Negative Autocorrelation in *Micro* MidPrice Differences

Guanlin Li

High-frequency financial data, particularly the differences (returns) of the Midprice, consistently exhibit significant negative first-order autocorrelation. *Why?*

1 Differenced White Noise Correlation

Let $\{U_t\}$ be a pure white noise process with:

$$\mathbb{E}[U_t] = 0, \quad \mathrm{Var}(U_t) = \sigma^2 > 0, \quad \mathrm{Cov}(U_t, U_{t-k}) = 0 \quad \forall k \neq 0. \tag{1}$$

The first-order difference of the white noise process is:

$$Z_t = U_t - U_{t-1}. \tag{2}$$

We compute the covariance between Z_t and its first lag Z_{t-1}:

$$\begin{aligned} \mathrm{Cov}(Y_t, Y_{t-1}) &= \mathrm{Cov}(U_t - U_{t-1},\ U_{t-1} - U_{t-2}) \\ &= -\mathrm{Var}(U_{t-1}) = -\sigma^2. \end{aligned} \tag{3}$$

All cross-terms vanish due to the white noise properties. The variance of Z_t is:

$$\mathrm{Var}(Z_t) = \mathrm{Var}(U_t) + \mathrm{Var}(U_{t-1}) = 2\sigma^2. \tag{4}$$

Therefore, the first-order autocorrelation coefficient is:

$$\rho_Z(1) = \frac{\mathrm{Cov}(Z_t, Z_{t-1})}{\sqrt{\mathrm{Var}(Z_t)\mathrm{Var}(Z_{t-1})}} = \frac{-\sigma^2}{2\sigma^2} = -\frac{1}{2}. \tag{5}$$

Any white noise series, once differenced, possesses a fixed theoretical first-order autocorrelation of exactly $\mathbf{-1/2}$. This is an inherent "saw-tooth" effect introduced by the differencing operator.

2 The Price Model with Microstructure Noise

We now embed the previous result into a standard financial price process, following the reduced-form framework commonly used in market microstructure literature [ASMZ05].

2.1 Observed Price vs. Efficient Price

Let X_t denote the unobservable "efficient" log-price, assumed to follow a pure random walk (drift is negligible at high frequencies):

$$dX_t = \sigma dW_t, \tag{6}$$

where σ is the volatility and W_t is standard Brownian motion. Let U_t represent the market microstructure noise (i.i.d., mean zero, variance a^2), arising from bid-ask spreads, price discreteness (tick size), and rounding errors. The observed log-price (e.g., the log of the Midprice) is:

$$\widetilde{X}_t = X_t + U_t. \tag{7}$$

Let the sampling interval be Δ. The observed high-frequency return is:

$$Y_i = \widetilde{X}_i - \widetilde{X}_{i-1} = (X_i - X_{i-1}) + (U_i - U_{i-1}). \tag{8}$$

The true price increment has variance $\sigma^2\Delta$, while the noise component is the difference of two i.i.d. noise terms. Since the efficient returns are uncorrelated across time, the autocovariance of Y_i is solely determined by the overlapping noise term:

$$\begin{aligned} \mathrm{Cov}(Y_i, Y_{i-1}) &= \mathrm{Cov}(U_i - U_{i-1},\ U_{i-1} - U_{i-2}) \\ &= -\mathrm{Var}(U_{i-1}) = -a^2. \end{aligned} \tag{9}$$

2.2 Total Variance of Observed Returns

The variance of the observed returns is the sum of the efficient price variance and the noise variance:

$$\mathrm{Var}(Y_i) = \sigma^2\Delta + 2a^2. \tag{10}$$

Combining Eq. 9 and Eq. 10, the theoretical first-order autocorrelation of the observed returns is:

$$\rho_Y(1) = \frac{\mathrm{Cov}(Y_i, Y_{i-1})}{\mathrm{Var}(Y_i)} = \frac{-a^2}{\sigma^2\Delta + 2a^2}. \tag{11}$$

Define the proportion of the total return variance attributable to microstructure noise as:

$$\pi(\Delta) = \frac{2a^2}{\sigma^2\Delta + 2a^2}. \tag{12}$$

As we sample at increasingly higher frequencies (e.g., tick-by-tick or millisecond data), the sampling interval $\Delta \to 0$. Consequently, $\lim_{\Delta\to 0}\pi(\Delta) = 1$. This implies that in the ultra-high-frequency limit, nearly 100% of the observed return variance is driven by microstructure noise, completely swamping the true price signal. Substituting $\Delta \to 0$ into Eq. 11 yields the asymptotic lower bound for the autocorrelation: $\lim_{\Delta\to 0}\rho_Y(1) = -1/2$.

2.3 Time-Scaling Properties

The dominance of microstructure noise at high frequencies is rooted in a fundamental difference in how the two components behave as the sampling interval Δ varies. The efficient price process X_t evolves continuously with variance proportional to the observation horizon. Specifically, for a return computed over an interval of length Δ:

$$\mathrm{Var}(X_t - X_{t-\Delta}) = \sigma^2\Delta. \tag{13}$$

This variance shrinks linearly with Δ. Hence, as we sample more frequently ($\Delta \to 0$), the magnitude of the true price movement over that interval tends to zero. By contrast, the microstructure noise U_t is a purely observational error attached to each recorded price. It does *not* accumulate or average out over shorter time spans. The variance contributed by the noise to a single observed return is:

$$\mathrm{Var}(U_t - U_{t-\Delta}) = 2a^2, \tag{14}$$

which is **independent of** Δ. It remains constant regardless of whether we sample every second or every minute. This asymmetric scaling property is the ultimate driver of the noise-to-signal explosion. As $\Delta \to 0$:

$$\frac{\text{Noise Variance}}{\text{True Price Variance}} = \frac{2a^2}{\sigma^2\Delta} \longrightarrow \infty. \tag{15}$$

This insight is central to the literature [ASMZ05] on realized volatility estimation. Sampling every millisecond does not yield more information about σ; it yields more information about a.

References

[ASMZ05] Yacine Ait-Sahalia, Per A Mykland, and Lan Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *The review of financial studies*, 18(2):351–416, 2005.