Optimal estimation of intraday 'true value' using the Glosten-Milgrom model

Isaac Drachman

We wish to statistically model the 'true value' of a traded asset over a time period solely by utilizing its order flow. The primary object of study is therefore μ_t, the true value after the t-th trade. We say that μ_t follows a distribution F_t. We let our best estimate be $\hat{\mu}_t = \mathbb{E}F_t$.

We consider the data (m_0, X), where m_0 is the initial midpoint price and X is a sequence of trades x_1, x_2, ..., x_T. The t-th trade x_t is the pair (d_t, p_t), where $d_t = +1$ or -1 for a market buy or a market sell, respectively, and p_t is price at which the trade occurred. We specify the first prior distribution F_0 as Normal with mean m_0 and std. deviation σ, which will be a parameter of our model. Our first best estimate is $\hat{\mu}_0 = m_0$.

From the Glosten-Milgrom microstructure model, we say that all traders are either perfectly informed (who know μ_t) or uninformed (noisy trading) with the following specifications (where b_t, a_t are bid/ask after trade t)

$$\begin{aligned} P(\text{trade made by informed trader}) &= \alpha \\ P(\mu_t > a_t \,|\, \text{informed bought}) &= 1 \\ P(\mu_t < b_t \,|\, \text{informed sold}) &= 1 \\ P(\text{uninformed buys when trades}) &= \eta \end{aligned}$$

After witnessing trade x_t we update F_t to a posterior distribution F_{t+1} using this[1]. We explicitly modify the probability distribution function for a trade with $d_t = +1$ (i.e. market buy)

$$f_{t+1}(p) = f_t(p) \times \begin{cases} (1-\alpha)\eta + \alpha & p > a_t \\ 1 - \alpha - (1-\alpha)\eta & p \leq a_t \end{cases}$$

After witnessing a trade with $d_t = -1$ (i.e. market sell)

$$f_{t+1}(p) = f_t(p) \times \begin{cases} (1-\alpha)\eta + \alpha & p < b_t \\ 1 - \alpha - (1-\alpha)\eta & p \geq b_t \end{cases}$$

We set a_t, $b_t = p_t$, i.e. we have that the price at which the trade occurred was on the bid/ask. These probability distributions then need to be normalized.

When coding this model we create the sequence of vectors $\{\mathbf{p}^t\}_{t\in\{0,1,\ldots,T\}}$ such that each $\mathbf{p}^t$ has $\lfloor 800\sigma \rfloor$ components with

$$p_i^t = P(\mu_t = m_0 - 4\sigma + i/100) = f_t(m_0 - 4\sigma + i/100)$$

This allows a discretization of μ_t and its distribution F_t into the minimum increment (tick) of one cent. We denote the price $m_0 - 4\sigma + i/100$ as v_i for convenience. The expectation $\hat{\mu}_t$ can then be computed by

$$\begin{aligned}\hat{\mu}_t = \mathbb{E}F_t &= \int_{-\infty}^{\infty} p\, f_t(p)\ \mathrm{d}p \\ &\approx \mathbf{v}\cdot\mathbf{p}^t = \sum_{i=0}^{\lfloor 800\sigma \rfloor} v_i\, f_t(v_i)\end{aligned}$$