BAC (Bank of America), we have that on average
\[ \rho = -0.072 \]
which is lower than on 1-minute data because the noise is begin averaged out.

In general we might find better weights but to preserve variance we want
\[ \frac{1}{k_n} \sum_{j=0}^{k_n-1} w_j^2 = 1, \text{ or } \sum_{j=0}^{k_n-1} w_j^2 = k_n. \]
The above is exactly what we did for the within-day volatility pattern.

Another way to pre-average 1-min to 5-min is to select weights
\[ w = (1 2 3 2 1)/S, \quad S = \sqrt{(1^2 + 2^2 + 3^2 + 2^2 + 1^2)/5} \]
With such weighting the BAC correlation becomes
\[ \rho = -0.056 \]
slightly weaker. After pre-averaging, compute BV, \( \tau_i \), and truncate for jumps.

References