Market MicroStructure Models

Jonathan Kinlay

Summary

This note summarizes some of the key research in the field of market microstructure and considers some of the models proposed by the researchers. Many of the ideas presented here have become widely adopted by high frequency trading firms and incorporated into their trading systems.

May 2011

Research Papers

Measuring the Information of Stock Trades
Joel Hasbrouck, Quantitative Finance, 1991, Journal of Finance, 46, 179-207

This paper suggests that the interactions of security trades and quote revisions be modeled as a vector autoregressive system. Within this framework, a trade's information effect may be meaningfully measured as the ultimate price impact of the trade innovation. Estimates for a sample of NYSE issues suggest: a trade's full price impact arrives only with a protracted lag; the impact is a positive and concave function of the trade size; large trades cause the spread to widen; trades occurring in the face of wide spreads have larger price impacts; and, information asymmetries are more significant for smaller firms.

The primary statistical technique employed here is vector autoregression applied to quote and trade data. While use of nonlinear functions of the variables permits a rich characterization of their dynamics, the estimation technique is essentially least squares. The basic computations are therefore linear and relatively tractable, important considerations when faced with systems characterized by large numbers of parameters and observations.

It is assumed that the efficient stock price m_t, i.e. the stock price conditioned on all publicly available information, evolves as follows:

$$m_t = m_{t-1} + \mathrm{z}\mathrm{v}_{2,t} + v_{1,t}$$

Here $v_{1,t}$and $v_{2,t}$are mutually and serially uncorrelated disturbance terms. We may regard the former as arising from public nontrade information, while the latter is the innovation in trades. The z coefficient reflects the private information conveyed by this innovation.

The second part of the model assumes that the quoted midprice q_t evolves according to the relationship:

$$q_t = m_{t-1} + a\,(q_{t-1} - m_{t-1}) + b x_t$$

where x_t is the signed trade at time t and $0 < a \leq 1$ and $b > 0$.

There are a couple of different ways to think about this second equation. We can regard a as a parameter measuring the speed of mean reversion of the quoted mid-price to the efficient price and the parameter b as a measure of momentum in the quoted midprice. But we can also think in terms of an inventory model, in which a market maker will seek to manage his inventory by adjusting his quoted prices, imperfectly if the case where a < 1.

Hasbrouck shows that this system of equations may be transformed into a Vector Autorgression (VAR) relationship between the midprice changes $r_t = q_t - q_{t-1}$ and the signed trades x_t. His analysis leads to the following conclusions:

1. The full impact of a trade on the security price is not felt instantaneously but with a protracted lag.
2. As a function of trade innovation size, the ultimate impact of the innovation on the quote is nonlinear, positive, and increasing, but concave.
3. The order flow is affected by prior quote revisions: there is Granger-Sims causality running from quote revisions to trades.
4. Spread size exhibits a response to trading activity. large trades in particular are associated with a widening of the spread. This is consistent with a market maker widening the spread after a large trade based on an increased likelihood that an information event has occurred.

The Next Tick on Nasdaq
Bruce Mizrach, Quantitative Finance, 2008, vol. 8, issue 1, pages 19-40

The Nasdaq stock market provides information about buying and selling interest in its limit order book. Using a vector autoregressive model of trades and returns, the author assesses the effect of the entire order book on the next tick. The author also determines the influence of individual market makers and electronic networks and find evidence that the identity of market participants can be useful information. Finally, the author produces a set of dynamic market price responses to buy and sell orders, and finds that these estimates vary with standard measures of liquidity.

Mizrach starts with the bivariate Vector Autoregression (VAR) model of intraday quote and trade evolution introduced by Hasbrouck (1991). His quote revision model is specified as:

$$r_t = a_{r,0} + \sum_{i=1}^{5} a_{r,i}\, r_{t-i} + \sum_{i=0}^{15} b_{r,i}\, x_{t-i}^0 + \epsilon_{r,t}$$

where r_t is the mid-price return and x_t^0 denotes the net sum of the sequence of signed trades since the last tick (a trade is considered a buy (sell) and is signed +1 (-1) if it is above (below) the midquote). As in Hasbrouck, the bivariate VAR includes a trade process with specification

$$x_t^0 = a_{x,0} + \sum_{i=1}^{5} a_{x,i}\, r_{t-i} + \sum_{i=0}^{15} b_{x,i}\, x_{t-i}^0 + \epsilon_{x,t}$$

The author estimates this model for First Health, noting the high level of autocorrelation in midprice returns and the fact that transactions are positively autocorrelated and highly predictable, with an R^2 of 0.74. He goes on to show how estimates of market impact can be derived from the model.

Having set the framework for the VAR model, Mizrach applies it for two groups of stocks, the Nasdaq 100 and a sample of 250 small cap stocks. He finds that for the Nasdaq sample there is positive feedback on quote revisions, even in the absence of trades. The sample of small cap stocks, on the other hand, have autoregressive midpoint quote dynamics that are more like the order driven NYSE sample in Hasbrouck, with negative first order autocorrelations in most cases.

Mizrach next extends the VAR model to include information beyond the inside quote and makes a number of very interesting observations:

- The most significant explanatory variable is the ratio of the number of bidders to offers on the first tier and is almost always more important than quoted depth.
- The entire demand curve impacts the next tick and is surprisingly persistent: only after 13 ticks does the order book's impact on the next tick become insignificant.
- The inside quote volume is highly significant in most cases.

- The Ax plays a statistically significant role in both the bid and ask on most stocks.
- Quoted depths away from the first tier are only occasionally significant.

The positive serial correlation in quote revisions for large cap stocks is consistent with the idea that uniformed market makers follow the quotes of informed market makers. The market also appears to recognize that some dealers and ECNs are more informed than others and is likely to adjust their quotes in line with only some of the most active quote providers. Finally, the author's analysis indicates that market impact studies need to take into account the entire state of the order book, not just the inside quote.

The Information Content of an Open Limit-Order Book
Charles Cao, Oliver Hansch, Xiaoxin Wang, Journal of Futures Markets Volume 29, Issue 1, pages 16-41, January 2009

Using data from the Australian Stock Exchange, the authors assess the information content of an open limit-order book with a particular focus on the incremental information contained in the limit orders behind the best bid and offer. The authors find that the order book is moderately informative—its contribution to price discovery is approximately 22%. The remaining 78% is from the best bid and offer prices on the book and the last transaction price imbalances between the demand and supply schedules along the book are significantly related to future short-term returns, even after controlling for the autocorrelations in return, the inside spread, and the trade imbalance.

The authors characterize the shape of the order book by reference to changes in the weighted prices at level 1 and from step $n_1$to n_2

$$\mathrm{WP}^1 = \frac{Q_1^d\,P_1^d + Q_1^s\,P_1^s}{Q_1^d + Q_1^s}$$

$$\mathrm{WP}^{n_1-n_2} = \frac{\sum_{j=n_1}^{n_2} Q_j^d\,P_j^d + Q_j^s\,P_j^s}{\sum_{j=n_1}^{n_2} Q_j^d + Q_j^s}$$

WP^1 will change if the height or the length of the order book changes, while $\mathrm{WP}^{n_1-n_2}$ changes is any of the steps between step $n_1$to $n_2$experience a change a change in the height, length, or both.

To address the question of whether the order book provides valuable information, the authors compare the informational content of WP^1 and WP^{1-10} . If the order book beyond the first step contains information about the value of the stock then WP^{1-10} can be expected to be a better indicator of value that WP^1. Alternatively, if the book does not contain useful information, introduces noise, or reacts sluggishly to new information, then WP^1will be more informative. The authors use an Error Correction Model (ECM) to assess the contribution of each price series to price discovery. The specification of the ECM models used by the authors is as follows:

$$\Delta\mathrm{WP}_t^1=\alpha_{1,0}\text{-}\alpha_1(\mathrm{WP}_{t-1}^1\text{-}\beta\mathrm{WP}_{t-1}^{1-10})+\sum_{i=1}^{p}\left(\gamma_{1,i}\Delta\mathrm{WP}_{t-i}^1+\delta_{1,i}\Delta\mathrm{WP}_{t-i}^{1-10}\right)+\eta_{1,t}$$

$$\Delta\mathrm{WP}_t^{1-10}=\alpha_{2,0}\text{-}\alpha_2(\mathrm{WP}_{t-1}^1\text{-}\beta\mathrm{WP}_{t-1}^{1-10})+\sum_{i=1}^{p}\left(\gamma_{2,i}\Delta\mathrm{WP}_{t-i}^1+\delta_{2,i}\Delta\mathrm{WP}_{t-i}^{1-10}\right)+\eta_{2,t}$$

Four cases are delineated:

- Case 1: $\alpha_1$is insignificant, but α_2 is significant $\Rightarrow\mathrm{WP}^1$ leads WP^{1-10} ;
- Case 2: $\alpha_1$is significant, but α_2 is insignificant $\Rightarrow\mathrm{WP}^{1-10}$ leads WP^1;
- Case 3: both $\alpha_1$and α_2 are significant $\Rightarrow$ both WP^1 and WP^{1-10} adjust downward towards long term equilibrium; and
- Case 1: neither $\alpha_1$or is significant $\Rightarrow$ neither WP^1 nor WP^{1-10} adjust downward towards long term equilibrium.

The authors find that the $\alpha_1$coefficients are significant in only a handful of cases, whereas the α_2 coefficients are significant for almost every stock. This implies that WP_t^{1-10} adjusts to WP^1 for virtually all firms, but the reverse is not true in most cases. Thus WP^1 leads WP^{1-10}in almost all cases.

However, the authors go on to show that, although the information from the first step is dominating, the order book beyond the first step is informative of the true value of the underlying stock: the contribution of the order book beyond the first step is of the order of 30%. They also find that imbalance information from steps 2 to 10 leads to an 11-18% increase in the adjusted R^2in comparison to the result using imbalance from step one only. Finally, the authors find that traders use the available information on the state of the book, including beyond the first step, in developing their order submission strategies.

The Logarithmic ACD Model: an Application to the Bid-Ask Quote Process for Three NYSE Stocks
Luc Bauwens and Pierre Giot, 2000, Annales D'Economie et de Statistique - No 6

This paper introduces the logarithmic autoregressive conditional duration (Log-ACD) model and compares it with the ACD model of ENGLE and RUSSELL [1998]. The logarithmic version allows additional variables to be introduced into the ACD model without sign restrictions on their coefficients. The authors apply the Log-ACD model to price durations relative to the bid-ask quote process of three securities listed on the New York Stock Exchange, and they investigate the influence of some characteristics of the trade process (trading intensity, average volume per trade and average spread) on the bid-ask quote process.

Let x_i be the duration between two quotes (a quote being a collection of data relative to a buy or a sell of a security on a stock exchange). The assumption introduced by ENGLE and RUSSELL [1998] is that the time dependence in the durations can be subsumed in their conditional expectations $\Psi_i = E(x_i \mid I_{i-1})$, in such a way that x_i / Ψ_i is independent and identically distributed.
I_{i-1} denotes the information set available at time *ti*-1 (*i.e.* at the beginning of duration *xi*), supposed to contain at least $\tilde{x}_{i-1}$ and $\tilde{\psi}_{i-1}$, where $\tilde{x}_{i-1}$ denotes x_i and its past values, and likewise for $\tilde{\psi}_{i-1}$.
The ACD model specifies the observed duration as a mixing process:

$$x_i = \Phi_i \epsilon_i$$

where the $\epsilon_i$are IID and follow a Weibull $(1, \gamma)$ distribution, while the Φ_i are proportional to the conditional expectation of the x_i.

A second equation specifies an autoregressive model for the (expected) conditional durations, which in the ACD(1,1) model takes the form:

$$\Psi_i = \omega + \alpha x_{i-1} + \beta \Psi_{i-1}$$

with constraints $\omega > 0$, $\beta \geqslant 0$, $\alpha \geqslant 0$ and $\alpha + \beta < 1$ to ensure the existence of the unconditional mean of durations and the positivity of conditional durations. The condition $\Psi_i = E(x_i \mid I_{i-1})$ leads to a third equation that links the first two:

$$\Psi_i = \Gamma\left(1 + \frac{1}{\gamma}\right) \Phi_i$$

The positivity constraints can be quite restrictive and can make it difficult to introduce further explanatory variables. Hence the authors consider a logarithmic version of the model in the form

$$x_i = e^{\Phi_i} \epsilon_i$$

with log conditional expectations $\Psi_i = \ln E(x_i \mid I_{i-1})$ and

$$\Psi_i = \omega + \alpha \ln x_{i-1} + \beta \Psi_{i-1},$$

or more generally : $\Psi_i = \omega + \alpha g(x_{i-1}, \epsilon_{i-1}) + \beta \Psi_{i-1}$

The authors proceed to apply several variants of the model to three different stocks, demonstrating how the model can easily be extended to incorporate potentially important explanatory microstructure variables such as market depth, and changes in the price or spread.

High Frequency Trading in a Limit Order Book
Marco Avellaneda & Sasha Stoikov, Quantitative Finance, 2008, vol. 8, issue 3, pages 217-224

The authors study a stock dealer’s strategy for submitting bid and ask quotes in a limit order book. The agent faces an inventory risk due to the diffusive nature of the stock’s mid-price and a transactions risk due to a Poisson arrival of market buy and sell orders. After setting up the agent’s problem in a maximal expected utility framework, they derive the solution in a two step procedure. First, the dealer computes a personal indifference valuation for the stock, given his current inventory. Second, he calibrates his bid and ask quotes to the market’s limit order book. The authors compare this ”inventory-based” strategy to a ”naive” best bid/best ask strategy by simulating stock price paths and displaying the P&L profiles of both strategies. They find that their strategy has a P&L profile that has both a higher return and lower variance than the benchmark strategy.

Operating within the framework of a continuous time GBM process, the authors define the concept of an indifference price, being the price at which holder of a stock portfolio would be indifferent to his current portfolio and his portfolio +/- one stock and show that this is given by

$$r(s, t) = s - \mathrm{q}\gamma\sigma^2(T - t)$$

where s is the stock price, q is the current portfolio inventory, γ is a coefficient of risk aversion, σ is the stock volatility and T is the investment horizon. The authors show that the asymptotically optimal spread is given by

$$\delta^a + \delta^b = \gamma\sigma^s(T - t) + \frac{2}{\gamma}\ln\left(1 + \frac{\gamma}{k}\right)$$

The authors run some simulation tests to demonstrate that an inventory strategy based on this spread formula significantly outperforms a comparable strategy using a symmetric spread.

Price Dynamics in a Markovian Limit Order Market
Rama Cont, Adrien de Larrard, Forthcoming

The authors propose and study a simple stochastic model for the dynamics of a limit order book, in which arrivals of market order, limit orders and order cancellations are described in terms of a Markovian queuing system. Through its analytical tractability, the model allows to obtain analytical expressions for various quantities of interest such as the distribution of the duration between price changes, the distribution and autocorrelation of price changes, and the probability of an upward move in the price, conditional on the state of the order book. The authors study the diffusion limit of the price process and express the volatility of price changes in terms of parameters describing the arrival rates of buy and sell orders and cancellations. These analytical results provide some insight into the relation between order ow and price dynamics in order-driven markets.

Based on the simplifying assumption that the spread is equal to one tick, the authors start with a representation of the state of the order book in the form of a triplet:

$$X_t = \left(s_t^b,\ q_t^b,\ q_t^a\right)$$

where $\mathrm{s}_\mathrm{t}^\mathrm{b}$ is the current top of book bid price, $\mathrm{q}_\mathrm{t}^\mathrm{b}$ is the size of the bid queue, representing the total size of limit orders at the bid, while $\mathrm{q}_\mathrm{t}^\mathrm{a}$ is the size of the ask queue.

The state X_t of the order book is modified by order book events, being market orders, limit orders and cancellations, which are assumed to occur according to independent Poisson processes with arrival rates μ, λ and θ respectively. It is assumed that once the bid queue is depleted the price decreases by one tick and when the ask queue depleted the price increases by one tick. The new queue size then corresponds to what was previously the number of orders sitting at the price immediately below (resp. above) the best bid (resp. ask). Instead of keeping track of these queues (and the corresponding order flow) at all price levels, the authors treat these sizes as stationary variables drawn from a certain distribution f on N^2. Here f(x,y) represents the probability of observing $\left(\mathrm{q}_\mathrm{t}^\mathrm{b},\ \mathrm{q}_\mathrm{t}^\mathrm{a}\right)$ =(x,y) right after a price increase (and we can envisage a similar, but possibly different probability function for price decreases).

In this simple, but elegant, framework the authors are able to derive expressions for various quantities of interest such as:

- The conditional distribution of the duration between price moves
- The probability of a price increase, given the state of the order book

- The dynamics of the price autocorrelations and the distribution and autocorrelations of price changes
- The volatility of the price

They show that all of these quantities may be characterized in terms of order flow statistics.

Price Dynamics in Limit Order Markets
Christine Parlour, The Review of Financial Studies, Vol. 11, No. 4 (Winter, 1998), pp. 789-816

This article presents a one-tick dynamic model of a limit order market. Agents choose to submit a limit order or a market order depending on the state of the limit order book. Each trader knows that her order will affect the order placement strategies of those who follow and the execution probability of her limit order is endogenous. All traders take this into account which, in equilibrium, generates systematic patterns in transaction prices and order placement strategies even with no asymmetric information.

In this simplified game-theoretic framework, the author proceeds to make a number of interesting propositions about how the probabilities of execution of limit buy and sell orders evolve over time. This is a widely-referenced article that is included here for the sake of completeness.