2009-08-31

Liquidity Modeling Using Order Book Data

Yi Li
Worcester Polytechnic Institute

Liquidity Modeling Using Order Book Data

by

Yi Li

A Project Report

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Financial Mathematics

August 2009

APPROVED:

Professor Marcel Y. Blais, Capstone Advisor

Professor Bogdan Vernescu, Head of Department

Abstract

On a stock exchange, trading activity has an impact on stock prices. Market agents place limit orders, which come in the form of bids and asks. These orders wait in the market to be executed when another agent agrees to fulfill the transaction. We examine an "inventory-based" quoting strategy model developed by Marco Avellaneda & Sasha Stoikov [1]. We expand on their work by developing a method to calibrate the model to market data using limit order data provided by Morgan Stanley [9]. We consider solving a least squares problem which fits the model to the data using a sensitivity parameter γ.

ACKNOWLEDGMENTS

I would like to gratitude to my advisor, Professor Marcel Blais for all his support as being a mentor through the time of my study at Worcester Polytechnic Institute. I would like to thank my family for loving me unconditionally with heart and soul. Also, I would like to thank all my friends for your support.

Contents

List of Figures

1 Introduction

On a stock exchange, such as the New York Stock Exchange (NYSE), the NASDAQ, or the London Stock Exchange, a trade occurs when one market agent buys a share of a stock for a specified price and another market agent sells a share of that same stock at the same specified price. The exchange exists to facilitate this mechanism. Trades can be initiated in different ways. There are two main types of orders that an agent can issue in an attempt to initiate a stock trade, market orders and limit orders.

A market order (unlimited order) is an order to buy or sell a number of shares of a stock immediately at the best available price. As opposed to a market order, a limit order is defined as an order placed with a brokerage to buy or sell a number of shares of a stock at a specified price (or better); however the execution is not immediate and may not necessarily happen. Limit orders are maitained by specialists who are also refered to as “dealers” [1], and a limit order book is a record of all the unexecuted limit orders in the market. In this paper, we will mainly focus on a model that uses optimal quotes submission strategies in a limit order book introduced by Marco Avellaneda & Sasha Stoikov [1].

Most basic mathematical finance models of security markets generally assume that at any given time t one can buy or sell any number of shares of a stock at the market price s_t without affecting that price. In other words, every market agent is a price taker. This framework does not account for any liquidity issues, such as dealer interactions and how the prices and quantities are determined through the mechanics of a trade. In reality, stock trading activities have an impact on the market price of a stock.

The quantity of a stock held in hand by a dealer is called inventory. An inventory can be liquidated by buying or selling. If a limit order is initiated by a dealer who wants to buy, the order is set at a bid price in the market which is usually below the mid-price. If a limit order is initiated by a seller, the price for the stock he sets in the market is the ask price. For instance, suppose in limit orders there are 1,000 shares of Microsoft stock at a price of $100 per share waiting to be sold in the market and 1,500 shares at a price of $100.5. Now a market order for buying 2,000 shares of Microsoft stock is placed. After the first 1,000 shares at $100 are exhausted, the buyer would have to buy another 1,000 shares at $100.5. The price per share for the entire trade is thus $100.25. This trade has increased the market price of the stock. In general, liquidity issues like this exsit as transactions occur. The concept of liquidity has been modeled in many different ways in the literature (see, Longstaff [14], Keynes [13], Engel [8], Baum [2]), and Çetin [5]).

In the paper by Marco Avellaneda & Sasha Stoikov [1], the authors presented an approach that incorporates liquidity. Avellaneda & Stoikov work in the framework where the price per share of an asset is a given by a function $S(t, x)$, where x represents the size of a trade. $x > 0$ indicates a buyer-initiated market order, and $x < 0$ indicates a seller-initiated market order[1]. On the grounds of a maximal expected utility framework, the paper suggests a two step procedure:

1. The dealer computes a personal reservation valuation for the stock, given current inventory.

2. The dealer calibrates bid and ask quotes to the market's limit order book.

This is an"inventory-based" strategy for submitting bid and ask quotes in a limit order book. The model introduced in Avellaneda & Stoikov [1] involves the use of a sensitivity parameter γ. We introduce a method that calibrates the model to order book data provided by Morgan Stanley[9] using a least-squares fitting approach.

2 Background

2.1 Basic Mathematical Finance Concepts

We begin with some important definitions that are fundamental concepts in mathematical finance.

Definition: (Brownian Motion) Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. For each $\omega \in \Omega$, suppose there is a continuous function $W(t)$ for $t \geq 0$ that satisfies $W(0) = 0$ and depends on ω. Then $W(t)$, $t \geq 0$, is a Brownian motion if for all $0 = t_0 < t_1 < \cdots < t_m$, the increments

$$W(t_1) = W(t_1) - W(t_0), W(t_2) - W(t_1), \ldots, W(t_m) - W(t_{m-1}) \quad (1)$$

are independent and each of these increments is normally distributed with

$$E[W(t_{i+1}) - W(t_i)] = 0, \quad (2)$$

$$Var[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i.[19] \quad (3)$$

Definition: (Black-Scholes Model) The Black-Scholes model consists of two assets with dynamics given by the stochastic differential equations

$$dB(t) = rB(t)dt, \quad (4)$$

$$dS(t) = \alpha S(t)dt + \sigma S(t)dW(t) \quad (5)$$

where r is the rate of interest, α is the local mean rate of return of S, often referred to as the drift, and σ is known as the volatility of S. [4]

Definition: (Convex Risk Measure[10]) A mapping $f : X \rightarrow R^d$ is called a convex risk measure if $f(0)$ is finite and if f satisfies the following conditions for all $x, y \in X$:

- Monotonicity: If $x \leq y$, then $f(x) \geq f(y)$.
- Translation Invariance: If $m \in R$, then $f(x + m) = f(x) - m$.
- Convexity: $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$, for $0 \leq 1$.

2.2 Liquidity

In this model the mid-price path of the stock is simulated using the Black-Scholes model [4] with constant volatility σ, and initial value $S_t = s$ under the assumptions that dealers have no opinion on the drift. The interest rate is assumed to be 0 for simplicity, so the stock price dynamics are given by

$$d(S_t) = \sigma d(W_t) \tag{6}$$

where W_t is a standard one-dimensional Brownian Motion. Avellaneda & Stoikov [1] choose a convex risk measure to set up a value function that interprets the agent's objective as

$$u(x, s, q, t) = \max_{\delta^a, \delta^b} E_t[-exp(-\gamma(X_T + q_T S_T)] \tag{7}$$

where δ^b and δ^a represent the differences between the mid-prices and bid-ask prices, X_T is dealer's wealth, q_T is the number of shares in the market waiting to be bought or sold in a limit order, and S_T is the mid-market price at time $t = T$.

The reservation bid price r_b is defined as the price where the dealer is indifferent between his or her current portfolio and his or her current portfolio plus one share of stock. Similarly, the reservation ask price r_a is defined as the price where the dealer is indifferent between his or her current portfolio and his or her current portfolio minus one stock [1].

There are two circumstances in a continuous-time setting, the finite horizon and the infinite horizon. The optimizing dealer with finite horizon has respective reservation bid and ask prices

$$r^a(s,q,t) = s + (1-2q)\frac{\gamma\sigma^2(T-t)}{2} \tag{8}$$

and

$$r^b(s,q,t) = s + (-1-2q)\frac{\gamma\sigma^2(T-t)}{2}. \tag{9}$$

The optimizing dealer with infinite horizon has the two prices

$$\bar{r}^a(s,q) = s + \frac{1}{\gamma}\ln(1 + \frac{(1-2q)\gamma^2\sigma^2}{2\omega - \gamma^2 q^2\sigma^2}) \tag{10}$$

and

$$\bar{r}^b(s,q) = s + \frac{1}{\gamma}\ln(1 + \frac{(-1-2q)\gamma^2\sigma^2}{2\omega - \gamma^2 q^2\sigma^2}) \tag{11}$$

where $\omega > \frac{1}{2}\gamma^2\sigma^2 q^2$. Here γ is a sensitivity parameter.

The average of these two prices is refered to as the reservation (or indifference) price.

$$r(s,q,t) = \frac{r^a + r^b}{2} \tag{12}$$

After working through the analysis a final formula for indifference or reservation price is derived in [1] to be

$$r(s,q,t) = s - q\gamma\sigma^2(T-t). \tag{13}$$

A $q > 0$ represents a long position. The reservation price r is below the mid-price, indicating an intent to sell. Otherwise $q < 0$ represents a short position, i.e. the reservation price is above the mid-price on the behalf of the willingness to buy.

3 Our Approach

We first establish the fundamental framework for our model. We will use a standard assumption that all trading takes place in a continuous-time setting. Avellaneda & Stoikov conducted their research based on a set of simulated stock price data using the Black-Scholes model. In this paper we will build a bid-ask spread model using real limit order data provided by Morgan Stanley[9]. Using the data we will compute daily stock prices, historical volatility, and address the sensitivity parameter γ. We will set up a least square approach to fit in γ and solve for the optimal solution γ^*.

3.1 Order Book Data

The order book data we used was provided by Robert Ferstenberg at Morgan Stanley[9]. The data set contains the complete posting of the top 10 bids and the top 10 asks, including both prices and sizes (number of shares at available at each price) for various stocks from 7/01/2003 to 12/23/2003.

Trading updates were extracted by Professor Marcel Y. Blais and restored in two types of matrices, one consisting of prices and the other consisting of

Notations	
tp	trade price
tt	trade time
ts	trade size
bap$x$	the x^th best ask price
bas$x$	corresponding size to the x^th best ask price
bbp$x$	the x^th best bid price
bbs$x$	corresponding size to the x^th best bid price

Table 1: Order Book Data Notation

sizes, they have the form as illustrated in table

Time	quotes
t_1	bp1 bp2 bp3 bp4 bp5 bp6 bp7 bp8 bp9 bp10
t_2	bp1 bp2 bp3 bp4 bp5 bp6 bp7 bp8 bp9 bp10
t_3	bp1 bp2 bp3 bp4 bp5 bp6 bp7 bp8 bp9 bp10
t_4	bp1 bp2 bp3 bp4 bp5 bp6 bp7 bp8 bp9 bp10
t_5	bp1 bp2 bp3 bp4 bp5 bp6 bp7 bp8 bp9 bp10
$\vdots$	$\dots$

Table 2: Price matrix P

With these matirces data processing is convenient and easy to handle. For example, if at time t_i, the ${t_i}^{th}$ row in the size matrix Q of a stock S contains negative entries, this indicates a negative number of shares waiting to be bought. This row thus represents a set of bid quotes, and each element in this row represents the size of a best bid. The ${t_i}^{th}$ row in the price matrix P gives the top 10 best bid prices corresponding to these sizes. Similarly, if the ${t_i}^{th}$ row in Q contains positive entries, then the corresponding row in P gives the top 10 ask prices. We use the current top 10 best bid prices and top 10 best ask prices to build the limit order book from this data set.

Time	quotes
t_1	bs1 bs2 bs3 bs4 bs5 bs6 bs7 bs8 bs9 bs10
t_2	bs1 bs2 bs3 bs4 bs5 bs6 bs7 bs8 bs9 bs10
t_3	bs1 bs2 bs3 bs4 bs5 bs6 bs7 bs8 bs9 bs10
t_4	bs1 bs2 bs3 bs4 bs5 bs6 bs7 bs8 bs9 bs10
t_5	bs1 bs2 bs3 bs4 bs5 bs6 bs7 bs8 bs9 bs10
$\vdots$	$\ldots$

Table 3: Size matrix Q

3.2 Limit Orders

A market order (unlimited order) is an order to buy or sell a number of shares of a stock immediately at the best available price. A limit order is defined as an order placed with a brokerage to buy or sell a number of shares of a stock at a specified price (or better); however the execution is not immediate and may not necessarily happen. Limit orders are maitained by specialists who are also refered to as "dealers" [1], and a limit order book is a record of all the unexecuted limit orders in the market.

The current shape of the limit order book determines the priority of limit order execution if a large market order arrives. For instance, suppose a seller wants to sell Q shares of a stock in the market. According to the distances between the bid pirce and the mid-pirce, limit orders with the highest bid prices will be automatically executed. As the highest bid quotes get exhausted the dealer would have to move on to the second highest bid quotes, as this trading procedure goes on, the price that the trades occur at decreases. As a result, an impact on the market price occurs since the transaction take place at a price lower than the mid-price.

Consider a more general case. Suppose there are $Q_B = \begin{bmatrix} q_{B_1} & \dots & q_{B_n} \end{bmatrix}$ aggregate bids at prices $P_B = \begin{bmatrix} P_{B_1} & \dots & P_{B_n} \end{bmatrix}$ at time t_0. If a seller comes to market at time t_0 and sells K shares, then the first q_{B_1} shares will be executed at a price of P_{B_1}. The seller will then sell q_{B_2} shares at a price of P_{B_2}, and continue moving on to the next highest price until he has sold all K shares. Thus the average price per share for this trade is given by

$$P_B^* = \frac{\sum_{i=1}^{n-1} q_i P_i + Q_n P_n}{\sum_{i=1}^{n-1} q_i + Q_n} \tag{14}$$

where $n = \inf\{k : \sum_{i=1}^{k-1} q_i \leq K < \sum_{i=1}^{k} q_i\}$ and $Q_n = K - \sum_{i=1}^{n-1} q_i$. [3]

The further away from the mid-price a limit order is, the less likely it is that the limit order will be executed. Limit orders, whether they are bids or asks, are more likely to be executed if the bid or ask price is close to the mid-price.

Ticker	bbs1	bbs2	bbs3	bbs4	bbs5	Time
nts	23660	20000	20405	7900	20404	162455
Ticker	bas1	bas2	bas3	bas4	bas5	Time
nts	12970	9973	16104	32100	7900	19450

Table 4: Sample Size Data

Take the above two tables for example, at time t_0 with a highest bid price of 965.5. There are 23,660 shares available at this bid price in limit orders in the market. If a seller places a market order for 50,000 shares, the highest

Ticker	bbp1	bbp2	bbp3	bbp4	bbp5	Time
nts	965.5	961	960.5	955	954.5	950
Ticker	bap1	bap2	bap3	bap4	bap5	Time
nts	980	980.5	982	983	984	990.5

Table 5: Sample Price Data

bid quotes with price 965.5 will be used up so the seller would have to move on to the 20,000 shares available at the second highest bid price of 961. This process is repeated until the market order is filled. The actual stock price per share s_i for the entire trade is

$$\begin{aligned} s_{50000} &= \frac{23660 * 965.5 + 20000 * 961 + (50000 - 23660 - 20000) * 960.5}{50000} \\ &= 963.066 \end{aligned}$$

After this trade occurs the bid quotes information in the order book is updated to

Ticker	bbs1	bbs2	bbs3	bbs4	bbs5	Time
nts	23660	20000	20405	7900	20404	162455
Ticker	bas1	bas2	bas3	bas4	bas5	Time
nts	50000	14065	32100	7900	0	19450

Table 6: Updated Sample Size Data

Since this trade was initiated by a market agent who placed a market

Ticker	bbp1	bbp2	bbp3	bbp4	bbp5	Time
nts	963.066	960.5	955	954.5	0	162455
Ticker	bap1	bap2	bap3	bap4	bap5	Time
nts	980	980.5	982	983	984	19450

Table 7: Updated Sample Price Data

order to sell shares of the stock, this type of trade is called a seller-initiated trade.

3.3 Model Description

In our model, t represents the current time, and τ represents how far back in time we look at historical data.. For any stock S, the model goes back for a period of time τ starting from the t^{th} day. Information from the size matrix Q and the price matrix P is used to evaluate historical volatility σ using

$$\sigma_{adn} = \frac{\sum (R_t - \overline{R_t})^2}{n-1}. \tag{15}$$

where R_t represents the daily return given by

$$R_t = \frac{s_t - s_{t-1}}{s_{t-1}}. \tag{16}$$

For example, if a dealer picks the 60^{th} day after an original start date t, our model computes the average volatility σ over the 30 business days[1] prior to day t. The most recent update of best bids and best asks will be used

[1]Business days here represent the days when trading behavior happens, excluding non-trade days such as weekends and holidays.

as the limit order book for that business day. The mid-point of the bid-ask spread is the daily stock price. As discussed in the previous section, when a market order for Q shares comes into the market a corresponding stock price per share s_Q can be calculated using (14).

4 Sensitivity Analysis

As a sensitivity parameter γ can be adjusted to amplify or to decrease the impact triggered by liquidity. Three figures of a stock reservation prices and mid-prices over the same time interval with different γ are provided. From Figures 1, 2, and 3 we can see that even a small change in γ results in significant a jump in reservation price. Our goal is to find an optimal γ which could better calibrate Marco Avellaneda & Sasha Stoikov's [1] model to the limit order book data.

As defined in [1] the value function is

$$v(x, s, q, t) = -e^{-\gamma x} e^{-\gamma q s} e^{\frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}}. \tag{17}$$

We start with the case $q = 0$. This gives us

$$v(x, s, 0, t) = E_t[-e^{-\gamma x}] = -e^{-\gamma x} \tag{18}$$

We then take the first and second derivatives of the value function to get for $x > 0$

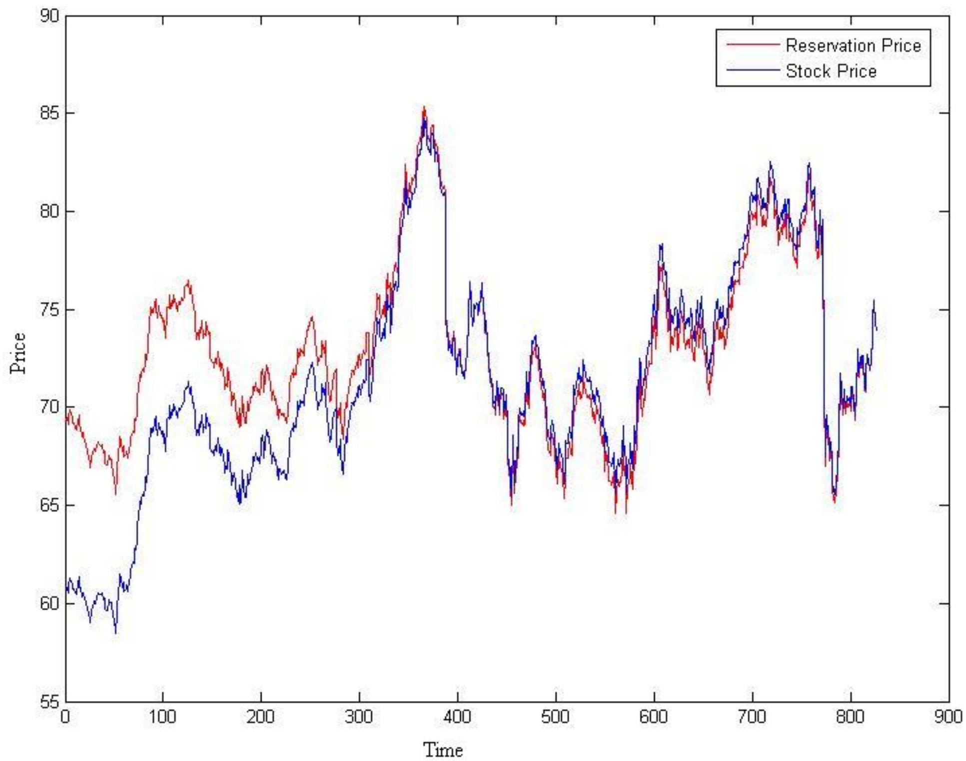

Figure 1: Indifference prices on sample stock price data when $\gamma = 0.0001$

$$\frac{\partial v}{\partial \gamma} = xe^{-\gamma x} > 0 \tag{19}$$

and

$$\frac{\partial^2 v}{\partial \gamma^2} = -x^2 e^{-\gamma x} < 0 \tag{20}$$

The first derivative of (18) with respect to γ is positive, and we see that the second derivative with respect to γ is negative regardless of x. This indicates that the value function curve is increasing and is concave-down in γ. As for

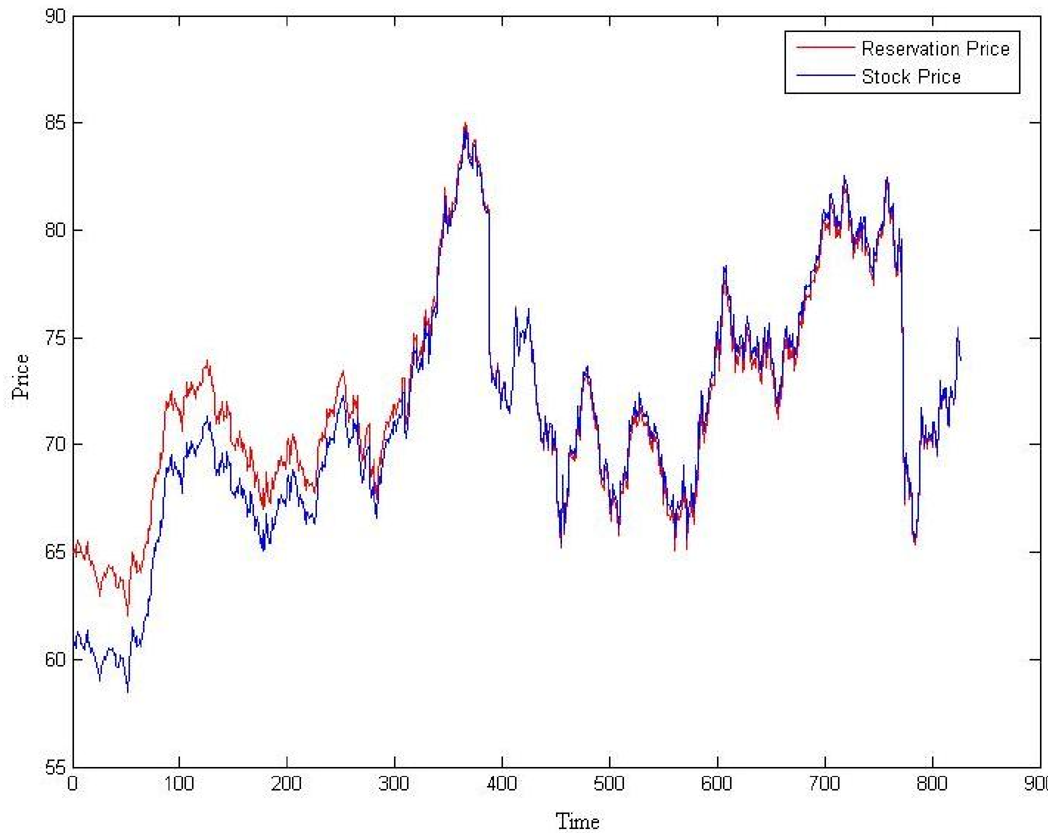

Figure 2: Indifference prices on sample stock price data when $\gamma = 0.00005$

the first derivative, if $x > 0$, then the function v is increasing; if $x < 0$, then v is decreasing. This means that v exhibits a diminishing sensitivity to γ as γ increases. Next we examine the sensitivity of v to γ in the general case $q \neq 0$.

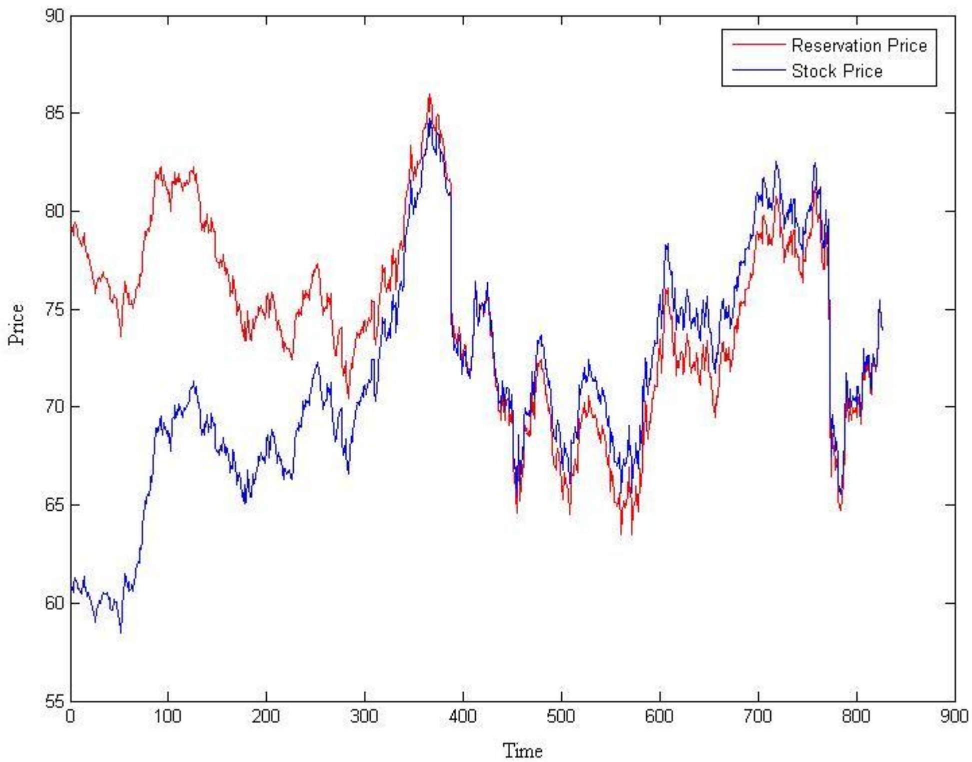

Figure 3: Indifference prices on sample stock price data when $\gamma = 0.005$

$$
\begin{aligned}
\frac{\partial v}{\partial \gamma} &= \frac{\partial v}{\partial \gamma} - e^{-\gamma x} e^{-\gamma q s} e^{\frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}} \\
&= \frac{\partial v}{\partial \gamma} - e^{-\gamma (x+qs)} e^{\frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}} \\
&= -(-(x+qs) e^{-\gamma (x+qs)} e^{\frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}} + \gamma q^2 \sigma^2 (T-t) e^{-\gamma (x+qs)} e^{\frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}}) \\
&= -e^{-\gamma x} e^{-\gamma q s} e^{\frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}} ((\gamma q^2 \sigma^2 (T-t) - x - sq)
\end{aligned}
$$

Plugging in the value function (21)

$$v(x,s,q,t) = -e^{-\gamma x} e^{-\gamma q s} e^{\frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}}. \tag{21}$$

we get the first derivative of v with respect to γ

$$\frac{\partial v}{\partial \gamma} = v(x,s,q,t)(\gamma q^2 \sigma^2 (T-t) - x - sq) \tag{22}$$

Consider the following binomial with respect to q

$$\gamma \sigma^2 (T-t) q^2 - sq - x. \tag{23}$$

From (21) we see that all three terms $e^{-\gamma x}$, $e^{-\gamma q s}$, $e^{\frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}}$ in v are positive since they have exponential form. The value function is negative due to the negative sign on the right-hand side of equation (21). If we want (22) to be positive then we need to set (23) to be negative.

When using (21), changes in γ can lead to a dramatic impact on the computation of reservation prices. We can use the limit order book data provided by Morgan Staley to calibrate the model to the market by fitting the data over γ as an independent variable.

5 Estimation of γ Using A Least-Squares Approach

When an agent comes to the market and initiates trades through market orders the market is affected. Trades waiting to happen in the limit order book might occur at a price that is higher or lower than the mid-price, which is considered to be the mid-market price (or stock price) in this paper. Therefore when the current best bids/asks are exhausted, the dealers would have to continue on buying/selling at a lower/higher price than the first best bids/asks [3]. Accordingly, the stock price might increase or decrease in response to this impact. In order to choose a meaningful value for the sensitivity parameter γ, we want to choose a γ so that transactions will not result in a drastic jump on the reservation prices chart.

The best ask prices and best bid prices regularly update to form the limit order book as time goes on. The order book data is built on the order of seconds. To simplify the data mining we rebuild the data on a daily basis. The terminal time T is the end of the last business day. We then use the limit order book to compute a sequence of reservation prices r_t with fixed timestamp t, which is t^{th} business day, and terminal time T using

$$r_t = s_t - q\gamma\sigma^2(T-t) \tag{24}$$

where q represents the number of shares traded, and it is an independent variable. A trade is initiated by a seller or buyer through a market order. As discussed before, for σ we use historical volatility. In this case s_t is stock price at time t which we assumed to be the midpoint of the bid-ask spread. For any fixed time t, we have a number of best bids and best asks. If we plot all quotes' prices in one chart over the inventory q, they result in a regression line is shown in Figure 4.

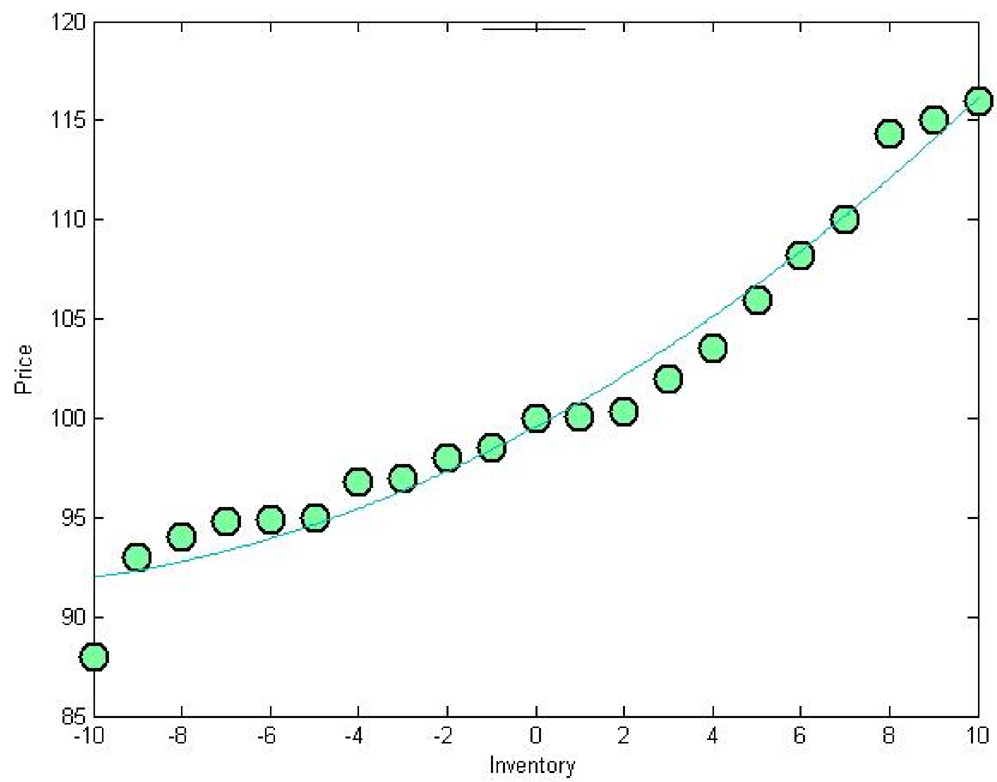

Figure 4: Regression

To find the best fit for the order book data to the Avellaneda & Stoikov [1] model , we use a least-squares fit to find γ^*. Our objective is to solve the following optimization problem,

$$\min_{\gamma} \sum (s_i - r_q(q_i))^2 \tag{25}$$

where s_i is the stock price per share at a given inventory q_i. The summation is over all data pairs (s_i, q_i) in the limit order book.

Our objective becomes

$$\min_{\gamma} \sum (s_i - r_q(q_i))^2 = \min_{\gamma} \sum (s_i - s_t + q_i \gamma \sigma (T-t))^2. \tag{26}$$

This corresponds to

$$\begin{aligned}
\min_{\gamma} \sum (s_i - r_q(q_i))^2 &= \min_{\gamma} \sum ((s_i - st + q_i \gamma \sigma (T-t))^2) \\
&= \min_{\gamma} \sum ((s_i - s_t)^2 + (q_i \sigma (T-t))^2 \gamma^2 - 2(s_t - s_i) q_i \sigma (T-t) \gamma) \\
&= \min_{\gamma} \sum ((q_i) \sigma * (T-t)^2 \gamma^2 - 2(s_t - s_i) q_i \sigma (T-t) \gamma + (s_i - s_t)^2)
\end{aligned}$$

Since this is a quadratic least-squares problem, the first order conditions are necessary and sufficient for a minimizer [17]. Our analysis gives the optimal value of γ as

$$\gamma_* = \frac{\sum (s_t - s_i) q_i \sigma (T-t)}{\sum (q_i \sigma (T-t))^2} \tag{27}$$

6 Numerical Results

We tested our model using the data provided by Morgan Stanley [9]. For each row, the value in the first column is the timestamp t. The other columns contain the top 10 bids or top 10 asks. If there are less than 10 quotes updated at time t then the rest of the column entries are filled with zeros. The order book data contains all updates on a basis of seconds. As in reality stock trading is concluded everyday, our model uses the data on a daily basis. As mentioned before we take the mid-points of the closing updates (i.e. the bid-ask spread) at the end of each day to be our daily stock prices.

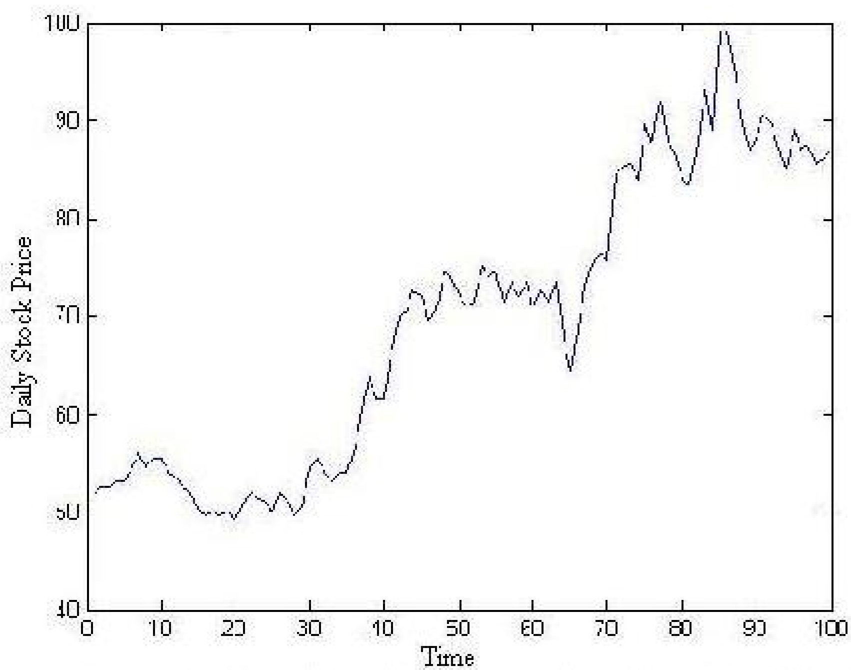

Figure 5: Mid-point stock price over 100 days

We run the model for the stock with ticker "adn", on the ${t_i}^{th}$ business day after 07/01/2003, and we look backwards for 100 days to get the historical stock price chart, shown in Figure 5. Then based on this information we can compute the historical volatility of "adn" using the daily data.

By separating the price matrix P into two matrices P_a and P_b with positive ask updates and negative bid updates separately, and applying the algorithm discussed in the previous section, we can calculate the stock price per share for a market order. If $q > 0$ then we should use the ask matrix P_a, P_b otherwise. Tables 8 and 9 give stock prices per share calculated for a set of different order sizes q.

Stock	qa1	qa2	qa3	qa4	qa5
adn	250000	375000	425000	525000	0
Stock	pa1	pa2	pa3	pa4	pa5
adn	83.0000	83.3333	83.5294	84.0000	0

Table 8: Price per share for asks at different order size q

Stock	qb1	qb2	qb3	qb4	qb5
adn	-75000	-125000	0	0	0
stock	pb1	pb2	pb3	pb4	pb5
adn	79.0000	78.6000	0	0	0

Table 9: Price per share for bids at different order size q

After obtaining stock prices s_t, limit order book data (s_i, q_i), and historical volatilities σ, an optimal γ^* can be computed using (27). We apply the Avellaneda & Stoikov [1] model with these results for stock "adn" to obtain the simulation depicted in Figure 7.

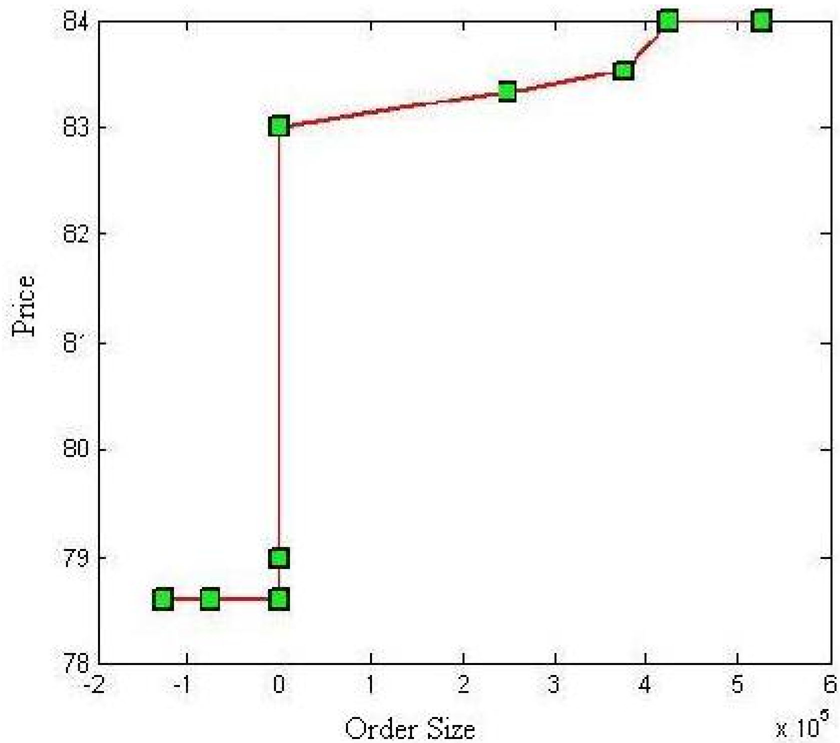

Figure 6: Price per share at different order size chart

The first day of the data set is 07/01/2003. We choose the 120^{th} business day after 07/01/2003 and set the range for historical data to be 100 days backwards. This is the timeframe used in our computations for each stock. Then we run the estimation using 200 stocks. Figure 8 and Figure 9 depict the distribution of the optimal γ^* values, while Figure 10 and figure 11 depict the distribution of the optimal γ^* without the outliers.

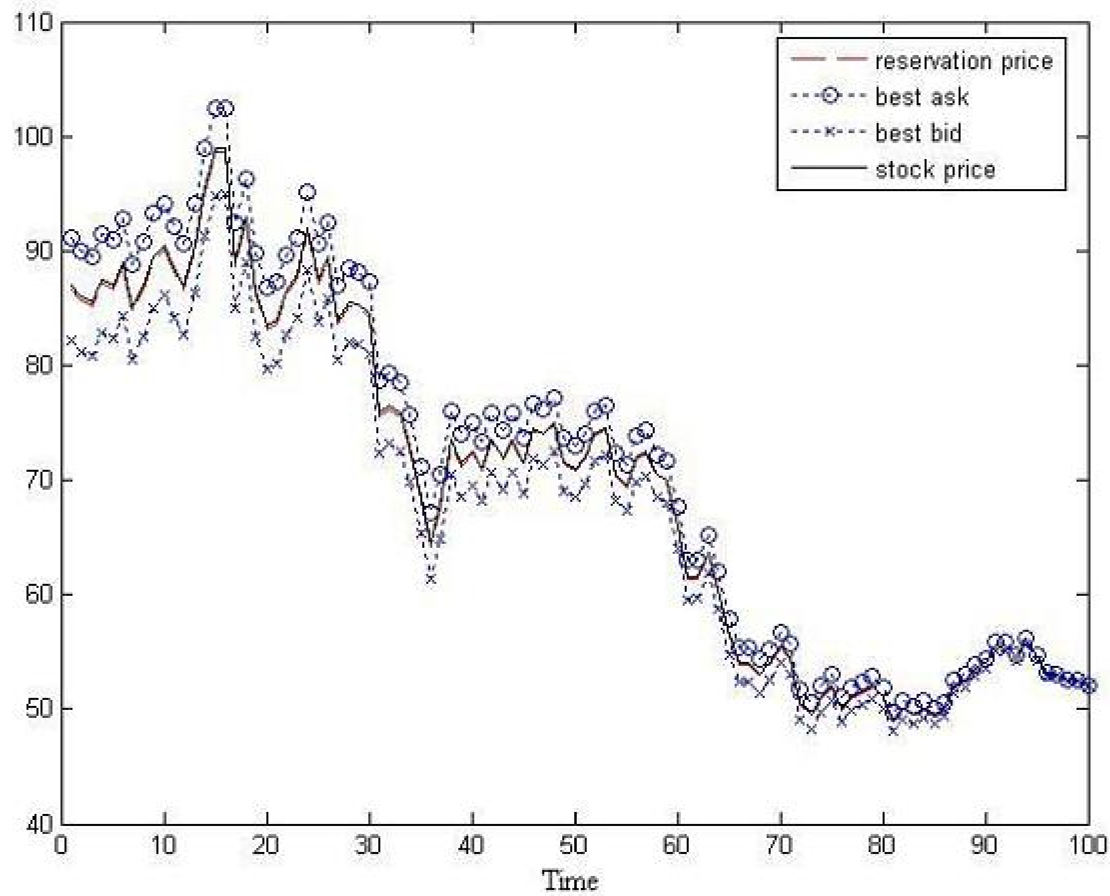

Figure 7: Numerical Experiment Plot

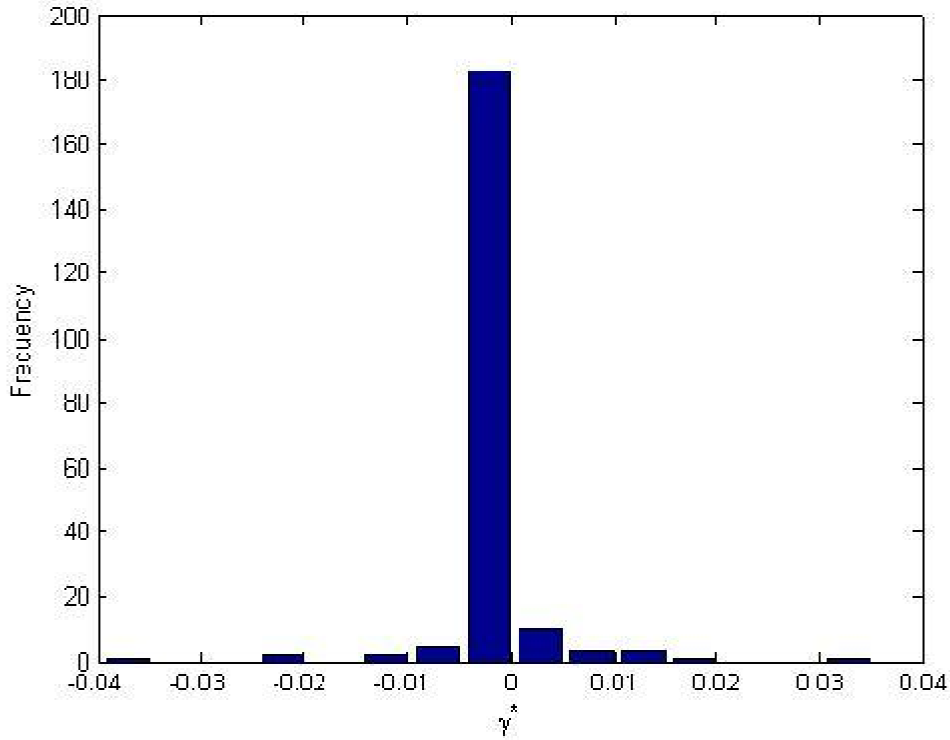

Figure 8: Histogram of γ^* with outliers

Statistics for γ^* in table 10.

γ^*	min	max	mean	median	std	range
	-0.0132	0.0178	0.0023	0.0023	0.009903	0.031

Table 10: Statistics for γ^*

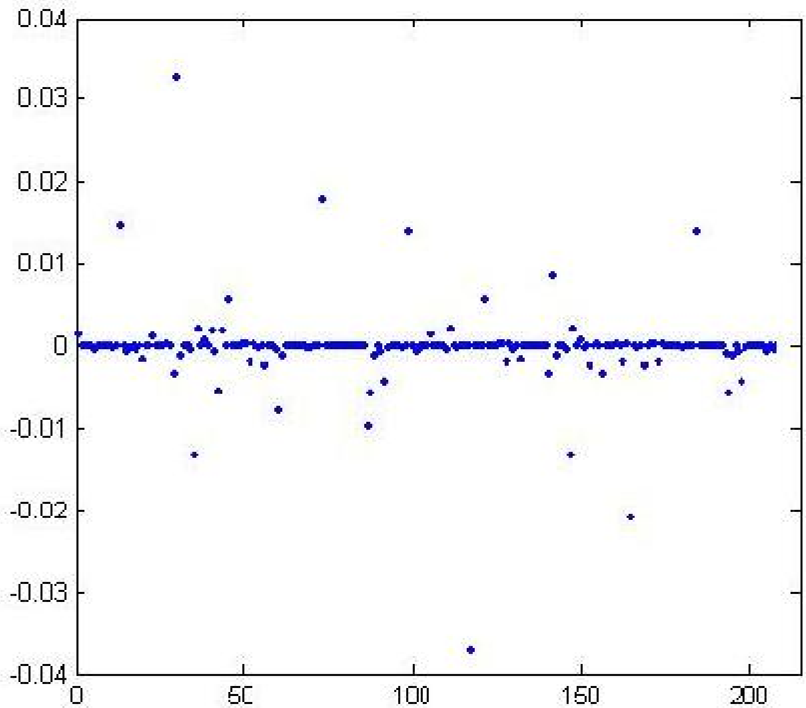

Figure 9: Scatter plot of γ^* with outliers

7 Conclusion

This paper concentrates on liquidity modeling with order book data based on the theory introduced by Avellaneda & Stoikov [1]. We utilized a least-squares approach to calibrate the model given by Avellaneda & Stoikov [1] to limit order book data through a sensitivity parameter γ. Our results give a method for choosing the values for γ when implementing the Avellaneda & Stoikov [1] model.

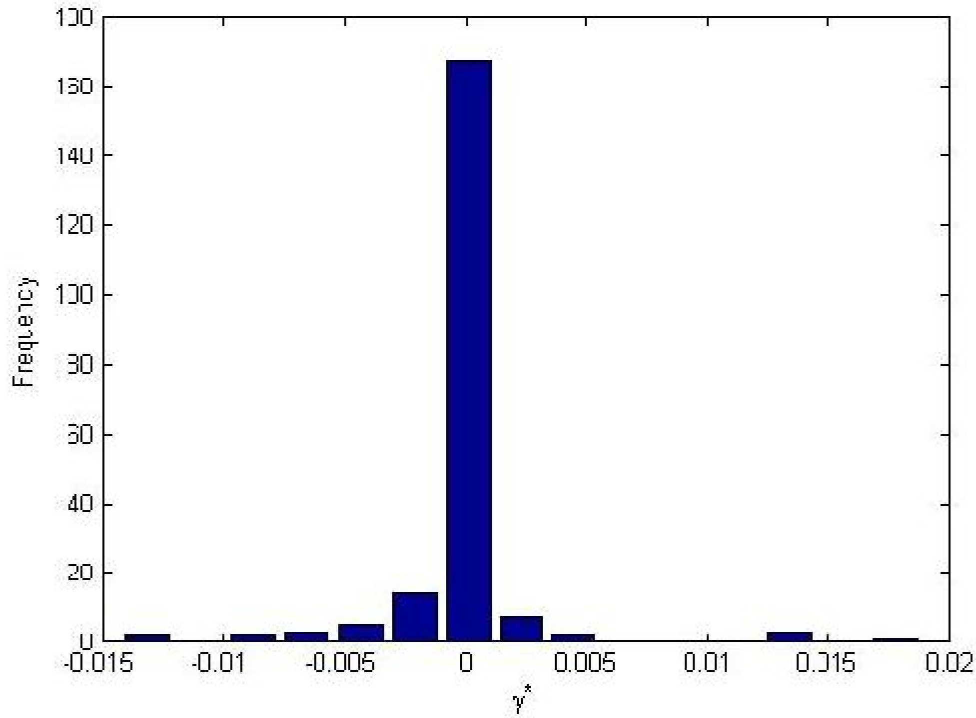

Figure 10: Histogram of γ^* without outliers

A Appendix

A.1 Order Book Data

The Morgan Stanley Data set includes data for 2184 commodities that trade on several different exchanges, including NYSE, AMEX, the London Stock Exchange, and Nasdaq. The data set includes standard tick data and order book data for July 1, 2003 - December 12,2003. In total the data set is approximately 12 gigabytes[3].

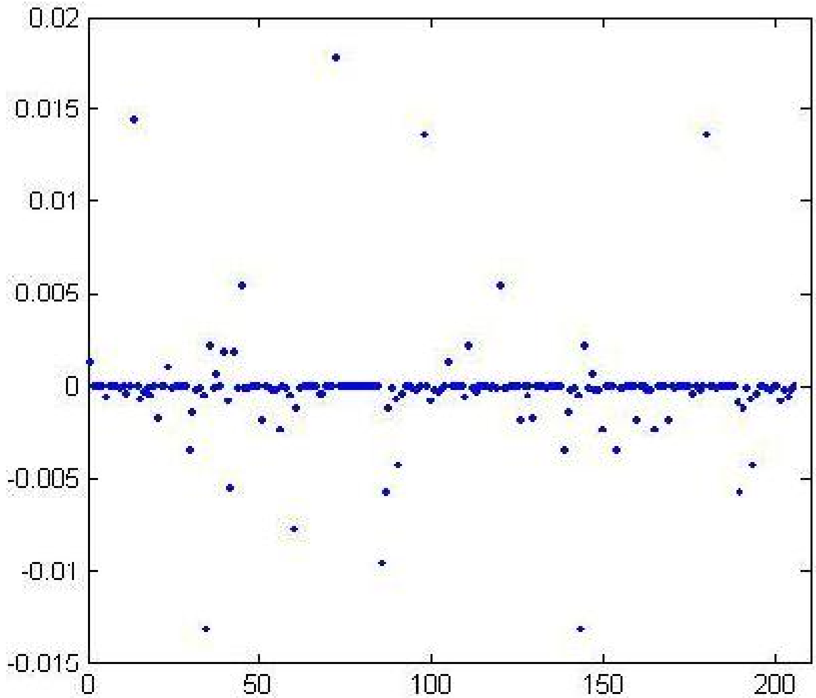

Figure 11: Scatter plot of γ^* without outliers

A.2 Matlab Code

The following function computes the timestamp on daily bases,

```
function [t,n,day] = time()
q=load('orderbookdatasizematrix.dat');     % order size
p=load('orderbookdatapricematrix.dat');    % order price
[m,l]=size(q);                             % m: days; l: shares
timestamp=zeros(m,1);
n=0;                                       % index: number of days
t=zeros();                                 % time point
```

```
day=zeros();                                     % business days matrix
for k=1:m-1
    timestamp(k)=q(k+1)-q(k);       % timestamp, jump indicates a day
    if (timestamp(k)>(1.0e+004)*3)  % sort out jumps
        n=n+1;               % number of days
        t(n)=k-3;           % go back for about 35 sec
    end
end
for i=1:n
    day(i,1)=i;
end
```

Computing for stock price per share

```
function [s_a,s_b,sa_i,sb_i,qa,qb,a1,a2,b1,b2]=share(t)
q=load('orderbookdatasizematrix.dat');    % order size
p=load('orderbookdatapricematrix.dat');   % order price
[m0,n]=size(q);pa=[];pb=[];qa=[];qb=[];m=t;
for i=1:m
    for j=2:n
        if (q(i,2)>0)
            pa1(i,1:n-1)=p(i,2:n);
            qa1(i,1:n-1)=q(i,2:n);
        elseif (q(i,2)<0)
            pb1(i,1:n-1)=p(i,2:n);
```

```
            qb1(i,1:n-1)=q(i,2:n);
        end
    end
end
pa=pa1(find(sum(pa1')),:);
qa=qa1(find(sum(qa1')),:);
pb=pb1(find(sum(pb1')),:);
qb=qb1(find(sum(qb1')),:);
[a1,a2]=size(pa);[b1,b2]=size(pb);
% s_a is the matrix of stock price per share for ask quotes, given qa as
% inventory, s_b is similarly defined.
for j=1:a2
    s_a(j)=(pa(a1,1:j)*qa(a1,1:j)')/sum(qa(a1,1:j));
end
for j=1:b2
    s_b(j)=(pb(b1,1:j)*qb(b1,1:j)')/sum(qb(b1,1:j));
end
% last update of asks and bids by the end of the day
for j=1:a2
    s_a(j)=(pa(a1,1:j)*qa(a1,1:j)')/sum(qa(a1,1:j));
    if qa(a1,j)~=0
        ai(j)=1;
    elseif qa(a1,j)==0
```

```
        ai(j)=0;
    end
    sa_i(j)=s_a(j)*ai(j);
end
for j=1:b2
    s_b(j)=(pb(b1,1:j)*qb(b1,1:j)')/sum(qb(b1,1:j));
    if qb(b1,j)~=0
        bi(j)=1;
    elseif qb(b1,j)==0
        bi(j)=0;
    end
    sb_i(j)=s_b(j)*bi(j);
end
```

This function gives us the historical data within selected range, evaluated with bid-ask spread

```
function [stock] = stock(t)
%% with the closing moment of each day, use the most recent updates to
%% compute for stock prices (mid-points), the last entry of stock matrix is
%% considered the stock price for the day.
q=load('orderbookdatasizematrix.dat');    % order size
p=load('orderbookdatapricematrix.dat');   % order price
[m0,n]=size(q);m=t;
pa=zeros();              % ask prices
```

```
pb=zeros();               % bid prices
stock=zeros();            % stock prices
for i=1:m                    % choose the most recent asks/bids
    if (q(i,2)<0)            % order size negative indicating bids
        pb=p(i,2:11);        % update bid prices at time i
        pa=pa;               % hold ask prices
    elseif (q(i,2)>0)        % order size positive indicating asks
        pa=p(i,2:11);        % update ask prices at time i
        pb=pb;               % hold bid prices
    end
    stock(i)=(pa(1)+pb(1))/2;
    if (i==1)                % eliminate 0 point, i.e. initial stock price
        stock(i)=pa(1)+pb(1);
    end
end
```

The following code takes previous σ, historical stock prices s_t, the most recent update ask price per share s_a and bid price per share s_b, then evaluates sensitivity parameter γ.

```
function [sigma,s_t,stp,s_a,s_b,gamma] = st(I,P)
% I-user's pick of starting day; P-user's pick of the
length of business days, i.e. how many days user wants to go back
[t,n,day] = testtime();
stp=0;
```

```
for i=1:n
   [stock] = teststock(t(i));
   [m1,m2] = size(stock);
   s_t1(i) = stock(m1,m2);
end
for j=1:P
       dret(j)=(s_t1(day(I-P+j))-s_t1(day(I-P+j-1)))/s_t1(day(I-P+j-1));
       s_t(j)=s_t1(day(I-P+j-1));
end
[n1,n2] = size(s_t);
stp = s_t(n1,n2);          % stock price on day I
sigma=std(dret);
[s_a,s_b,sa_i,sb_i,qa,qb,a1,a2,b1,b2]=testshare(t(I));
% compute for gamma
gamma=0;
for l=1:a2
    gamma0a(l)=((stp-s_a(l))*qa(a1,l)*sigma*(n-T));
    gamma1a(l)=(qa(a1,l)*sigma*(n-T));
end
gamma1a2=0;
gamma1a2=gamma1a*gamma1a';
for l=1:b2
    gamma0b(l)=((stp-s_b(l))*qb(b1,l)*sigma*(n-T));
```

```
    gamma1b(l)=(qb(b1,l)*sigma*(n-T));
end
gamma1b2=0;
gamma1b2=gamma1b*gamma1b';
gamma=(sum(gamma0a)+sum(gamma0b))/(gamma1a2+gamma1b2);
```

References

[1] Avellaneda, Marco, and Stoikov, Sasha, 2006, "High-Frequency Trading in A Limit Order Book", *Quantitative Finance*, Vol. 8, No. 3, 2008, 217-224.

[2] Baum, D., 2004, Seminar talk in the Financial Engineering Seminar at Cornell University, September 3, 2004.

[3] Blais, Y. Marcel, 2005, "Liquidity And Modeling The Stochastic Supply Curve For A Stock Price", *Ph. D. Dissertation*, Cornell University, August, 2005.

[4] Bjork, Tomas, 2004, "Arbitrage Theory in Continuous Time", second edition, Oxford Finance, 2004.

[5] Çetin, U., R. Jarrow, and P. Protter, 2004, "Liquidity Risk and Arbitrage Pricing Theory", *Finance and Stochastics*, 8, 311-341.

[6] Çetin, U., R. Jarrow, P. Protter, and M. Warachka, 2004, "Pricing Options in an Extended Black Scholes Economy with Illiquidity: Theory and Empirical Evidence", To appear in *Review of Financial Studies*, 2005.

[7] Ellis, K., R. Michaely, and M. O'Hara "The Accuracy of Trade Classification Rules: Evidence from Nasdaq" *The Journal of Financial and Quantitative Analysis* Vol 35, No. 4, 2000, 529–551.

[8] Engel, C., 1992, "The Risk Premium and the Liquidity Premium in Foreign Exchange Markets", *International Economic Review*, Vol. 33, No. 4, 871-879.

[9] Ferstenberg, R. and H. Alexander *Morgan Stanley Tick Data*, July 01 2003 - Dec. 19 2003.

[10] Follmer, Hans, and Schied, Alexander, 2008, "Convex and Coherent Risk Measures", October 8, 2008.

[11] Gopikrishnan, Parameswaran , Plerou, Vasiliki, Gabaix, Xavier, and Stanley, H. Eugene, , 2003, "A theory of power-law distributions in financial market fluctuations" *Letters to Nature* Vol. 423, 15, May, 2003, 267–270.

[12] Ho, Thomas, and Stoll, Hans R. 1980, "On Dealer Markets Under Competition", *The Journal of Finance*, Vol. 35, No. 2, May, 1980, 259-267.

[13] Keynes, J.M., 1930, *A Treatise on Money, Vol. II*, Macmillan, London.

[14] Longstaff, F.A., 2001, "Optimal Portfolio Choice and the Valuation of Illiquid Securities" *Review of Financial Studies*, Oxford University Press for Financial Studies, Vol. 12, No. 2, 407-431.

[15] Maslov, Sergei, and Mills, Mark, "Price Fluctuations From The Order Book Perspective–Empirical Facts and A Simple Model" *PHYSICA A*, Vol. 299, 2001, 234-246.

[16] Pennings, Joost M. E., and Smidts, Ale, 2003, "The Shape of Utility Functions and Organizational Behavior" *Management Science*, Vol. 49, No. 9, 2003, pp. 1251-1263.

[17] Peressini, A. L., Sullivan, F. E., and Uhl, Jr., J. J., 1988, "The Mathematics of Nonlinear Programming", Springer, 1998.

[18] Stoll, Hans R., "Market Microstructure" working paper, 2002.

[19] Shreve, Steven E., 2004, "Stochastic Calculus for Finance II: Contunuous-Time Models", Springer, 2004.

[20] Weber, Phillip and Rosenow, Bernd, “Order Book Approach to Price Impact” Germany, 2003.