

ALGORITHMIC FINANCE

Forecasting prices from level-I quotes in the presence of hidden liquidity

Marco Avellaneda; Josh Reed; Sasha Stoikov

Algorithmic Finance (2011), 1:1, 35-43

DOI: 10.3233/AF-2011-004

Abstract, HTML, and PDF:

<http://algorithmicfinance.org/1-1/pp35-43>

Aims and Scope *Algorithmic Finance is a high-quality academic research journal that seeks to bridge computer science and finance, including high frequency and algorithmic trading, statistical arbitrage, momentum and other algorithmic portfolio management strategies, machine learning and computational financial intelligence, agent-based finance, complexity and market efficiency, algorithmic analysis on derivatives, behavioral finance and investor heuristics, and news analytics.*

Managing Editor

Philip Maymin, NYU-Polytechnic Institute

Deputy Managing Editor

Jayaram Muthuswamy, Kent State University

Advisory Board

Kenneth J. Arrow, Stanford University
Herman Chernoff, Harvard University
David S. Johnson, AT&T Labs Research
Leonid Levin, Boston University
Myron Scholes, Stanford University
Michael Sipser, Massachusetts Institute of Technology
Richard Thaler, University of Chicago
Stephen Wolfram, Wolfram Research

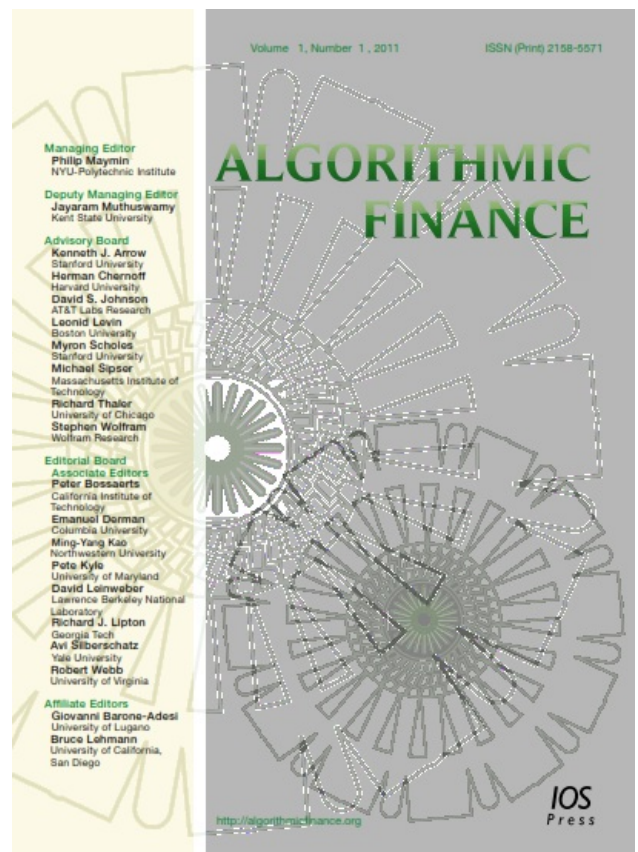
Editorial Board

Associate Editors

Peter Bossaerts, California Institute of Technology
Emanuel Derman, Columbia University
Ming-Yang Kao, Northwestern University
Pete Kyle, University of Maryland
David Leinweber, Lawrence Berkeley National Laboratory
Richard J. Lipton, Georgia Tech
Avi Silberschatz, Yale University
Robert Webb, University of Virginia

Affiliate Editors

Giovanni Barone-Adesi, University of Lugano
Bruce Lehmann, University of California, San Diego



Subscription, submission, and other info:
www.iospress.nl/journal/algorithmic-finance

Unique Features of the Journal *Open access:* Online articles are freely available to all. *No submission fees:* There is no cost to submit articles for review. There will also be no publication or author fee for at least the first two volumes. *Authors retain copyright:* Authors may repost their versions of the papers on preprint archives, or anywhere else, at any time. *Enhanced content:* Enhanced, interactive, computable content will accompany papers whenever possible. Possibilities include code, datasets, videos, and live calculations. *Comments:* Algorithmic Finance is the first journal in the Financial Economics Network of SSRN to allow comments. *Archives:* The journal is published by [IOS Press](http://www.iospress.nl). In addition, the journal maintains an [archive on SSRN.com](http://www.ssrn.com). *Legal:* While the journal does reserve the right to change these features at any time without notice, the intent will always be to provide the world's most freely and quickly available research on algorithmic finance. *ISSN:* Online ISSN: 2157-6203. Print ISSN: 2158-5571.

Forecasting prices from level-I quotes in the presence of hidden liquidity

Marco Avellaneda*, Josh Reed[†] and Sasha Stoikov[‡]

**Courant Institute, New York University and Finance Concepts LLC, NY, USA*

[†]*Stern School of Business, New York University, NY, USA*

[‡]*Cornell Financial Engineering Manhattan, NY, USA*

Email: sfs33@cornell.edu

Abstract. Bid and ask sizes at the top of the order book provide information on short-term price moves. Drawing from classical descriptions of the order book in terms of queues and order-arrival rates (Smith et al., 2003), we consider a diffusion model for the evolution of the best bid/ask queues. We compute the probability that the next price move is upward, conditional on the best bid/ask sizes, the *hidden liquidity* in the market and the correlation between changes in the bid/ask sizes. The model can be useful, among other things, to rank trading venues in terms of the “information content” of their quotes and to estimate hidden liquidity in a market based on high-frequency data. We illustrate the approach with an empirical study of a few stocks using quotes from various exchanges.

1. Introduction

The term “order book” (OB) is generally used to describe the bid and ask prices and sizes in continuous-auction exchanges, such as NYSE-ARCA, BATS or NASDAQ. A distinction is often made between Level I quotes, i.e. the best bid/ask prices and sizes, and Level II quotes, which consist of all prices and sizes available in the order book. In either case, the OB provides information on market depth, allowing traders to estimate the impact of their trades. A question of obvious interest, given the high degree of transparency of OB data, is whether the order book provides any information on short-term price moves.

The role of private information on the behavior of agents in financial markets has been a central theme of the market microstructure literature. In recent years, with the emergence of competing electronic trading venues (ECN’s and Dark Pools) for the same asset and algorithmic trading, questions related to the quality, speed and transparency of information on various exchanges have become ever more relevant for regulators and practitioners alike. Parlour and Seppi (2008) offer a comprehensive review of the

theoretical microstructure literature in light of these, now prevalent, limit order markets.

In the empirical microstructure literature, the focus is on the net informational content of measurable quantities in the limit order book, rather than the information content of particular agents in the market. For instance, the influence of variables such as the sizes at the best quotes on future price moves have been shown to be statistically significant (see Harris and Panchapagesan (2005), Hellström and Simonsen (2009), Cao, Hansch, and Wang (2009)). All these studies show that this effect is particularly strong at short time intervals, on the order of 1–5 minutes. In fact, Hasbrouck (1995) shows how one may compare the relative information content of various time series variables on the efficient price of an asset. Hasbrouck used these econometric methods to analyze the initial stages of the US stock market “fragmenting” into regional exchanges, and at the time found “that the preponderance of the price discovery takes place in the New York Stock Exchange (NYSE) (a median 92.7 percent information share)”. Using similar techniques, Cao, Hansch and Wang find that “the order book

beyond the first step (best quotes) is informative - its information share is about 30%”.

Here, we propose a modeling approach that allows one to measure and compare the information content of order books as well as make short term price forecasts (on the order of seconds). We ask a simple, fundamental, question about the OB. Can we forecast the direction of the next price movement, based on bid and ask sizes? The degree to which this can be done given the OB could be called the *information content* of the OB. For example, if the sizes of queues do not provide information, then, if ΔP denotes the next price move,

$$\begin{aligned} \text{Prob.}\{\Delta P > 0 \mid OB\} \\ = \text{Prob.}\{\Delta P < 0 \mid OB\} = 0.5, \end{aligned}$$

where the probabilities are conditional on observing the OB. If the OB is “informative”, we expect that

$$\text{Prob.}\{\Delta P > 0 \mid OB\} = p(OB),$$

i.e. that the order book provides a forecast of next price move in the form of a conditional probability. The information contained in the OB, if any, should tell us to what extent $p(OB)$ differs from 0.5 based on the observation of limit orders in the book and on the statistics of the queue sizes as they vary in time. We will define an upward price move in terms of (i) the depletion of the best ask queue and (ii) the arrival of a new bid order at that price level. Conversely a downward price move will be defined in terms of (i) the depletion of the best bid queue and (ii) the arrival of a new ask order at that price level. Such a careful distinction in defining a price move ΔP is important, since prices are very granular when observed at a high frequency and the time at which these two possible transitions occur are random.

Our approach is inspired by Markov-type models for the order book, first proposed by Smith et al. (2003) and more recently studied in Cont et al. (2010). These models are high-dimensional Markov processes with a state-space consisting of vectors (bid price, bid size) and (ask price, ask size), and of Poisson-arrival rates for market, limit and cancellation orders. They are often referred picturesquely as “zero-intelligence models”, because orders arrive randomly, rather than being submitted by rational traders with a budget, utility objective, memory, etc. Needless to say, a full description of order books as a Markov

process gives rise to a highly complex system and the solution of the model (in any sense) is often problematic and of questionable value in practice. In fact, one might argue that “zero-intelligence” may not characterize the fashion in which continuous auctions are conducted. Traders, often aided by sophisticated computer algorithms, position their orders to take advantage of situations observed in the order book as well as to fill large block orders on behalf of customers. Rules such as first-in-first-out, and the possibility of capturing rebates for posting limit orders (adding liquidity) result in markets in which there is a significant amount of strategizing *conditionally* on the state of the OB.

For this reason, we choose to simplify such models by considering instead a reduced, diffusion-type dynamics for bid and ask sizes and focusing on the top of the book, instead of on the entire OB. Thus, our goal is to create diffusion models, inspired by SFGK or CST, that can be used to forecast the direction of stock-price moves based on measurable statistical quantities. In contrast to CST, we explicitly model bid and ask quotes with some *hidden liquidity*, i.e. sizes that are not shown in the OB, but which may influence the probability of an upward move in the price. The idea of estimating hidden liquidity is not new to the trading literature, though there is no clear consensus on its definition or on how to estimate it. For instance, Burghadt et al. (2006), estimate the magnitude of hidden liquidity by comparing sweep-to-fill prices to VWAP prices. Moreover in the spirit of our computations of $p(OB)$ they show a strong link between order imbalance and the probability that the next traded price is at the bid or at the offer. Also note that there is a close relationship between our probabilities of an upward/downward move and the notion of micro-price (the size weighted mid-price) well known to practitioners (see Gatheral and Oomen).

2. Modeling Level I quotes

In the Markov model of CST, the OB has two distinguished queues representing the sizes at best bid and the best ask levels, which are separated by the minimum tick size. Market, limit and cancellation orders arrive at both queues according to Poisson processes. One of the following two events must then happen first:

1. The ask queue is depleted and the best ask price goes up by one tick and the price “moves up”.

2. The bid queue is depleted and the best bid price goes down by one tick and the price “moves down”.

The dynamics leading to a price change may thus be viewed as a “race to the bottom”: the queue that hits zero first causes the price to move in that direction.

As it turns out, the predictions of such models are not consistent with market observations. If they were, this would imply that if the best ask size becomes much smaller than the best bid size, the probability that the next price move is upward should approach 100%. However, empirical analysis (see Section 4) shows that this probability does not increase to unity as the ask size goes to zero.

2.1. Hidden liquidity

We hypothesize that this happens for two reasons: first, markets are fragmented; liquidity is typically posted on various exchanges. In the U.S. for example, Reg NMS requires that all market orders be routed to the venue with the best price. Moreover, limit orders that could be immediately executed at their limit price on another market need to be rerouted to those venues. Thus, one needs to consider the possibility that once the best ask on an exchange is depleted, the price will not necessarily go up, since an ask order at that price may still be available on another market and a new bid cannot arrive until that price is cleared on all markets. The second reason is the existence of trading algorithms that split large orders into smaller ones that replenish the best quotes as soon as they are depleted (“iceberg orders”). In the sequel, we will model this by assuming that there is a fixed hidden liquidity (size) behind the best bid and ask quotes. This hidden liquidity may correspond to iceberg orders or orders present on another exchange. Quotes on other exchanges, although not technically hidden from anybody, may be subject to latencies and therefore only available to some traders with the fastest data feeds. The main adjustable parameter in our model, the *hidden liquidity*, will be an important indicator of the information content of the OB.

In summary, the main new idea in our interpretation of the OB models is that we do not immediately assume that a *true* change in price occurs when either of the queues first hits zero. Rather, we take the following view. We postulate that a price transition takes place whenever the first of two events happens:

1. The size for the best ask price goes to zero and a new bid order appears at that price. This can only happen if the hidden liquidity at that price is depleted, i.e. all ask orders on all exchanges are cleared at that price and iceberg orders are exhausted.
2. Alternatively, the size for the best bid price goes to zero and a new ask order appears at that price.

2.2. The discrete Poisson model

Adopting the language of queuing theory, we refer to the number of shares offered at the lowest ask price as the *ask queue*. Similarly, the number of shares bid at the highest bid price is called the *bid queue*. In the spirit of CST, we view these queues as following a continuous time Markov chain (CTMC) where time is continuous and share quantities are discrete, consistently with a minimum order size. We adopt the following notation,

$$\begin{aligned}
 h &= \text{minimum order size} \\
 \lambda &= \text{arrival rate of limit orders at the bid} \\
 \mu &= \text{arrival rate of market orders or cancellations at the bid} \\
 \eta &= \text{arrival rate of simultaneous cancellations at the bid and limit orders at the ask}
 \end{aligned} \tag{1}$$

and assume that the arrival rates on the opposite side of the book are identical. Empirically, we know that the queue sizes are negatively correlated. This may be due to the presence of market makers who simultaneously adjust their quotes on the bid and the ask when their assessment of the fair price changes. Therefore, it is convenient to incorporate correlation between the bid and ask queues by introducing the diagonal transition rates η .

The model for the top of the order book is a continuous-time discrete space process in which the evolution of the queues follows a Markov process in which a state is (X_t, Y_t) , where X_t = bid queue size and Y_t = ask queue size at time t . Each state can transition into eight neighboring states by increasing or decreasing the queue sizes by h .

We may write the moments of the process (X_t, Y_t) in terms of the transition rates

$$\begin{aligned}
E[X_{t+\Delta t} - X_t | X_t, Y_t] &= E[Y_{t+\Delta t} - Y_t | X_t, Y_t] \\
&= h(\lambda - \mu) \Delta t + o(\Delta t) \\
E[(X_{t+\Delta t} - X_t)^2 | X_t, Y_t] &= E[(Y_{t+\Delta t} - Y_t)^2 | X_t, Y_t] \\
&= h^2(\lambda + \mu + 2\eta) \Delta t + o(\Delta t) \\
E[(X_{t+\Delta t} - X_t)(Y_{t+\Delta t} - Y_t) | X_t, Y_t] &= h^2(2\eta) \Delta t + o(\Delta t).
\end{aligned} \tag{2}$$

If we assume, for simplicity, that $\lambda = \mu$ the drifts, variances and correlations of queue sizes simplify to

$$\begin{aligned}
m_X &= m_Y = 0 \\
\sigma_X^2 &= \sigma_Y^2 = 2h^2(\lambda + \eta) \\
\rho &= \frac{-\eta}{\lambda + \eta}
\end{aligned} \tag{3}$$

3. Diffusion approximation

3.1. Probability of the ask queue depleting

Let $\langle X \rangle$ and $\langle Y \rangle$ denote, respectively, the median size of the queues X_t, Y_t . We define the coarse-grained variables

$$x = X / \langle X \rangle, \quad y = Y / \langle Y \rangle,$$

which measure the queue sizes “macroscopically”.

The process (x_t, y_t) can be approximated by the diffusion

$$\begin{aligned}
dx_t &= \sigma dW_t^{(1)} \\
dy_t &= \sigma dW_t^{(2)} \\
E(dW^{(1)} dW^{(2)}) &= \rho dt,
\end{aligned} \tag{4}$$

where

$$\sigma^2 = \frac{2h^2(\lambda + \eta)}{\langle X \rangle^2}, \tag{5}$$

ρ is defined in (2.7), and $W^{(1)}, W^{(2)}$ are standard Brownian motions.

This diffusion limit is an approximation of the discrete model in the sense of heavy traffic limits (see Cont, 2011 and Cont and Larrard, 2011 for more details). Essentially, this limit is valid if the average queue sizes are much larger than the typical quantity of shares traded and the frequency of orders per unit time is high, *i.e.* $\langle X \rangle = \langle Y \rangle \gg h$ and $\lambda, \eta \gg 1$.

We are interested in computing the function $u(x, y)$ representing probability that the ask size hits zero before the bid size hits zero, given that we observe the (standardized) bid/ask sizes (x, y) . From diffusion theory, this function satisfies the differential equation

$$\sigma^2(u_{xx} + 2\rho u_{xy} + u_{yy}) = 0, \quad x > 0, y > 0, \tag{6}$$

or, simply,

$$u_{xx} + 2\rho u_{xy} + u_{yy} = 0 \quad \text{for } x > 0, y > 0. \tag{7}$$

If we assume, naively, that the order-book fully represents the liquidity in the market at a particular price level, then the mid-price will move up once the ask queue is depleted *i.e.* when $y_t = 0$ for the first time, since no more sellers are present at that level. In this case, the probability that the price will increase corresponds to the probability that the diffusion (7) exits the quadrant $\{(x, y); x > 0, y > 0\}$ through the x -axis. The corresponding boundary conditions for $u(x, y)$ are therefore

$$\begin{aligned}
u(0, y) &= 0, \quad \text{for } y > 0, \\
u(x, 0) &= 1, \quad \text{for } x > 0.
\end{aligned} \tag{8}$$

Note that the processes X_t and Y_t do not have a drift and therefore do not have a stationary distribution. A drift could be introduced, as in CST, by modeling market orders and cancellations separately, resulting in queues sizes that revert around an equilibrium level. However, such microstructural distinctions are not essential when we fit the model to transactions data, since we are interested in very short term predictions rather than equilibrium distributions.

3.2. Probability of an upward move

What is essential, however, is to make a distinction between an ask queue being depleted and a “genuine” price move where a new bid order appears at price.

We know that an upward price move might not take place when the ask queue is depleted, due to additional liquidity at that level, which we call *hidden liquidity*. This hidden liquidity can be attributed to either iceberg orders or by virtue of a Reg-NMS-type mechanism in which there are other markets that still post liquidity on the ask-side at the same level and which must be honored before the mid price can move up.

A simple way to model this is to assume that there is an additional amount of liquidity, denoted by H , representing the fraction of average book size ($< X >$ or $< Y >$) which is “hidden” or absent from the book. A true price transition takes place if the hidden liquidity is exhausted. In other words, we observe queues of size x or y but the “true” size of the queues are $x + H$ and $y + H$. Thus, if we denote by $p(x, y; H)$ the probability of an upward price move conditional on the observed queue sizes (x, y) and the hidden liquidity parameter H , we have

$$\sigma^2 (p_{xx} + 2\rho p_{xy} + p_{yy}) = 0, \quad x > -H, y > -H,$$

with the boundary condition

$$\begin{aligned} p(-H, y) &= 0, \quad \text{for } y > -H, \\ p(x, -H) &= 1, \quad \text{for } x > -H. \end{aligned}$$

In other words we can solve the problem with boundary conditions at zero and use the relation

$$p(x, y; H) = u(x + H, y + H), \quad (9)$$

where $u(x, y)$ satisfies the diffusion equation (7) on the first quadrant of the (x, y) plane with boundary conditions (8). One could also obtain an effect similar to that of our hidden liquidity parameter by considering mixed Von-Neumann/Dirichlet boundary conditions at zero or bid and ask processes with jumps. Our modeling choice is motivated by the simplicity of our closed form solution.

3.3. Solution

Theorem 3.1. *The probability of an upward move in the mid price is given by*

$$p(x, y; H) = u(x + H, y + H), \quad (10)$$

where

$$u(x, y) = \frac{1}{2} \left(1 - \frac{\text{Arctan} \left(\sqrt{\frac{1+\rho}{1-\rho}} \frac{y-x}{y+x} \right)}{\text{Arctan} \left(\sqrt{\frac{1+\rho}{1-\rho}} \right)} \right). \quad (11)$$

The proof of the above result is in the Appendix.

Remarks.

1. The probability depends on the hidden liquidity H and the correlation ρ , but not on the volatility of queue sizes σ .

2. As H approaches ∞ , $p(x, y, H)$ approaches 0.5. In other words, as the hidden liquidity grows, the bid and ask sizes offer no information on the probability of an upward price move.

3. If we set $\rho = 0$ the above expression simplifies to

$$u(x, y) = \frac{2}{\pi} \text{Arctan} \left(\frac{x}{y} \right). \quad (12)$$

4. As ρ approaches -1 , the numerator and denominator in (22) both tend to zero. The limit as $\rho \rightarrow -1$ is

$$u(x, y) = \frac{x}{x + y}.$$

5. If we consider the sector $y < x$, i.e. the sector for which the ask queue is smaller than the bid queue, $u(x, y)$ is an increasing function of ρ . In fact, setting $\xi = \frac{y-x}{y+x}$ and $\alpha = \sqrt{\frac{1+\rho}{1-\rho}}$,

$$\frac{\partial u}{\partial \rho} = -\frac{1 + \alpha^2}{1 + \alpha^2 \xi^2} \frac{1}{(1 - \rho)^2} \frac{1}{2\xi}.$$

This is a positive quantity since ξ is negative in the sector. Therefore, the assumption $\rho = -1$ will underestimate the probability of an up-tick if the “true” correlation was different than -1 .

4. Data analysis

In this section, we study the information content of the best quotes for the tickers QQQQ, XLF, JPM, and AAPL over the first five trading days in 2010 (i.e. Jan 4-8). All four tickers are traded on various exchanges, and this allows us to compare the information content of these venues. In other words we will be computing the probability

$$\text{Prob.}\{\Delta P > 0 \mid OB\} = p(OB)$$

discussed in the introduction, for ΔP defined to be the next midprice move, and for OB defined to be the pair of bid and ask sizes (x, y) .

In our data analysis, we focus on the hidden liquidity for the perfectly negatively correlated queues model, i.e.

$$p(x, y; H) = \frac{x + H}{x + y + 2H} \quad (13)$$

which we estimate by minimizing square errors with respect to the empirical probabilities. We make this choice because the parameter H has a more important impact on p than ρ and optimization routines often converged towards the $\rho = -1$ case in our data set.

In practice, when performing our data analysis, we find it easier to bucket the data in deciles of queue sizes, rather than normalizing by the mean queue size, as we did in Section 3. The implied hidden liquidity parameter we compute in the sequel can therefore be interpreted as a fraction of the maximum observed queue size.

4.1. Data description

The data comes from the WRDS database, more specifically the consolidated quotes of the NYSE-TAQ data set. Each row has a timestamp (between the hours of 10:00 and 16:00, rounded to the nearest second), a bid price, an ask price, a bid size, an ask size and an exchange flag, indicating if the quote was on NASDAQ (T), NYSE-ARCA (P) or BATS (Z), see Table 1 for a sample of the data. There are other regional exchanges, but for the purpose of this study, we focus on these venues as they have significantly more than one quote per second.

In table 2, we present some summary statistics for the tickers QQQQ, XLF, JPM and AAPL, across the three exchanges. The tickers QQQQ, XLF and JPM are ideal candidates, because their bid-ask spread is almost always one tick (or one cent) wide, much like our

stochastic model. We also pick AAPL, whose spread most often trades at 3 cents (or three ticks wide), due to AAPL's relatively high stock price. Though our model does not strictly consider spreads greater than one, we use it to fit our model, conditional on the spread, i.e. $OB = (x, y, s)$ where s is the spread in cents.

4.2. Estimation procedure

1. We split the data set into three subsets, one for each exchange. Items 2-6 are repeated separately for each exchange and each ticker.
2. We remove zero and negative spreads.
3. We "bucket" the bid and ask sizes, by taking deciles of the bid and ask size and normalizing queue sizes so that (i, j) represents the i th decile of the bid size and the j th decile of the ask size respectively.
4. For each bucket (i, j) , we compute the empirical probability that the price goes up u_{ij} . This is done by looking forward to the next mid price change and computing the empirical percentage of occurrences of (i, j) that ended up going up, before going down.
5. We count the number of occurrences of the (i, j) bucket, and denote this distribution d_{ij} .
6. We minimize least squares for the negatively correlated queues model, i.e.

$$\min_H \sum_{i,j} \left[\left(u_{ij} - \frac{i + H}{i + j + 2H} \right)^2 d_{ij} \right] \quad (14)$$

and obtain an implied hidden liquidity H for each exchange.

4.3. Results

We first illustrate the predictions of our model for the ticker XLF on the Nasdaq exchange (T). We report the empirical probabilities of an up move, given the

Table 1
A sample of the raw data

symbol	date	time	bid	ask	bsize	asize	exchange
QQQQ	2010-01-04	09:30:23	46.32	46.33	258	242	T
QQQQ	2010-01-04	09:30:23	46.32	46.33	260	242	T
QQQQ	2010-01-04	09:30:23	46.32	46.33	264	242	T
QQQQ	2010-01-04	09:30:24	46.32	46.33	210	271	P
QQQQ	2010-01-04	09:30:24	46.32	46.33	210	271	P
QQQQ	2010-01-04	09:30:24	46.32	46.33	161	271	P

Table 2
Summary statistics

Ticker	Exchange	num quotes	quotes/sec	avg(spread)	avg(bsize + asize)	avg(price)
XLF	NASDAQ	0.7M	7	0.010	8797	15.02
XLF	NYSE	0.4M	4	0.010	10463	15.01
XLF	BATS	0.4M	4	0.011	7505	14.99
QQQQ	NASDAQ	2.7M	25	0.010	1455	46.30
QQQQ	NYSE	4.0M	36	0.011	1152	46.27
QQQQ	BATS	1.6M	15	0.011	1055	46.28
JPM	NASDAQ	1.2M	11	0.011	87	43.81
JPM	NYSE	0.7M	6	0.012	47	43.77
JPM	BATS	0.6M	5	0.014	39	43.82
AAPL	NASDAQ	1.3M	13	0.034	9.1	212.50
AAPL	NYSE	0.4M	4	0.046	5.7	212.66
AAPL	BATS	0.6M	6	0.054	4.5	212.43

Table 3

Empirical vs. Model probabilities for the probability of an upward move (XLF), on Nasdaq (T). Rows represent bid size percentiles (i), columns represent ask size percentiles (j). The model is given by $p(i, j) = \frac{i+H}{i+j+H}$ with $H = 0.15$

decile sizes	0.1 < 1250	0.2 < 1958	0.3 < 2753	0.4 < 3841	0.5 < 4835	0.6 < 5438	0.7 < 5820	0.8 < 6216	0.9 < 6742
0.1	0.50	0.38	0.25	0.25	0.32	0.26	0.23	0.23	0.15
0.2	0.61	0.50	0.47	0.41	0.36	0.40	0.38	0.27	0.20
0.3	0.75	0.53	0.50	0.43	0.39	0.37	0.43	0.39	0.28
0.4	0.74	0.58	0.57	0.50	0.42	0.42	0.47	0.46	0.37
0.5	0.68	0.64	0.61	0.58	0.50	0.51	0.48	0.49	0.41
0.6	0.74	0.60	0.63	0.58	0.49	0.50	0.50	0.50	0.49
0.7	0.78	0.62	0.57	0.53	0.52	0.50	0.50	0.60	0.53
0.8	0.77	0.73	0.61	0.54	0.51	0.50	0.40	0.50	0.42
0.9	0.85	0.79	0.72	0.63	0.60	0.51	0.47	0.57	0.50
decile sizes	0.1 = 1250	0.2 = 1958	0.3 = 2753	0.4 = 3841	0.5 = 4835	0.6 = 5438	0.7 = 5820	0.8 = 6216	0.9 = 6742
0.1	0.50	0.42	0.36	0.31	0.28	0.25	0.23	0.21	0.19
0.2	0.58	0.50	0.44	0.39	0.35	0.32	0.29	0.27	0.25
0.3	0.64	0.56	0.50	0.45	0.41	0.37	0.35	0.32	0.30
0.4	0.69	0.61	0.55	0.50	0.46	0.42	0.39	0.37	0.34
0.5	0.72	0.65	0.59	0.54	0.50	0.46	0.43	0.41	0.38
0.6	0.75	0.68	0.63	0.58	0.54	0.50	0.47	0.44	0.42
0.7	0.77	0.71	0.65	0.61	0.57	0.53	0.50	0.47	0.45
0.8	0.79	0.73	0.68	0.63	0.59	0.56	0.53	0.50	0.47
0.9	0.81	0.75	0.70	0.66	0.62	0.58	0.55	0.53	0.50

bid and ask sizes in table 3, as well as the model probabilities, given by equation (13) with H estimated with the procedure described above. Notice that even for very large bid sizes and small ask sizes (say the 90th percentile of sizes at the bid and the 10th percentile of sizes at the ask) the empirical probability of the mid price moving upward is high (0.85) but

not arbitrary close to one. The same is true of our model, which assumes there is a hidden liquidity H behind both quotes. We interpret H as a measure of the information content of the bid and ask sizes: the smaller H is, the more size matters. The larger the H , the closer all probabilities will be to 0.5, even for drastic size imbalances.

Table 4
Implied hidden liquidity across tickers and exchanges

Ticker	NASDAQ	NYSE	BATS
XLF	0.15	0.17	0.17
QQQQ	0.21	0.04	0.18
JPM	0.17	0.17	0.11
AAPL $s = 1$	0.16	0.90	0.65
AAPL $s = 2$	0.31	0.60	0.64
AAPL $s = 3$	0.31	0.69	0.63

In table 4, we display the hidden liquidity H for the four tickers and three exchanges. These results indicate that size is most important for

- XLF on NASDAQ,
- QQQQ on NYSE-ARCA
- JPM on BATS.

Finally we calculate H for AAPL, for different values of the bid-ask spread ($s = 1, 2$ and 3 cents). We find that

- AAPL quotes are more informative on NASDAQ, and that they matter most when the spread is small.

Modeling stocks with larger spreads may require more sophisticated models of the order book, possibly including Level II information. Since a majority of US equities trade at average spreads of several cents, we consider this avenue worthy of future research.

5. Conclusions

Based on a diffusion model of the liquidity at the top of the order book, we proposed closed-form solutions for the probability of a price uptick conditional on Level-I quotes. The probability is a function of the bid size, the ask size and an adjustable parameter, H , the *hidden liquidity*. The advantage of this simple model is that it can be fitted to high-frequency data and produces an *implied* hidden liquidity parameter, obtained by fitting tick data (from WRDS) to the proposed formulas. The result is that we can classify different markets in terms of their hidden liquidity or, equivalently, how informative the Level I quotes of a stock are in terms of forecasting the next price move (up or down). If the hidden size is small (compared to the typical size shown in the market under consideration), we say that the best quotes are informative. Statistical analysis for different stocks shows the following results:

- for XLF (SPDR Financial ETF), NASDAQ has the least hidden liquidity;
- for QQQQ (Powershares Nasdaq-100 Tracker), NYSE-ARCA has the least hidden liquidity;
- for JPM (J.P. Morgan & Co.) BATS has the least hidden liquidity and
- for AAPL (Apple Inc.) NASDAQ has the least hidden liquidity.

We used only 5 days of data for these calculations and a study of the stability of our hidden liquidity parameter over longer periods remains to be done. Nevertheless, the approach presented here seems to provide a way of comparing trading venues, in terms of their information content and hidden liquidity, and hence on the possibility of forecasting price changes from their orderbook data.

Appendix

Solution of the PDE (7) for general ρ

Proposition 1. *Let $\Omega(X, Y)$ be a harmonic function. Let us set*

$$v(\zeta, \eta) = \Omega\left(\frac{\zeta}{\sigma_1}, \frac{\eta}{\sigma_2}\right).$$

Then,

$$\sigma_1^2 v_{\zeta\zeta} + \sigma_2^2 v_{\eta\eta} = 0. \quad (15)$$

Proof. By the chain rule, $\sigma_1^2 v_{\zeta\zeta} = \sigma_1^2 \frac{\Omega_{XX}}{\sigma_1^2} = \Omega_{XX}$. The same holds for the η -derivative. Add and use harmonicity of Ω . \square

Proposition 2. *Let Ω be a harmonic function. Then*

$$u(x, y) = \Omega\left(\frac{y+x}{\sqrt{2}\sqrt{1+\rho}}, \frac{y-x}{\sqrt{2}\sqrt{1-\rho}}\right) \quad (16)$$

satisfies

$$u_{xx} + 2\rho u_{xy} + u_{yy} = 0. \quad (17)$$

Proof. Let $\sigma_1 = \sqrt{1+\rho}$, $\sigma_2 = \sqrt{1-\rho}$ and set $\zeta = \frac{y+x}{\sqrt{2}}$, $\eta = \frac{y-x}{\sqrt{2}}$. Clearly, by Proposition 1, $v(\zeta, \eta) \equiv \Omega(\frac{\zeta}{\sigma_1}, \frac{\eta}{\sigma_2})$ satisfies

$$\sigma_1^2 v_{\zeta\zeta} + \sigma_2^2 v_{\eta\eta} = 0. \quad (18)$$

Since $u(x, y) = v\left(\frac{y+x}{\sqrt{2}}, \frac{y-x}{\sqrt{2}}\right)$, we have, after differentiating twice the function u

$$\begin{aligned} u_{xx} &= \frac{1}{2}v_{\zeta,\zeta} + \frac{1}{2}v_{\eta,\eta} - v_{\zeta\eta} \\ u_{yy} &= \frac{1}{2}v_{\zeta,\zeta} + \frac{1}{2}v_{\eta,\eta} + v_{\zeta\eta} \\ u_{xy} &= \frac{1}{2}v_{\zeta,\zeta} - \frac{1}{2}v_{\eta,\eta}. \end{aligned} \quad (19)$$

Adding the first two terms and then adding the third one multiplied by 2ρ gives

$$\begin{aligned} u_{xx} + 2\rho u_{xy} + u_{yy} &= \frac{1}{2}v_{\zeta,\zeta} + \frac{1}{2}v_{\eta,\eta} - v_{\zeta\eta} \\ &\quad + 2\rho \left(\frac{1}{2}v_{\zeta,\zeta} - \frac{1}{2}v_{\eta,\eta} \right) \\ &\quad + \frac{1}{2}v_{\zeta,\zeta} + \frac{1}{2}v_{\eta,\eta} + v_{\zeta\eta} \\ &= v_{\zeta,\zeta} + \rho(v_{\zeta,\zeta} - v_{\eta,\eta}) + v_{\eta,\eta} \\ &= (1+\rho)v_{\zeta,\zeta} + (1-\rho)v_{\eta,\eta} \\ &= \sigma_1^2 v_{\zeta\zeta} + \sigma_2^2 v_{\eta\eta} \\ &= 0. \end{aligned} \quad (20)$$

□

Theorem 5.1. *The probability of an upward move in the mid price is given by*

$$p(x, y; H) = u(x + H, y + H), \quad (21)$$

where

$$u(x, y) = \frac{1}{2} \left(1 - \frac{\text{Arctan} \left(\sqrt{\frac{1+\rho}{1-\rho}} \frac{y-x}{y+x} \right)}{\text{Arctan} \left(\sqrt{\frac{1+\rho}{1-\rho}} \right)} \right). \quad (22)$$

Proof. Use $\Omega(X, Y) = \text{Arctan}(Y/X)$ and apply Proposition 2, using

$$X = \frac{y+x}{\sqrt{2}\sqrt{1+\rho}}; \quad Y = \frac{y-x}{\sqrt{2}\sqrt{1-\rho}}.$$

It follows that the function

$$u(x, y) = \frac{1}{2} \left(1 - \frac{\text{Arctan} \left(\sqrt{\frac{1+\rho}{1-\rho}} \frac{y-x}{y+x} \right)}{\text{Arctan} \left(\sqrt{\frac{1+\rho}{1-\rho}} \right)} \right) \quad (23)$$

satisfies

equation (3.4). Furthermore, we have $u(x, 0) = 1$ and $u(0, y) = 0$. □

References

- Burghadt, G., Hanweck, J., Lei, L., 2006. Measuring Market Impact and Liquidity. *The Journal of Trading* 1 (4) 70–84.
- Cao, C., Hansch, O., Wang, X., 2009. The information content of an open limit order book, *The Journal of Futures Markets* 29(1), 16–41.
- Cont, R., Larrard, A., 2011. Price dynamics in a Markovian limit order market. Working Paper.
- Cont, R., 2011. Statistical modeling of high frequency financial data: facts, models and challenges. *IEEE Signal Processing* 28 (5), 16–25.
- Cont, R., Stoikov, S., Talreja, R., 2010. A stochastic model for order book dynamics. *Operations Research* 58 (3) 549–563.
- Gatheral, J., Oomen, R., 2010. Zero-intelligence realized variance estimation. *Finance and Stochastics* 24(2), 249–283.
- Harris, L., Panchapagesan, V. 2005. The information content of the limit order book: evidence from the NYSE specialist trading decisions. *Journal of Financial Markets* 8, 25–67.
- Hasbrouck, J., 1995. One security, many markets: determining the contributions to price discovery. *Journal of Finance* 1 (4), 1175–1199.
- Hellström, J., Simonsen, O., 2009. Does the open limit order book reveal information about short-run stock price movements? Working paper.
- Parlour, C., Seppi, D., 2008. Limit order markets: a survey. In: Boot, A.W.A., Thakor, A.V. (Eds.), *Handbook of Financial Intermediation & Banking*.
- Smith, E., Farmer, J.D., Gillemot, L., Krishnamurthy, S., 2003. Statistical theory of the continuous double auction. *Quantitative Finance* 3, 481–514.