

# Stochastic Models of Market Microstructure

Vidyadhar G. Kulkarni  
Department of Statistics and Operations Research  
University of North Carolina  
Chapel Hill, NC 27599-3260

January 26, 2012

# Contents

<b>1</b>	<b>Terminology of Financial Markets</b>	<b>4</b>
<b>2</b>	<b>Preliminaries from Microeconomics</b>	<b>8</b>
2.1	Utility Theory . . . . .	8
2.2	Partial Equilibrium Model . . . . .	9
2.3	General Equilibrium Model . . . . .	10
2.4	Disequilibrium Models . . . . .	13
2.4.1	Discrete Time Model . . . . .	14
2.4.2	Continuous Time Model . . . . .	14
<b>3</b>	<b>Preliminaries from Stochastic processes</b>	<b>16</b>
3.1	Univariate Time Series . . . . .	16
3.1.1	Definitions . . . . .	16
3.1.2	Common Time Series Models . . . . .	17
3.1.3	Properties of ACVF . . . . .	18
3.1.4	Lag Operator . . . . .	18
3.1.5	Estimation and Prediction . . . . .	19
3.2	Multivariate Time Series . . . . .	19
3.2.1	Estimation and Prediction . . . . .	20
3.3	Continuous Time Markov Chains . . . . .	21
3.4	Semi-Markov Decision Processes . . . . .	21
<b>4</b>	<b>Call Markets</b>	<b>22</b>

4.1	Market Clearing Price Model . . . . .	23
4.2	Utility Based Model. . . . .	24
4.3	Market Maker's Role . . . . .	27
4.4	Strategic Customers: Grossman Model . . . . .	28
4.5	Strategic Customers: Kyle Model . . . . .	30
4.6	A Real Call Market . . . . .	30
4.7	Model of a Real Call Market . . . . .	31
4.8	Game Theoretic Models of Call Markets . . . . .	35
4.9	Stochastic Characteristics of a Call Market . . . . .	37
<b>5</b>	<b>Quote-Driven Markets: Dealers as Inventory Managers</b>	<b>38</b>
5.1	A Static Model. . . . .	39
5.2	A Dynamic Model . . . . .	41
<b>6</b>	<b>Quote-Driven Markets: Dealers as Utility Maximizers</b>	<b>45</b>
6.1	Single Period Model . . . . .	45
6.2	Multiple Period Model . . . . .	47
<b>7</b>	<b>Quote-Driven Markets: Transaction Costs</b>	<b>52</b>
7.1	Roll Model . . . . .	52
7.2	Roll Model: Correlated Transactions . . . . .	53
7.3	Roll Model: Trade Influence . . . . .	54
7.4	Roll Model: Further Extensions . . . . .	55
7.5	Multivariate Roll Model . . . . .	56
<b>8</b>	<b>Quote-Driven Markets: Informed Traders</b>	<b>58</b>
8.1	Single-Trade Copeland-Galai Model . . . . .	58
8.2	Sequential Glosten-Milgrom Model . . . . .	60
<b>9</b>	<b>Order-Driven markets: Dynamic Programming Models</b>	<b>64</b>
9.1	Single Trader, Single Period . . . . .	64

9.2	Single Trader, Multiple Periods . . . . .	66
9.3	Parlour Model . . . . .	67
9.4	Foucault Model . . . . .	70
<b>10</b>	<b>Order-Driven Markets: Simple Queueing Models</b>	<b>73</b>
10.1	The Garman Model . . . . .	73
10.2	Special Case: A Clearing Market . . . . .	75
10.3	Extensions to the Garman Model . . . . .	77
10.4	Luckock Model . . . . .	78
10.4.1	Separable Markets. . . . .	80
10.4.2	Constant Elasticities . . . . .	80
<b>11</b>	<b>One-Sided Order Books: Strategic Traders</b>	<b>82</b>
11.1	Closed Sell Side Book: No information . . . . .	82
11.2	Closed Sell Side Book: Partial Information . . . . .	85
11.3	Open Sell Side Book: Seppi Model . . . . .	85
11.4	Open Sell Side Book: Rosu Model . . . . .	87
<b>12</b>	<b>Optimal Execution Strategies</b>	<b>91</b>
12.1	Optimal Splitting: Bertsimas Model . . . . .	91
12.1.1	Additive Permanent Price Impact . . . . .	92
12.1.2	Additive Permanent Price Impact with Information . . . . .	93
12.1.3	Multiplicative Temporary Price Impact . . . . .	94
12.2	Optimal Splitting and Timing: OW model . . . . .	95
12.2.1	Special Case: Discrete Trading Times . . . . .	96
12.3	Optimal Order Placements: Harris Model . . . . .	98

# Chapter 1

## Terminology of Financial Markets

(The material in chapter is based on the first six chapters of Harris [21].)

**Financial Markets** offer a mechanism to trade financial instruments called **securities**: stocks, bonds, options, futures, swaps, foreign exchange, commodities, etc. The actual trade may occur in a physical place (called **trading floors**): NYSE, Chicago Merc, or on an **electronic exchange**: Nasdaq, Euronext,; or over the counter: Goldman Sachs. See

[http://www.tdd.lt/slnews/Stock\\_Exchanges/Stock.Exchanges.htm](http://www.tdd.lt/slnews/Stock_Exchanges/Stock.Exchanges.htm)

for a list of world's financial markets. (Not all the links in it work though!)

The main trading problem is matching compatible buyers and sellers. Together they are called **traders**. The motivation behind trading can be investing, borrowing, income generation via day trading, gambling. Examples: individuals, money market fund managers, institutions, brokers, dealers, etc.

**Brokers** match buyers and sellers, and make money by charging commissions. They do not own any inventory themselves. These may be physical locations or electronic programs. Examples: Fidelity, Merrill Lynch, Dreyfuss, Ameritrade, eTrade, etc.

**Dealers** buy from the sellers (at the bid price) and sell to the buyers (at the ask price) from the inventory that they own. They make money by setting the ask price slightly higher than the bid price. The difference ask-bid is what is called the bid-ask spread. Dealers exist for each security on each exchange.

Every security listed on an exchange has a designated dealer on that exchange. This dealer is called the **market maker** or the **specialist**. His job is to provide liquidity for the market in that security, i.e., be available to buy or sell to other traders.

Traders go through brokers and brokers go through dealers to trade securities on exchanges. Every trade is eventually **settled** by the specialist or market maker in that security.

**Broker-Dealers** are hybrid of brokers and dealers. Examples: Merrill Lynch, Goldman Sachs, Morgan Stanley, Credit-Suisse, etc.

Why do traders trade? Dealers and brokers trade because it is their business to make money from trading. Liquidity traders sell securities to move income from the future (the dividends, for example) to the present, and buy to move funds from the present to the future. Informed traders trade because they have inside information, or superior knowledge about the future of a security than others, and so hope to increase their wealth at the cost of those who do not have this information. Some trade for risk management purposes, as through derivative markets, or for portfolio balancing requirements. Some may trade just to gamble or speculate! All traders trade by placing an order in the market.

An **order** is a concise statement of intent from the trader about a trade. It also serves as an instruction for the broker, and is visible to all brokers and dealers dealing with that security. It specifies what instrument to trade, how much to trade, whether to buy or sell, at what price. An order that has been submitted but not yet executed is called a **standing orders**.

When traders want to buy, they specify the buying price called **bid**. When traders want to sell they specify the selling price called **ask** or **offer**. There may be many standing orders for a security. The highest bid price among the is called the **market bid**, and the lowest ask price or offer price is called the **market ask** or **market offer**.

A **market order** is an order to execute a trade at the market price. A **limit order** is to trade at a price no worse than the **limit price**. A limit buy order with a limit bid price of 100 can execute when the price decreases to 100 or below. It is submitted when the current market price is above 100. A limit sell order has a limit sell price of 100 and can execute at that price or higher. It is submitted when the current market ask is less than 100.

A **stop order** is a standing order with a stop price. A stop sell order with stop price of 100 is executed as soon as the the market ask price declines to 100. It is submitted when the current the

current ask price is above 100. Hence such orders are also called **stop loss** orders.

The orders are executed during **trading sessions**. A **continuous trading session** allows the orders to be executed any time the market is open. Most current markets have continuous trading sessions. On the other hand, in **call markets** all standing orders are executed at a fixed time one or more times during a day. Call markets are used for the initial auction of government bonds, T-notes, etc.

Many continuous markets open with an opening call when they clear all the accumulated orders that arrive while the market is closed. They then switch over to continuous trading.

The orders are maintained in an **order book** and executed by using an **execution system**. There are three main execution systems: quote-driven, order-driven or brokered.

In a **quote-driven** execution system a dealer provides a quoted ask and bid price, which the dealer changes continuously in order to match supply and demand. The order book is maintained and visible only to the dealer. The traders only see the ask and bid price (and sometimes the most recent trade). All the orders must execute at this price. Examples: Nasdaq (organized by a dealer association), NYSE (organized by an exchange), eTrade (organized by an electronic vendor).

In an **order-driven** execution system, traders can settle their orders among themselves, following a fixed order precedence rule. Typically the orders are ranked by price, followed by time of arrival, followed by size. The order book is visible to all registered traders. Example: most electronic exchanges, eBay.

In a **brokered** execution system, brokers match buyers and sellers through their proprietary contacts.

Through the trading sessions and order execution systems the markets provide a method of **price discovery**, that is, finding a price that will match supply and demand.

An example of a page in an order book at 12:30:00 pm on a particular day:

order number	bids	ask	size	time	comment
<i>S5</i>		33.17	400	12.07.15	
<i>S4</i>		33.17	600	12.07.15	
<i>S3</i>		33.17	200	12.05.42	
<i>S2</i>		32.98	300	12.15.24	
<i>S1</i>		32.95	800	12.08.47	Market ask
<i>B1</i>	32.93		1200	12.03.00	Market bid
<i>B2</i>	32.01		200	12.25.30	
<i>B3</i>	32.00		300	12.18.28	
<i>B4</i>	31.50		800	12.08.19	

The order execution sequence is *S1*, *S2*, *S3*, *S4*, *S5* among the sell orders, and *B1*, *B2*, *B3*, *B4* among the buy orders. Thus the current market ask is 32.95, and the market bid is 32.93. The bid-ask spread is 2 cents. Consider three possible events:

1. A buy order comes in at 12:30:00 for 1000 shares at 33.00, with partial execution allowed. This buy order will be executed immediately: all 800 shares of *S1* and 200 shares of *S2* will get sold at 32.98 if the exchange follows **uniform pricing**. If the exchange follows **discriminatory pricing** the buyer will get 800 shares at 32.95, and 200 at 32.98. The buyer benefits more from discriminatory pricing. The order book will be modified to reflect this transaction. The market ask will now change to 32.98.
2. A sell order comes in 12:30:00 for 1500 shares at 32.90, with partial execution allowed. This sell order will be partially executed immediately by *B1* at 32.93, and the remaining buy order will stay on the books at 32.90 for 300 shares. In this case the uniform pricing and the discriminatory pricing produce the same transaction price. The market ask now reduces to 32.90.
3. A market buy order comes in for 2000 shares. It is satisfied by 800 shares at 32.95 from *S1*, 300 shares at 32.98 from *S2*, 200 shares at 33.17 from *S3*, 600 shares at 33.17 from *S4* and 100 shares at 33.17 from *S5*. Market ask now moves to 33.17.



## Chapter 2

# Preliminaries from Microeconomics

(The material in chapter is based on Chapters 8, 14 and 16 of Nicholson [35].)

In this chapter we shall study some simple models of price discovery. Historically, the most puzzling question was why the most essential things in life were free or almost free (like air, water), while the least essential things were very expensive (like diamonds). The issue was settled with the introduction of the concept of marginal cost and marginal value. Air is free because the value of an additional unit of air is zero. The price is related to the marginal value, not the actual value.

### 2.1 Utility Theory

Consider the following hypothetical experiment. You have a choice between two games, the first one pays a guaranteed payoff of  $\$x$ , while the second pays one dollar with probability .5, and zero dollars with probability .5. Thus the expected payoff from the second game is .5. At what value of  $c$  of  $x$  will you be indifferent between the two games? You are

1. **risk-averse** if  $c > .5$ ,
2. **risk neutral** if  $c = .5$ , and
3. **risk-seeking** (or risk-loving) if  $c < .5$ .

$c$  is called the **certainty equivalent**, and  $c - .5$  is called the **risk premium**.

Economists like to think in terms of utility  $U(w)$  of wealth  $w$ , rather than wealth itself. How can one pick a utility function? The following simple conceptual experiment works well if we know that  $w \in [a, b]$ . We can reasonably assume that  $U$  is an increasing function of  $w$ , and set  $U(a) = 0$  and  $U(b) = 1$ . For  $a < w < b$  consider two options: under option 1 the individual gets a fixed reward  $w$

with probability 1, and under option 2 the individual gets reward  $b$  with probability  $u$ , and reward  $a$  with probability  $1 - u$ . Clearly, if  $u$  is 0, the individual prefers option 1, and if  $u = 1$ , he prefers option 2. It seems reasonable to assume that as  $u$  increases from zero to one, there will be a value of  $u$  when he is indifferent between these two options. We take this value of  $u$  as the utility  $U(w)$  of  $w$ .

In general we take utility as a given increasing function  $U : (-\infty, \infty) \rightarrow (-\infty, \infty)$ , and do not worry about how this function is derived. Utility is linear for a risk neutral person, concave for a risk averse person, and convex for a risk seeking person. A measure of **risk aversion**, introduced by Pratt [40], is defined mathematically as follows

$$r(w) = -\frac{U''(w)}{U'(w)}. \quad (2.1)$$

**Example 2.1** For  $U(w) = a + bw$ ,  $b > 0$ , we have risk aversion  $r(w) = 0$ . For  $U(w) = \ln(w)$ , we have  $r(w) = 1/w$ , a decreasing risk aversion. For  $U(w) = -e^{-\rho w}$ , we have  $r(w) = \rho$ , a constant risk aversion.  $\rho$  is called the risk aversion factor.

A related quantity called **relative risk aversion** is defined by

$$rr(w) = wr(w) = -w\frac{U''(w)}{U'(w)}. \quad (2.2)$$

The logarithmic utility has constant relative risk aversion.

## 2.2 Partial Equilibrium Model

Consider the supply and demand for a single product. Let  $p$  be the market price of this product. Let  $D(p)$  be the demand and  $S(p)$  be the supply for the product if its market price is  $p$ . There are many ways to interpret  $D(p)$  and  $S(p)$ . One can think of  $D(p)$  as the expected number of consumers (each consuming one item) who are willing to pay at least  $p$  for the product, and  $S(p)$  as the expected number of producers (each producing one item) who are willing to sell the product at price  $p$  or less. One can also think of  $D^{-1}(q)$  as the marginal utility (in dollars) if one more unit of the product is consumed, and  $S^{-1}(q)$  as the marginal cost of producing one item. In all these interpretations it seems reasonable to assume that  $D$  is a decreasing function of  $p$ , and  $S$  is an increasing function of  $p$ . Assume that they are continuous. Then the partial equilibrium model says that the market price  $p^*$  is such that

$$D(p^*) = S(p^*). \quad (2.3)$$

Another way of stating the same result is to define the excess demand function

$$ED(p) = D(p) - S(p).$$

Then Equation 2.3 can be written in an equivalent way as

$$ED(p^*) = 0.$$

This model was first proposed by Alfred Marshall (1842-1924) (*principles of Economics*, 1920, Book 5, Chapters 1,2, and 3). Under certain reasonable conditions there will be a unique price  $p^*$  that satisfies Equation 2.3. This is called the partial equilibrium model because it does not take into account the effect of other products in the market on the supply and demand for this product.

We define the consumer surplus as the total net utility derived by the consumers if they consume  $q$  units at price  $p$ . It is given by

$$\int_0^q D^{-1}(u)du - pq.$$

Similarly the producer surplus is defined as the total net utility derived by the producers if they produce and sell  $q$  units at price  $p$ . It is given by

$$pq - \int_0^q S^{-1}(u)du.$$

It can be shown that the total surplus (consumer surplus + producer surplus) is maximized at the quantity  $q^*$  satisfying

$$q^* = D(p^*) = S(p^*),$$

where  $p^*$  is from Equation 2.3.

## 2.3 General Equilibrium Model

Now consider a market for  $n$  products (including labor, capital, etc). Let  $p_i$  be the price and  $S_i$  be the supply of the  $i$ th product,  $1 \leq i \leq n$ . Let  $p = [p_1, p_2, \dots, p_n]$ . Let  $D_i(p)$  be the demand for product  $i$  if the price vector is  $p$ . A price vector  $p^*$  is called an equilibrium price if

$$D_i(p^*) = S_i, \quad 1 \leq i \leq n.$$

Will there always exist an equilibrium non-negative price  $p^*$ ? If it exists, will it be unique? This model was first studied by Walrus (1834-1919) (*Elements of Pure Economics*, translated by W. Jaffe, 1954).

A modern proof of existence is based on the following fixed point theorem:

**Theorem 2.1** *Let  $A$  be a closed, bounded, convex set, and let  $G : A \rightarrow A$  be a continuous mapping of  $A$  into  $A$ . Then there exists an  $a \in A$  such that*

$$G(a) = a.$$

Define the excess demand functions

$$ED_i(p) = D_i(p) - S_i, \quad 1 \leq i \leq n.$$

Walras made three assumptions.

1.  $ED_i(p)$  is homogeneous in  $p$ , i.e.,

$$ED_i(\lambda p) = ED_i(p), \quad 1 \leq i \leq n, \lambda > 0.$$

Thus if all prices double, demands will remain unaffected. Hence it suffices to assume that

$$\sum_{i=1}^n p_i = 1.$$

2.  $ED_i(p)$  is a continuous function of  $p$ , for all  $1 \leq i \leq n$ .
3. The total value of excess demand is zero at all  $p$ :

$$\sum_{i=1}^n p_i ED_i(p) = 0. \tag{2.4}$$

The last assumption is called Walras's Law, and is equivalent to the assumption that the model represents a closed economy, and wealth is neither created, nor destroyed. We can see an intuitive justification for this law as follows: Let the market consist of  $n$  products and  $m$  players. Suppose the price vector  $p$  is fixed. At these prices let

$D_{ij}$  = demand for product  $i$  by player  $j$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ,

$S_{ij}$  = supply of product  $i$  by player  $j$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ,

$D_i = \sum_{j=1}^m D_{ij}$  = total demand for product  $i$ ,  $1 \leq i \leq n$ ,

$S_i = \sum_{j=1}^m S_{ij}$  = total supply of product  $i$ ,  $1 \leq i \leq n$ ,

$ED_i = D_i - S_i$  = excess demand for product  $i$ ,  $1 \leq i \leq n$ .

Since the players cannot produce wealth out of nothing, they must pay for their demands from their income from selling the supply, that is,

$$\sum_{i=1}^n p_i D_{ij} = \sum_{i=1}^n p_i S_{ij}, \quad 1 \leq j \leq m.$$

Summing over all players, we get

$$\sum_{i=1}^n p_i D_i = \sum_{i=1}^n p_i S_i$$

which yields Equation 2.4.

Next we give a formal definition of the equilibrium price vector.

**Definition 2.1** A price vector  $p^* = [p_1^*, p_2^*, \dots, p_n^*]$  is called an equilibrium price vector if, for all  $1 \leq i \leq n$ ,

$$ED_i(p^*) = 0 \quad \text{if } p_i^* > 0,$$

and

$$ED_i(p^*) \leq 0, \quad \text{if } p_i^* = 0.$$

Any good with zero price is called a free good, and can have negative excess demand, because there is no way to increase the demand by lowering the price any further.

**Theorem 2.2** If assumptions 1, 2, and 3 hold, then there exists at least one equilibrium price vector.

**Proof:** Let

$$A = \{p \in R^n : p_i \geq 0, 1 \leq i \leq n, \sum_{i=1}^n p_i = 1\}.$$

Note that  $A$  is a closed, bounded and convex set. Define the mapping  $F : A \rightarrow A$  by

$$F(p) = [F_1(p), F_2(p), \dots, F_n(p)],$$

where

$$F_i(p) = \max(p_i + ED_i(p), 0), \quad 1 \leq i \leq n. \quad (2.5)$$

Thus if  $ED_i(p) > 0$ , then  $F_i(p) > p_i$ , and if  $ED_i(p) < 0$ , then  $F_i(p) < p_i$ . This is reasonable: if the excess demand for a product is positive, its price should rise; and if the excess demand for a product is negative, its price should fall.

Next we show that not all  $F_i(p)$  as defined in Equation 2.5 can be zero if  $p \in A$ . For if they were, we must have

$$p_i + ED_i(p) \leq 0, \quad 1 \leq i \leq n.$$

Now multiply the above equation by  $p_i$  and sum over all  $i$ . Then we get

$$\sum_{i=1}^n p_i^2 + \sum_{i=1}^n p_i ED_i(p) \leq 0.$$

However, the second term is zero due to Walrus's Law. Hence

$$\sum_{i=1}^n p_i^2 \leq 0.$$

But this cannot be true since  $p \in A$ . Thus at least one of the  $F_i(p)$ 's must be positive. This allows us to define

$$G_i(p) = F_i(p) / \sum_{j=1}^n F_j(p), \quad 1 \leq i \leq n.$$

Thus

$$\sum_{i=1}^n G_i(p) = 1.$$

Hence  $G(p) = [G_1(p), G_2(p), \dots, G_n(p)] \in A$ . Thus  $G$  satisfies the conditions of Theorem 2.1. Hence there exists a  $p^* \in A$  such that  $G(p^*) = p^*$ . Now let  $I = \{i : p_i^* > 0\}$ . Then, for  $i \notin I$  we have

$$F_i(p^*) / \sum_{j=1}^n F_j(p^*) = 0.$$

This implies that  $F_i(p^*) = 0$ , which implies that  $ED_i(p^*) \leq 0$ . Thus we have shown that

$$p_i^* = 0 \Rightarrow ED_i(p^*) \leq 0. \quad (2.6)$$

Next, if  $p_i^* > 0$ , we must have  $F_i(p^*) = p_i^* + ED_i(p^*) > 0$ . Hence

$$p_i^* = (p_i^* + ED_i(p^*)) / \sum_{j \in I} (p_j^* + ED_j(p^*)), \quad i \in I.$$

This implies

$$p_i^* \sum_{j \in I} (p_j^* + ED_j(p^*)) = p_i^* + ED_i(p^*), \quad i \in I.$$

Using the fact that  $\sum_{j \in I} p_j^* = 1$  and simplifying, we get

$$ED_i(p^*) = p_i^* c, \quad i \in I, \quad (2.7)$$

where

$$c = \sum_{j \in I} ED_j(p^*).$$

Multiplying Equation 2.7 by  $p_i$  and summing over all  $i \in I$  we get

$$\sum_{i \in I} p_i^* ED_i(p^*) = c \sum_{i \in I} (p_i^*)^2.$$

However, Walrus's law states that the left hand side is zero. Also the sum on the right hand side is positive. Hence  $c$  must be zero. Then Equation 2.7 implies that

$$ED_i(p^*) = 0, \quad i \in I. \quad (2.8)$$

Equations 2.6 and 2.8 imply that  $p^*$  is an equilibrium price vector. This completes the proof. ■

## 2.4 Disequilibrium Models

The above equilibrium models assume that the producers and consumers simultaneously know the supply and demand curves and instantaneously settle on the correct price. However, in practice how does this occur? To answer this question we need a dynamic model. We describe a discrete time version and a continuous time version below with the help of the one-dimensional partial equilibrium model.

### 2.4.1 Discrete Time Model

Suppose time is discrete,  $t = 0, 1, 2, \dots$ . At each time  $t$ , the suppliers have to base their production before the price  $p_t$  in period  $t$  is known. They base this on their forecast of the price at time  $t$ , namely,  $\tilde{p}_t$ . That is, they produce  $S_t = S(\tilde{p}_t)$ . Now the real price  $p_t$  is set so that the demand matches supply and the entire supply is cleared. This price is given by  $p_t = D^{-1}(S_t)$ . The system then moves to time  $t+1$ . The behavior of this model is heavily dependent on the producers' forecast  $\tilde{p}_t$ . We describe two alternatives:

**The Cobweb Model:** The producers assume that the price from the previous period will hold in this period as well, i.e.

$$\tilde{p}_t = p_{t-1}.$$

Suppose we start with an arbitrary  $p_0$ . Then  $S_1 = S(p_0)$ , and  $p_1 = D^{-1}(S_1)$ , etc. This produces the standard cobweb model. Note that  $p_t$  may not converge to the equilibrium price  $p^*$  under this model.

**Example 2.2** Assume

$$D(p) = c - dp, \quad S(p) = a + bp,$$

with  $a, b, c, d > 0$  and  $c > a$ . Then the cobweb model produces the difference equation

$$p_t = (c - a)/d - (b/d)p_{t-1}, \quad t \geq 1.$$

This can be solved to get

$$p_t = (p_0 - p^*)(-b/d)^t + p^*,$$

where

$$p^* = (c - a)/(b + d).$$

Thus  $p_t \rightarrow p^*$  if  $0 \leq b/d < 1$ .

**Rational Expectation Model:** Since the producers know the supply and the demand curve, they can compute  $p^*$ , and then use

$$\tilde{p}_t = p^*.$$

This will result in convergence to  $p^*$  in one step. The idea of using all available information to form the price forecasts is called the rational expectation method, and is used quite often in the analysis of financial models.

### 2.4.2 Continuous Time Model

There are two main classical models of how the price  $p(t)$  at time  $t$  varies with  $t$ .

**Walrasian Approach:** Let  $ED(p) = D(p) - S(p)$  be the excess demand at price  $p$ . It makes sense to assume that if  $ED(p) > 0$ , the price should increase and if it is  $< 0$ , the price should decrease. The Walrasian model incorporates this intuition in the following mathematical model:

$$\frac{dp(t)}{dt} = kED(p(t)), \quad t \geq 0, \quad (2.9)$$

for some given constant  $k > 0$ . Note that at the equilibrium price  $p^*$ , the excess demand is zero, and hence if  $p(t) = p^*$  at any time  $t = t_0$ , we will have  $p(t) = p^*$  for all  $t \geq t_0$ . In general, if  $p(0)$  is close to  $p^*$ , we can use Taylor expansion to write

$$ED(p) = ED(p^*) + ED'(p^*)(p - p^*) = ED'(p^*)(p - p^*).$$

Hence the solution to Equation 2.9 is given by

$$p(t) = (p(0) - p^*)e^{kED'(p^*)t} + p^*.$$

Thus  $p(t) \rightarrow p^*$  if  $ED'(p^*) < 0$ . This will be the case if we assume that  $ED(p)$  is a decreasing function of  $p$ . This approach to  $p^*$  is called the *tatonnement* (“groping”) process.

**Marshallian Approach:** Instead of adjusting prices in response to imbalance in supply and demand, the Marshallian approach is to adjust supply and demand in response to the price imbalance. Let  $q$  be a given quantity. The price that the consumers are willing to pay to consume this quantity is  $D^{-1}(q)$ , and the producers are willing to pay is  $S^{-1}(q)$ . Clearly, if  $D^{-1}(q) > S^{-1}(q)$ , the quantity  $q$  should be increased, or else it should be decreased. The Marshallian model incorporate this intuition in the following mathematical model:

$$\frac{dq(t)}{dt} = k(D^{-1}(q(t)) - S^{-1}(q(t))), \quad t \geq 0, \quad (2.10)$$

for some given constant  $k > 0$ .

Both these models are rather simple minded, and very few markets exist that are well described by them.



## Chapter 3

# Preliminaries from Stochastic processes

### 3.1 Univariate Time Series

(The material in chapter is based on Chapters 1, 2, and 3 of Brockwell and Davis [5].)

#### 3.1.1 Definitions

A discrete time stochastic process  $\{X_t\} = \{X_t, t = 0, \pm 1, \pm 2, \dots\}$  is called a time series. It is called strictly stationary if the joint distribution of  $(X_t, X_{t+1}, \dots, X_{t+k})$  is independent of  $t$  for all  $k$ . As a consequence, we see that if a time series is strictly stationary  $E(X_t)$  is a constant, and  $\text{Cov}(X_t, X_{t+h})$  depends only on  $h$ . We shall use the notation

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}), \quad h = 0, \pm 1, \pm 2, \dots$$

Clearly  $\gamma(h) = \gamma(-h)$  for a strictly stationary time series.

A time series is called weakly stationary, or covariance stationary, if  $E(X_t) = \mu$ , and  $\text{Cov}(X_t, X_{t+h}) = \gamma(h) < \infty$  for all  $t$ . The time series is called Gaussian if  $(X_t, X_{t+1}, \dots, X_{t+k})$  is multivariate Normal for each  $k$ . For a Gaussian time series, strict and weak stationarity are equivalent concepts, since the covariances completely describe the joint distribution of a constant mean multivariate Normal random variable.

From now on we use the term stationary to mean weakly stationary. The function  $\gamma$  is called the auto-covariance function (ACVF). Let

$$\rho(h) = \gamma(h)/\gamma(0), \quad h = 0, \pm 1, \pm 2, \dots$$

The function  $\rho$  is called the auto-correlation function.

### 3.1.2 Common Time Series Models

Let  $\{\epsilon_t\}$  be a sequence of uncorrelated random variables with mean zero and variance  $\sigma_\epsilon^2 < \infty$ . This is a stationary time series, called *white noise*. If the random variables are iid, the series is strictly stationary.

Let  $X_0 = 0$  and  $X_{t+1} = X_t + \epsilon_t$ ,  $t \geq 0$ . Then  $\{X_t\}$  is not a stationary time series, since  $\text{Var}(X_t) = t\sigma_\epsilon^2$  changes with  $t$ .

Let

$$X_t = \epsilon_t + \theta\epsilon_{t-1}, \quad t = 0, \pm 1, \pm 2, \dots$$

Then  $\{X_t\}$  is a stationary time series with ACVF

$$\gamma(h) = \begin{cases} \sigma_\epsilon^2(1 + \theta^2) & \text{if } h = 0, \\ \sigma_\epsilon^2\theta & \text{if } h = \pm 1, \\ 0 & \text{if } |h| > 1. \end{cases}$$

This time series called first order moving average, denoted by MA(1). One can similarly define MA( $q$ ), the  $q$ th order moving average process  $\{X_t\}$  as follows:

$$X_t = \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}, \quad t = 0, \pm 1, \pm 2, \dots$$

Its ACVF is given by

$$\gamma(h) = \begin{cases} \sigma_\epsilon^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|} & \text{if } |h| \leq q, \\ 0 & \text{if } |h| > q, \end{cases}$$

where  $\theta_0 = 1$ .

Let  $|\phi| < 1$  and define

$$X_t = \phi X_{t-1} + \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots$$

Then  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  is a stationary time series with ACVF

$$\gamma(h) = \phi^{|h|} \sigma_\epsilon^2 / (1 - \phi^2), \quad h = 0, \pm 1, \pm 2, \dots$$

This time series called first order auto-regressive, denoted by AR(1). One can similarly define AR( $p$ ), the  $p$ th order auto-regressive process  $\{X_t\}$  as follows:

$$X_t = \epsilon_t + \sum_{i=1}^p \theta_i X_{t-i}, \quad t = 0, \pm 1, \pm 2, \dots$$

ACVF of AR( $p$ ) process is more involved.

### 3.1.3 Properties of ACVF

Can any symmetric function  $\gamma$  be an ACVF of some stationary time series? The answer is given by the following theorem.

**Theorem 3.1** *A function  $\gamma$  defined on all integers is an ACVF of a stationary time series if and only if it is even and non-negative definite, i.e.,*

$$\gamma(h) = \gamma(-h), \quad h = 0, 1, 2, \dots,$$

$$\sum_{i=1}^n \sum_{j=1}^n a_i \gamma(i-j) a_j \geq 0, \quad \text{for all } a = [a_1, a_2, \dots, a_n] \in R^n, \quad n \geq 1.$$

Now construct an  $n \times n$  symmetric matrix  $\Gamma_n = [\gamma_{i-j}]_{i,j=1:n}$ . Then the above condition is equivalent to saying that  $\Gamma_n$  is positive definite for each  $n$ . Using the above theorem we can show that the function

$$\gamma(h) = \begin{cases} 1 & \text{if } h = 0, \\ \rho & \text{if } h = \pm 1, \\ 0 & \text{if } |h| > 1 \end{cases}$$

is an ACVF if and only if  $|\rho| \leq 1/2$ . Thus for an MA(1) time series we have  $\gamma(0) + 2\gamma(1) \geq 0$ .

### 3.1.4 Lag Operator

It is convenient to define an operator  $B$  (for backward) as follows:

$$BX_t = X_{t-1}.$$

We can define all powers of  $B$  as follows:

$$B^k X_t = X_{t-k}, \quad k = 0, \pm 1, \pm 2, \dots.$$

Let

$$\phi(x) = 1 - \sum_{i=1}^p \phi_i x^i, \quad \theta(x) = 1 + \sum_{i=1}^q \theta_i x^i.$$

Then the AR( $p$ ) model can be succinctly written as

$$\phi(B)X_t = \epsilon_t,$$

and the MA( $q$ ) model is given by

$$X_t = \theta(B)\epsilon_t.$$

One can combine both models to create an ARMA( $p, q$ ) model defined as follows:

$$\phi(B)X_t = \theta(B)\epsilon_t,$$

assuming  $\phi(x)$  and  $\theta(x)$  have no common factors. If we further assume that all solutions of  $\phi(x) = 0$  and  $\theta(x) = 0$  are strictly outside the unit circle, then  $\{X_t\}$  is *causal*, i.e.,  $X_t$  can be written as

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},$$

and  $\{X_t\}$  is *invertible*, i.e.,  $\epsilon_t$  can be written as

$$\epsilon_t = \sum_{i=0}^{\infty} \pi_i X_{t-i},$$

for some  $\{\psi_i\}$  and  $\{\pi_i\}$  satisfying

$$\sum_{i=0}^{\infty} |\psi_i| < \infty, \quad \sum_{i=0}^{\infty} |\pi_i| < \infty.$$

### 3.1.5 Estimation and Prediction

Let  $\{X_1, X_2, \dots, X_n\}$  be the data from a time series. The sample mean is

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

The sample ACVF is given by

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{i=1}^{n-|h|} (X_{i+|h|} - \bar{X}_n)(X_i - \bar{X}_n), \quad -n < h < n.$$

One can show that the estimated ACVF  $\hat{\gamma}$  is a valid ACVF of a time series, i.e., it satisfies the conditions of Theorem 3.1. In particular, for an MA(1) time series, we have  $\hat{\gamma}(0) + 2\hat{\gamma}(1) \geq 0$ .

Next we consider the problem of predicting  $X_{n+h}$  given the history  $X_1, X_2, \dots, X_n$ . The best linear predictor is given by

$$\hat{X}_{n+h} = \mu + \sum_{i=1}^n a_i (X_i - \mu),$$

where  $a = [a_1, a_2, \dots, a_n]$  is

$$\Gamma_n a = [\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1)]'.$$

## 3.2 Multivariate Time Series

(The material in this section is based on Chapter 7 of Brockwell and Davis [5].)

Let  $\{X_t\} = \{X_t, t = 0, \pm 1, \pm 2, \dots\}$  be a stationary multivariate time series with

$$X_t = [X_{t1}, X_{t2}, \dots, X_{tm}]'.$$

Let

$$\mu = \mathbb{E}(X_t) = [\mu_1, \mu_2, \dots, \mu_m]'$$

and

$$\Gamma(h) = [\gamma_{ij}(h)]_{i,j=1:m} = \mathbb{E}((X_{t+h} - \mu)(X_t - \mu)'), \quad h = 0, \pm 1, \pm 2, \dots$$

The basic properties of  $\Gamma(\cdot)$  are

1.  $\Gamma(h) = \gamma(-h)$ ,
2.  $\gamma_{ij}(h)^2 \leq \gamma_{ii}(0)\gamma_{jj}(0)$ ,  $i, j = 1 : m$ ,
3.  $\gamma_{ii}(\cdot)$  is a valid ACVF of a univariate time series for  $i = 1 : m$ ,
4.  $\sum_{i,j=1:n} a_i' \Gamma(i-j) a_j \geq 0$  for all  $n \geq 1$  and  $a_1, a_2, \dots, a_n \in R^m$ .

A multivariate ARMA( $p, q$ ) process is define as

$$\Phi(B)X_t = \Theta(B)\epsilon_t,$$

where

$$\Phi(z) = I - \sum_{i=1}^p \Phi_i z^i, \quad \Theta(z) = I + \sum_{i=1}^q \Theta_i z^i.$$

Here  $\Phi_i$  and  $\Theta_i$  are  $m \times m$  matrices, and  $B$  is the backward shift operator applied to the vectors, with powers given by  $B^h X_t = X_{t-h}$ . As before, the time series is causal and invertible if and only if

$$\det(\Phi(z)) \neq 0, \quad \det(\Theta(z)) \neq 0, \quad \text{for all } |z| \leq 1.$$

The matrix ACVF is easy to write down for a MA( $q$ ) process

$$X_t = \sum_{i=0}^q \Theta_i \epsilon_{t-i},$$

where  $\theta_0 = I$ , and  $\epsilon_t$  has variance covariance matrix  $\Sigma$ . It is given by

$$\Gamma(h) = \sum_{j=0}^{q-|h|} \Theta_{j+|h|} \Sigma \Theta_j', \quad h = 0, \pm 1, \pm 2, \dots$$

### 3.2.1 Estimation and Prediction

The parameters  $\mu$  and  $\Gamma(h)$  of a multivariate time series are estimated as follows:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

$$\hat{\Gamma}(h) = \frac{1}{n} \sum_{i=1}^{n-|h|} (X_{i+|h|} - \bar{X}_n)(X_i - \bar{X}_n)', \quad -n < h < n.$$

The best linear predictor  $\hat{X}_{n+1}$  of  $X_{n+1}$  based on the observations  $X_i, i = 1 : n$  is given by

$$\hat{X}_{n+1} = \sum_{i=1}^n \Phi_{ni} X_i,$$

where the matrices  $\Phi_{ni}, i = 1 : n$  can be obtained by solving

$$\sum_{j=1}^n \Phi_{nj} \Gamma(i-j) = \Gamma(i), \quad i = 1 : n.$$

Formulas for  $\hat{X}_{n+h}$  for  $h \geq 2$  are more involved.

### 3.3 Continuous Time Markov Chains

### 3.4 Semi-Markov Decision Processes

## Chapter 4

# Call Markets

On Friday, Dec 11, 2009, at 4:00pm Dow Jones Industrial Average (DJIA) closed at 10,471.50. It opened at 10,494.63 at 9am, Monday, Dec 13, 2009? What made the index change over the weekend? DJIA is the sum of the stock price of thirty companies (called the Blue chips), divided by a constant, which is currently about .13. Thus the change in DJIA must be due to the change in the stock price of at least one of the blue chip companies during the non-trading hours over the weekend. For example, over the same period, Cisco stock closed at 23.77 and opened at 23.99. What caused this change? The market maker for each of these stocks has collected the submitted orders over the weekend, and set up the opening price on Monday to satisfy various market conditions, personal profit motives, and legal obligations. We shall explore mathematical models of this mechanism in this chapter.

One of the services provided by a financial market is the mechanism to discover the true value of a traded security. In order to make this statement precise, we first need to define “true value”, sometimes called the “fundamental value”.

**Definition 4.1 Fundamental value** *of a financial security is the expected (discounted) value of all the present and future costs and benefits associated with holding the security indefinitely. The market value of a security is the price at which it can be traded on the market.*

The basic idea behind the design of a market is to align the market price as close to the fundamental value as possible. This is difficult to do and to verify since the fundamental value is hard to estimate. So we build mathematical models of the market mechanisms to see if it facilitates, at least in theory, the discovery of the true fundamental value. Clearly, we need a model of how the fundamental value affects the behavior of a trader. We begin with a simple model.

## 4.1 Market Clearing Price Model

The material of this section is partly based on the algebraic illustration in Section 10.3 of Harris [21]. Suppose a market for a single security has  $n$  traders. The fundamental value of the security is  $V$ , a fixed but unknown real number. (We can also think of  $V$  as the salvage value of the security after the trade is done.) Let  $F_i$  be the forecast of this value by the  $i$ th trader, and assume that

$$F_i = V + \epsilon_i, \quad 1 \leq i \leq n,$$

where  $\{\epsilon_i, 1 \leq i \leq n\}$  are independent random variables with mean zero and variance  $1/\psi_i$ . ( $\psi_i$  is called the precision of the  $i$ th forecast.) Thus we assume that the forecast is an unbiased estimator of  $V$ . The precision  $\psi_i$  is a measure of how well informed the  $i$ th trader is. The larger the precision, the more well informed is the trader. Now suppose the excess demand  $ED_i$  of the  $i$ th trader for this security is given by

$$ED_i(p) = a(F_i - p), \quad 1 \leq i \leq n, \quad (4.1)$$

where  $p$  is the market price for this security, and  $a > 0$  is a fixed constant. Thus if  $F_i > p$ , then  $ED_i > 0$ , and the trader buys this quantity (the trader takes a long position), while if  $F_i < p$ , then  $ED_i < 0$ , the trader sells  $-ED_i$  (the trader takes a short position).

Finally, suppose the net supply of this security is zero, i.e., we can only swap the securities among the  $n$  traders. The excess demand at price  $p$  is

$$ED(p) = \sum_{i=1}^n ED_i(p).$$

The theory of partial equilibrium of the previous chapter says that the “correct” price  $p^*$  is such that

$$ED(p^*) = 0.$$

We have

$$ED(p) = \sum_{i=1}^n ED_i(p) = a \sum_{i=1}^n (F_i - p) = a \sum_{i=1}^n F_i - nap = 0.$$

This implies

$$p^* = \frac{1}{n} \sum_{i=1}^n F_i = V + \delta$$

where

$$\delta = \frac{1}{n} \sum_{i=1}^n \epsilon_i$$

is the market error. If the forecasting errors are independent, we get

$$\text{Var}(\delta) = \frac{1}{n^2} \sum_{i=1}^n 1/\psi_i.$$



If this quantity goes to zero as  $n \rightarrow \infty$ , the market price gets close to the fundamental value as the number of traders increases.

Now let  $W_i$  be the wealth of the  $i$ th trader after the trades are settled. Initially each trader's wealth is zero, and after the market clears, the wealth of trader  $i$  increases by the fundamental value of the new acquisition and decreases by the amount the trader pays for it. Thus we get

$$W_i = a(F_i - p)V - a(F_i - p)p = -a(\epsilon_i - \delta)\delta.$$

Thus we have

$$\sum_{i=1}^n W_i = 0,$$

as expected, since the total wealth has been just redistributed among the  $n$  traders. The expected value of the  $i$ th trader's wealth can be computed to be

$$\mathbf{E}(W_i) = -a\mathbf{E}(\epsilon_i - \delta)\delta = \text{Var}(\delta) - \frac{1}{n\psi_i}.$$

Thus, the higher the  $\psi_i$ , higher the expected wealth after the transaction. In other words, the wealth moves from less informed traders to more informed traders. This also makes intuitive sense.

The total volume of trade is

$$T = \sum_{i=1}^n |a(F_i - p^*)| = \sum_{i=1}^n a|\epsilon_i - \delta|.$$

We need further distributional assumption on  $\epsilon$ 's to compute the  $\mathbf{E}(T)$ . Suppose  $\epsilon_i$  are independent Normal random variables. Then,  $\epsilon_i - \delta$  is a Normal random variable with mean zero and variance

$$\theta_i^2 = \frac{n-2}{n\psi_i} + \text{Var}(\delta).$$

Then we see that

$$\mathbf{E}(T) = a\sqrt{2/\pi} \sum_{i=1}^n \theta_i.$$

A natural question is: why do we assume that the traders choose the excess demand function of Equation 4.1? This leads us to the utility based model in the next section.

## 4.2 Utility Based Model.

The material of this section is partly based on Section 1 of Madhavan and Panchapkesan [32]. We start with a simple model of the fundamental value and the information possessed by the traders as in the previous section.

As before, suppose the market consists of  $n$  traders, who want to trade in a security whose true fundamental value is  $V$ . Unlike in the previous section, we assume that  $V \sim N(\mu, 1/\zeta)$ . Let  $F_i$  be the forecast received by trader  $i$ . As in Section 4.1 we use the forecast model:  $F_i = V + \epsilon_i$ , where  $\{\epsilon_i, i = 1, \dots, n\}$  are independent  $N(0, 1/\psi_i)$  random variables. Here  $\zeta$  is the precision of the fundamental value, and  $\psi_i$  is the precision of the  $i$ th trader's forecast.

Using the properties of the Normal distribution we can show that, given  $F_i$ ,  $V \sim N(\mu_i, \sigma_i^2)$ , where, using  $\gamma_i = \zeta/(\zeta + \psi_i)$ ,

$$\mu_i = \mu\gamma_i + F_i(1 - \gamma_i), \quad (4.2)$$

$$\sigma_i^2 = (\zeta + \psi_i)^{-1}. \quad (4.3)$$

Next, let  $c_i$  be the initial cash position of trader  $i$ . It may be positive or negative. Also, let  $e_i$  be the number of shares of the security held initially by trader  $i$ . The decision variable for trader  $i$  is  $q_i$ , the number of shares to be traded (positive for buy, negative for sell). If the trade occurs at price  $p$ , the terminal wealth of trader  $i$  will be

$$W_i = V(q_i + e_i) + c_i - pq_i, \quad 1 \leq i \leq n.$$

Now suppose the utility function of the  $i$ th trader is

$$U_i(w) = -e^{-\rho_i w}.$$

This is an increasing concave utility function with constant risk aversion factor  $\rho_i$ . The  $i$ th trader determines the optimal  $q_i$  by maximizing the expected utility of the terminal wealth  $W_i$ . Now, given  $F_i$ ,  $W_i$  is a Normal random variable with mean

$$\alpha_i = \mu_i(q_i + e_i) + c_i - pq_i,$$

and variance

$$\beta_i^2 = \sigma_i^2(q_i + e_i)^2.$$

Using the formula for the moment generating function of a Normal random variable, we get

$$\mathbf{E}(U_i(W_i|F_i)) = -\exp\left(-\rho_i\alpha_i + \frac{1}{2}\rho_i^2\beta_i^2\right).$$

Maximizing this is equivalent to finding the  $q_i$  that maximizes

$$\rho_i\alpha_i - \frac{1}{2}\rho_i^2\beta_i^2.$$

This yields the optimal  $q_i = q_i(p)$  as

$$q_i(p) = a_i - b_i p, \quad (4.4)$$

where

$$b_i = \frac{1}{\rho_i\sigma_i^2}, \quad a_i = \mu_i b_i - e_i.$$

This is indeed a linear function of  $p$ ! Thus the total excess demand is

$$ED(p) = \sum_{i=1}^n \mu_i b_i - \sum_{i=1}^n e_i - \left( \sum_{i=1}^n b_i \right) p \sum_{i=1}^n \mu_i b_i - \left( \sum_{i=1}^n b_i \right) p.$$

The equilibrium price is set so that

$$ED(p^*) = 0.$$

This yields

$$p^* = \frac{\sum_{i=1}^n \mu_i b_i}{\sum_{i=1}^n b_i} - \frac{\sum_{i=1}^n e_i}{\sum_{i=1}^n b_i} = \mu \frac{\sum_1^n \gamma_i / (\rho_i \sigma_i^2)}{\sum_1^n 1 / (\rho_i \sigma_i^2)} + \frac{\sum_1^n F_i (1 - \gamma_i) / (\rho_i \sigma_i^2)}{\sum_1^n 1 / (\rho_i \sigma_i^2)} - \frac{\sum_{i=1}^n e_i}{\sum_{i=1}^n b_i}. \quad (4.5)$$

Thus the market clearing price is weighted average of  $\mu_i$ 's, the conditional expected value estimated by trader  $i$ , with an additional term representing the hedging preferences. Now, the expected value of the signal  $F_i$  is  $\mu$ . Hence, the expected value of the above price is  $\mu$ , the true expected value of  $V$ , plus the bias term involving  $e_i$ 's. If we assume that  $e_i$ 's are zero mean random variables, the  $p^*$  becomes an unbiased estimator of  $\mu$ . Thus this market clearing mechanism will indeed lead to a correct price discovery (at least under the assumptions of this model)!

**Special Cases 1: Uninformed Traders.** Setting  $\psi_i = 0$  is equivalent to assuming that the  $i$ th trader is uninformed. In this case  $\mu_i = \mu$  and  $\sigma_i^2 = 1/\zeta$ , that is, the uninformed trader bases his order schedule on the prior distribution of  $V$ .

**Special Cases 2: Common Information.** Another special case is where all traders get their signal from a single source. This can be modeled by setting  $\psi_i = \psi$  and  $F_i = F$  for all  $1 \leq i \leq n$ . In this case, the equilibrium price reduces to

$$p^* = \mu \frac{\zeta}{\zeta + \psi} + F \frac{\psi}{\zeta + \psi} - \frac{\sum_{i=1}^n e_i}{n\rho(\zeta + \psi)},$$

which is just the conditional expected value of  $V$  given the signal, with a hedging term. This dramatically displays the influence of rating agencies, or TV shows hawking a particular stock.

**Special Cases 3: Identically Informed Traders.** Suppose the quality of the information for all the traders is identical. We model this by setting  $\psi_i = \psi$  for all traders. This gives

$$\gamma_i / \sigma_i^2 = \zeta,$$

and Equation 4.5 reduces to:

$$p^* = \mu \frac{\zeta}{1 + \zeta} + \frac{\psi}{1 + \zeta} \bar{F}_w - \frac{\sum_{i=1}^n e_i}{n\rho(\zeta + \psi)} \quad (4.6)$$

where  $\bar{F}_w$  is the weighted average given by

$$\bar{F}_w = \frac{\sum_1^n F_i / \rho_i}{\sum_1^n 1 / \rho_i}.$$

Thus the price is a weighted average of  $\mu$  and  $\bar{F}_w$ .

**Question:** Costly Information. We have assumed that traders get their information at no additional cost. However, in reality, information is costly. What is the effect of the cost of obtaining the information on order placing and pricing? This issue is addressed by Grossman and Stiglitz [18], and Kyle [30]. We shall treat these models in Section 4.4. ok

**Question:** What is the expected total trade volume? ■

### 4.3 Market Maker's Role

The models in Sections 4.1 and 4.2 have significant deficiency: it assumes that there is no market maker, the price is set by an automated system, and all traders are price takers. This is a reasonable model for ECN (Electronic crossing network) that operates a single auction call. However, it is not a good model for NYSE opening call since there the specialist or the market maker plays a significant role that is very different from the other traders.

In this section we extend the analysis of the model of Section 4.2 to include a market maker by simply calling him trader 0. He is similar to the other traders in all respects except that he decides the opening (clearing) price  $p$ . He then trades  $q_0(p)$  from his inventory of  $e_0$  to ensure that

$$\sum_{i=0}^n q_i(p) = 0.$$

Furthermore, he can see the entire order book as it builds leading up to the call time. Thus he can estimate the signals that the other traders are getting and can use them to construct his own signal  $F_0$ . In addition, one of the obligation he has as a market maker is to maintain price continuity, a vaguely defined objective. Essentially it is duty to see that the opening price (the clearing price established at the call) does not vary too much from the previous close price  $p_c$ . We incorporate this cost by a quadratic deviation cost with weight parameter  $\delta > 0$ . Thus the terminal wealth of the market maker is given by

$$W_0 = V(q_0 + e_0) + c_0 - pq_0 - \delta(p - p_c)^2.$$

The market maker's job is to decide  $p$  and  $q_0$  to maximize the expected value of the utility of his terminal wealth  $E(U_0(W_0|F_0))$ .

Next we discuss how the market maker arrives at his forecast  $F_0$  of  $V$ . Unlike the traders, the market maker has no independent source of the forecast. However, he builds it by observing the

demand schedules submitted by the traders. For example, from the  $i$ th trader's schedule he can compute

$$y_i = \frac{a_i/b_i - \mu\gamma_i}{1 - \gamma_i} = F_i - \frac{e_i}{b_i(1 - \gamma_i)}.$$

From this information he creates the forecast

$$F_0 = \sum_{i=1}^n y_i/n.$$

Since  $F_i \sim N(V, 1/\psi_i)$ , and  $e_i$ 's are assumed to be  $N(0, \theta_i^2)$ , we see that  $F_0 \sim N(V, 1/\psi_0)$  where

$$\frac{1}{\psi_0} = \frac{1}{n^2} \left[ \sum_{i=1}^n \frac{1}{\psi_i} + \sum_{i=1}^n \frac{\theta_i^2}{b_i^2(1 - \gamma_i)^2} \right].$$

The market maker decides  $p$  to set

$$q_0 + \sum_{i=1}^n q_i(p) = q_0 + \sum_{i=1}^n \mu_i b_i - \sum_{i=1}^n e_i - \left( \sum_{i=1}^n b_i \right) p = 0.$$

This gives

$$p = p^* + \lambda q_0 \tag{4.7}$$

where  $p^*$  is given by Equation 4.5, and

$$\lambda = 1 / \sum_{i=1}^n b_i.$$

Now, as in the common trader's case, the utility maximization problem for the market maker is equivalent to maximizing

$$\rho_0 \alpha_0 - \frac{1}{2} \rho_0^2 \beta_0^2 - \delta(p - p_c)^2.$$

This can be solved to get

$$q_0 = \frac{\mu_0 - p^* - \rho_0 \sigma_0^2 e_0 - 2\delta\lambda(p^* - p_c)}{2\lambda + \rho_0 \sigma_0^2 + 2\delta\lambda^2}. \tag{4.8}$$

Using this  $q_0$  in Equation 4.7 we get the opening price set by the market maker.

**Question:** Check these calculations. ■

## 4.4 Strategic Customers: Grossman Model

The traders in Section 4.2 behave in a non-strategic fashion. That is, they assume that the market price  $p$  is given, and their behavior does not affect it. However, clearly the market price is the result of their combined behavior. We need a game theoretic analysis to study their behavior correctly.

We begin with a simple model analyzed in Grossman [17].

Let  $F = [F_1, F_2, \dots, F_n]$  be the vector of information signals received by the  $n$  traders. From Equation 4.5, we see that the equilibrium price is a function of  $F$ . We denote it as  $p^*(F)$  to make this dependence clear. After the market clearing price is announced, will the  $i$ th trader still think of  $q_i(p^*(F))$  as given by Equation 4.4 as the optimal trade size? Or, given an opportunity, will he change? The price is a true equilibrium price if, even after given the choice of changing the order, the trader won't actually do so. This equilibrium price is called the Nash equilibrium price (or the rational expectation price), and we shall denote it as  $p^{**} = p^{**}(F)$ . Let  $E(U_i(q_i(p)))$  be the expected value of the utility of the terminal wealth when the  $i$ th trader uses an order schedule  $q_i(p)$ . We shall say that  $(q_i^{**}(p^{**}), p^{**})$  is a Nash equilibrium if, for any given  $F$ ,

$$E(U_i(q_i^{**}(p^{**}))) \geq E(U_i(q_i(p))), \quad i = 1, 2, \dots, n$$

and

$$\sum_{i=1}^n q_i^{**}(p^{**}) = 0.$$

The main result in Grossman [17] is given in the next theorem.

**Theorem 4.1** *Suppose  $\psi_i = 1$  for  $1 \leq i \leq n$ . The Nash equilibrium price is given by*

$$p^{**} = \alpha_0 + \alpha_1 \bar{F}, \quad (4.9)$$

where

$$\begin{aligned} \bar{F} &= \frac{\sum_1^n F_i}{n}, \\ \alpha_1 &= \frac{n}{\zeta + n}, \\ \alpha_0 &= \mu \frac{\zeta}{\zeta + n} - \frac{\sum_1^n e_i}{(\zeta + n) \sum_1^n 1/\rho_i}, \end{aligned}$$

and the order sizes are given by

$$q_i^{**} = \frac{1}{\rho_i} \frac{\sum_1^n e_i}{\sum_1^n 1/\rho_i} - e_i. \quad (4.10)$$

**Question:** Provide the details from [17]. ■

Note that the Nash price depends only on the average signal, and the order size does not depend on the information  $F_i$  at all! This has a disturbing implication: since the information in real life is not free, the traders will be unwilling to pay for it, since their decision to trade will not depend on the information at all! In a later paper Grossman and Stiglitz [18] explore this problem further. Kyle [30] defines an imperfect competition that leads to a slightly modified notion of market equilibrium price, that avoids the conclusion of the Grossman model.

## 4.5 Strategic Customers: Kyle Model

In this section we consider the model of a single informed customer considered by Kyle [29]. The fundamental value  $V$  is normally distributed with mean  $\mu$  and precision  $1/\zeta$ . The strategic customer can observe the actual value of  $V$ , and bases his order size  $Q$  on the observed value. Thus we can think of him as an insider. There are liquidity traders who place a combined order of size  $U$  that is normally distributed with mean zero and variance  $\sigma_u^2$ . The insider does not know  $U$ . The market clearing price  $P$  depends on  $Q + U$ . The strategic trader maximizes his expected profit given by

$$\pi(P, Q) = \mathbb{E}((V - P)Q).$$

unlike in the previous models, the price is determined by market efficiency, that is,

$$P = \mathbb{E}(V|Q + U).$$

The next theorem gives the optimal strategy for the strategic trader. First some notation:

$$\beta = \sqrt{\sigma_u^2 \zeta}, \quad \lambda = 2/\beta.$$

**Theorem 4.2** *The optimal strategy for the strategic trader is*

$$Q = \beta(V - \mu).$$

*The market clearing price under this strategy is given by*

$$P = \mu + \lambda(Q + U).$$

**Proof:**

## 4.6 A Real Call Market

In the above models we assumed that the “market” (with or without a market maker) has the excess demand functions available to compute the equilibrium price. How does a market obtain this function in practice? No trader is going to actually reveal his excess demand function as a function of the price. In this section we consider a single auction call market and study how it solves this problem.

A single auction call market operates as follows: Buyers and sellers submit their orders (price, quantity, buy or sell) to a central location up to a pre-specified time, say 9:30am. (This may be done electronically). At that time the market maker aggregates all the orders and announces a single price  $p$  at which all trades are settled. The buyers who wanted to buy at price  $p$  or higher get their order filled, and so do the sellers who wanted to sell at price  $p$  or lower. The unsatisfied

orders are then withdrawn or have to wait for the next call. The NYSE opens with single auction call like this for each stock.

Suppose there are  $n$  sellers and  $m$  buyers. Let  $S_i$  be the order size and  $a_i$  the asking price of the order from the  $i$ -th seller. Similarly let  $D_i$  be the order size and  $b_i$  be the bid price from the  $i$ th buyer. We assume that the buy orders are ranked according to decreasing bid prices, with ties broken in favor of earlier orders, and further ties broken in favor of larger orders. Similarly, the sell orders are ranked by increasing ask prices, followed by arrival time, followed by decreasing sizes. We assume that orders can be filled partially. From this data the market maker constructs the supply and demand functions as follows:

$$S(p) = \sum_{i=1}^n S_i 1_{\{a_i \leq p\}}, \quad p \geq 0,$$

$$D(p) = \sum_{i=1}^m D_i 1_{\{b_i \geq p\}}, \quad p \geq 0.$$

Note that  $D$  is left continuous non-increasing while  $S$  is right continuous non-decreasing. Next the excess demand function is computed as

$$ED(p) = D(p) - S(p).$$

Now define

$$A = \{p \geq 0 : ED(p) = 0\}.$$

If  $A$  is not empty, it is an interval. It may be closed or open on either end. Any price in  $A$  can be chosen to be the clearing price, and typically the market has a published rule about how it is chosen. If  $A$  is empty, the market clearing price  $p$  is defined to be

$$p = \sup\{p \geq 0 : ED(p) > 0\} = \inf\{p \geq 0 : ED(p) < 0\}.$$

At the announced call time at most one buy or at most one sell order will be partially filled. All other orders are either completely filled or completely unfilled. As we saw in the previous chapter, this price maximizes the sum of the surpluses for the buyers and the sellers.

## 4.7 Model of a Real Call Market

The model of Section 4.2 does not reflect the real call market described in Section 4.6. In that model every trader submits a linear demand schedule  $q_i(p)$ , and all orders are traded at  $p$ . However, in the real call market the  $i$ th trader submits a step function of  $p$  as a demand schedule: namely, buy  $D_i$  at  $p \leq b_i$  and buy nothing at  $p > b_i$ , or sell  $S_i$  at  $p \geq a_i$  and sell nothing for  $p < a_i$ . In such a call market there is a distinct possibility that a trader's order may not execute. How does a



trader decide upon what  $(D_i, b_i)$  or  $(S_i, a_i)$  to use? This is a rather complex problem and a partial answer is provided by Ho et al [24]. We discuss simplified version of their results here.

The main obstacle to the analysis is that the trader does not know what the market clearing price will be. Clearly a trader may place multiple orders in order to recreate the linear demand function of the previous section. However, this is not how the traders behave in reality. In order to capture the realistic behavior, and keep the analysis tractable we shall concentrate on a single trader who faces an uncertain market clearing price  $P$ , and assume that he is allowed to place a single buy limit order  $(D, b)$  or a single sell limit order  $(S, a)$ . The market is so large that his order has no effect on  $P$ .

First consider a buy order of size  $D > 0$  at bid price  $b$ . If the clearing price  $P \leq b$  this order will be executed and the terminal wealth will be

$$W = (D + e)V + c - DP.$$

If  $P > b$  the order will not execute and the terminal wealth will be

$$W = eV + c.$$

Now suppose the trader knows the distribution of  $P$ :

$$G(p) = \mathbb{P}(P \leq p),$$

and his utility function is given by

$$U(w) = -\exp(-\rho w),$$

for a given risk aversion factor  $\rho > 0$ . Then the expected utility of the terminal wealth for using the buy order  $(D, b)$  can be computed as

$$C_B(D, b) = U_0(1 - G(b) + e^{-\rho D(\alpha - \beta D)} \int_0^b e^{\rho D x} dG(x))$$

where

$$U_0 = -\exp(-\rho(\mu_v e + c) + \rho^2 \sigma_v^2 e^2 / 2),$$

$$\alpha = \mu_v - \rho \sigma_v^2 e, \quad \beta = \frac{1}{2} \rho \sigma_v^2.$$

The trader wants to pick  $(D, b)$  to maximize  $C_B(D, b)$ .

We treat two important cases:

(1).  $P$  is a discrete random variable. In this case,  $C_B(D, b)$  is a step function of  $b$ , and the optimal

$b$  will be one of the values that  $P$  can take. In particular, if  $P$  is a constant  $p < \alpha$ , then the optimal  $b$  is  $p$ , and the optimal  $D$  is satisfies

$$\alpha - 2\beta D = p.$$

Thus the optimal buy order is  $((\alpha - p)/(2\beta), p)$ . If  $p > \alpha$ , placing a buy order will reduce the expected utility.

(2)  $P$  is a continuous random variable with positive density  $g$  over  $[0, \infty)$ . Then  $C_B(D, b)$  is a differentiable function of  $b$  over  $[0, \infty)$ . Using routine calculations one can show that for a fixed  $D > 0$ , the optimal  $b$  satisfies

$$b = \alpha - \beta D. \quad (4.11)$$

Using this one can show that the optimal  $D$  satisfies

$$(\alpha - 2\beta D) \int_0^{\alpha - \beta D} e^{\rho D x} g(x) dx = \int_0^{\alpha - \beta D} e^{\rho D x} x g(x) dx. \quad (4.12)$$

It can be shown if  $\alpha > 0$ , there is a  $D \in [0, \alpha/2\beta]$  that satisfies the above equation. The optimal bid price is given by substituting this  $D$  in Equation 4.11. If  $\alpha < 0$ , placing a buy order will reduce the expected utility.

One can repeat the above analysis for a sell order  $(S, a)$  for  $S > 0$ . If the clearing price  $P \geq a$  this order will execute and the terminal wealth will be

$$W = (-S + e)V + c + SP.$$

If  $P < a$  the order will not execute and the terminal wealth will be

$$W = eV + c.$$

Then the expected utility of the terminal wealth for using the sell order  $(S, a)$  can be computed as

$$C_A(S, a) = U_0(G(a-) + e^{-\rho S(\alpha - \beta S)} \int_a^\infty e^{-\rho S x} dG(x)).$$

The trader wants to pick  $(S, a)$  to maximize  $C_A(S, a)$ .

Following the same steps as in the analysis of the buy order we see that if the clearing price is a constant  $p > \max(0, -\alpha)$ , the optimal ask price is  $p$ , and the optimal order size is  $(\alpha + p)/(2\beta)$ . On the other hand, if  $P$  is a continuous random variable with positive density over  $[0, \infty)$ , the optimal  $S$  is given by

$$(2\beta S - \alpha) \int_{\beta S - \alpha}^\infty e^{-\rho S x} g(x) dx = \int_{\beta S - \alpha}^\infty e^{-\rho S x} x g(x) dx. \quad (4.13)$$

and the optimal ask price is given by

$$a = \beta S - \alpha. \quad (4.14)$$

Unfortunately, neither Equation 4.12 nor 4.13 can be solved in a closed form. Ho et al [24] assume that  $P$  is normally distributed, but that does not help in solving this equation, and it allows negative prices, a troublesome deficiency.

Clearly the market clearing price is the result of the interaction of the buy and sell orders of all the participating customers. What type of behavior can we expect of the market clearing price if all the participants behaved in the optimal fashion as described above? This analysis gets rather complicated if there are a finite number of traders. Hence we consider the limiting case of an infinite number of traders below.

To keep the analysis simple, we shall assume that all traders have the same information about the distribution of  $P$  and  $V$ , and have the same coefficient of risk aversion. They only differ in their initial endowment  $e$ . (We can ignore the initial cash reserves, since they do not play any role in their behavior.) This implies that the  $\alpha$  values differ from customer to customer, but  $\beta$  values are the same. We shall call a customer to be of type  $t \in (-\infty, \infty)$  if his  $\alpha$  value is  $t$ . Let  $H(t)$  be the number of customers of type  $t$  or less. We shall assume that there is a function  $h(t)$  so that

$$H(t) = \int_{-\infty}^t h(x)dx.$$

We shall denote the optimal buy order of a type  $t$  customer by  $(D(t), b(t))$  and the optimal sell order as  $(S(t), a(t))$ . If the buy order does not exist, we shall define  $(D(t), b(t)) = (0, 0)$  and if the sell order does not exist, we shall set  $(S(t), a(t)) = (0, 0)$ . Then we can construct the supply and demand curves for this market as follows:

$$\text{Supply}(p) = \int_{t:a(t) \leq p} S(t)h(t)dt,$$

$$\text{Demand}(p) = \int_{t:b(t) \geq p} D(t)h(t)dt.$$

The excess demand function is then given by

$$ED(p) = \text{Demand}(p) - \text{Supply}(p).$$

Then, from the theory of partial equilibrium, the market clearing price  $p^*$  satisfies

$$ED(p^*) = 0.$$

We do not give the details of the calculations here (See Ho et al [24]) because this result says that the market price is deterministic. This is rather unsatisfactory. If this result holds, then we only need the analysis of a single trader using deterministic prices. However, that analysis shows that all the bid and ask prices would be identical, a situation far from reality. This implies is that we need to use game theory to analyze this problem. We do that in the next section.

## 4.8 Game Theoretic Models of Call Markets

In this section we study game theoretic models of call markets. See Harsanyi [19] for the general theory of multi person games with incomplete information. Also see Wilson [51] for a good overview of this area.

A call market is a form of auction. In the auction terminology it is a special case of a sealed bid double auction, where each buyer and seller submits a sealed bid, and the final clearing price is settled according to a given rule. This requires a different model of how the buyers and seller act. In the previous section the buyers and sellers had a common prior distribution of the clearing price, and a private valuation of the value of the asset. In this section we shall assume that there is a common knowledge about the distribution of the asset value. However, each trader gets his own signal about actual value that is known only to himself. His decision to trade depends on his own signal, the common knowledge about the value, and the number of traders involved. We make this more precise below.

We begin with a market with one buyer and one seller, each wanting to buy or sell one unit, as studied by Chatterjee and Samuelson [6] and Satterthwaite and Williams [44]. (These papers treat assume that the traders are risk neutral. Here we explain their results for traders with constant risk aversion.) Let trader 1 be the buyer and trader 2 the seller. As in Section 4.2, let  $V \sim N(\mu, 1/\zeta)$  be the fundamental value of the security, and  $F_i \sim N(V, 1/\psi_i)$  be the signal received by trader  $i$ . Given  $F_i$ , the  $i$ th trader's distribution of  $V$  is  $N(\mu_i, \sigma_i^2)$ , where  $\mu_i$  and  $\sigma_i^2$  are as in Equations 4.2 and 4.3. We shall assume that the parameters  $\mu, \zeta, \psi_1$  and  $\psi_2$  are known to all. However the signal  $F_i$  is known only to trader  $i$ .

After seeing their respective signals the first trader puts in a bid price of  $b$ , and the second trader puts in an ask price of  $a$ . If  $b < a$  no trade occurs. If  $b \geq a$ , a trade occurs at price  $p \in [a, b]$ . Initially the first trader's wealth is 0, and after the market call his wealth is

$$W_1 = \begin{cases} V - p & \text{if } b \geq a \\ 0 & \text{if } b < a \end{cases},$$

with expected utility

$$\mathbb{E}(e^{-\rho_1 W_1} | F_1) = \begin{cases} -\exp(-\rho_1(\mu_1 - p) + \frac{1}{2}\rho_1^2\sigma_1^2) & \text{if } b \geq a \\ -1 & \text{if } b < a. \end{cases} \quad (4.15)$$

Similarly, initially the second trader's worth is  $V$ , and after the market call it changes to

$$W_2 = \begin{cases} p & \text{if } b \geq a \\ V & \text{if } b < a, \end{cases}$$

with expected utility

$$\mathbf{E}(e^{-\rho_1 W_2} | F_2) = \begin{cases} -\exp(-\rho_2 p) & \text{if } b \geq a \\ -\exp(-\rho_2 \mu_2 + \frac{1}{2} \rho_2^2 \sigma_2^2) & \text{if } b < a. \end{cases} \quad (4.16)$$

Each user sets his prices to maximize his own expected utility after the market call.

Using Equations 4.15 and 4.16 we see that the buyer's reservation price is

$$p_1(F_1) = \mu_1 - \frac{1}{2} \rho_1 \sigma_1^2 = \mu \gamma_1 + F_1(1 - \gamma_1) - \frac{1}{2} \rho_1 \sigma_1^2, \quad (4.17)$$

which is a linear function of the buyer's signal  $F_1$ . The buyer won't buy at a price more than  $p_1(F_1)$ . Similarly, the seller's reservation price is

$$p_2(F_2) = \mu_2 - \frac{1}{2} \rho_2 \sigma_2^2 = \mu \gamma_2 + F_2(1 - \gamma_2) - \frac{1}{2} \rho_2 \sigma_2^2. \quad (4.18)$$

The seller won't sell at a price less than  $p_2(F_2)$ . Thus a trade takes place if and only if  $p_2(F_2) \leq p \leq p_1(F_1)$ , if such a price exists. The distribution of  $F_i \sim N(\mu, 1/\zeta + 1/\psi_i)$ , is known to trader  $j \neq i$ . The buyer's strategy is to select his bid price based on his signal  $F_1$  and his knowledge of the distribution of  $F_2$ . We shall specify the buyer's strategy as a function  $b = B(F_1)$ . Similarly the seller's strategy is a function  $a = A(F_2)$ .

Now suppose the market uses the following rule to settle the trading price when  $a \leq b$ :  $p = p(a, b) = kb + (1 - k)a$ . For given functions  $A$  and  $B$ , the buyer's expected utility at the end of the trading session is given by

$$u_1(b, F_1) = -\mathbf{E}(\exp(-\rho_1(\mu_1 - p(A(F_2), b))) + \frac{1}{2} \rho_1^2 \sigma_1^2) \mathbf{1}\{b \geq A(F_2)\}) - \mathbf{P}(b < A(F_2)), \quad (4.19)$$

where the expectation is with respect to  $F_2$ . The seller's expected utility is given by

$$u_2(a, F_2) = -\mathbf{E}(\exp(-\rho_2 p(a, B(F_1))) \mathbf{1}\{a \leq B(F_1)\}) - \exp(-\rho_2 \mu_2 + \frac{1}{2} \rho_2^2 \sigma_2^2) \mathbf{P}(a > B(F_1))), \quad (4.20)$$

where the expectation is with respect to  $F_1$ .

Now, if  $k = 1$ ,  $p(a, b) = b$ , and a trade occurs only at the buyer's price. Hence the optimal strategy for the seller is to use  $A(F_2) = p_2(F_2)$ , which is a linear increasing function of  $F_2$  as given in Equation 4.18. Substituting in Equation 4.19, we get  $u_1(b, F_1)$  as a scalar function of  $b$ . The optimal value of  $b$  can now be computed. Similarly, if  $k = 0$ , a trade occurs only at the seller's price, hence the optimal strategy for the buyer is to use  $B(F_1) = p_1(F_1)$ , a linear increasing function of  $F_1$  as given in Equation 4.17. Using this in Equation 4.20 we can find the optimal ask strategy  $a = A(F_2)$ . If  $0 < k < 1$ , the situation is more complex.

**Question:** Compute the optimal strategies  $A$  and  $B$  when  $0 < k < 1$ . ■

Rustichini et al [43] extend the results of [6] and [44] to the case of  $m$  buyers and  $n$  sellers, each trading one unit of an asset. They assume all the buyers are identical and independent, and all the sellers are identical and independent.

**Question:** Derive their results for the traders with constant risk aversion as we have done above. ■

## 4.9 Stochastic Characteristics of a Call Market

What will be the stochastic characteristics of a call market where traders behave as predicted by the game theoretic analysis? For example, how many units will change hands? At what price? How many trades will go unsatisfied? Questions like these are answered by Mendelson [34] using a simple model. We restate his results for the simplest tractable case.

Let  $N$  be the number of buyers and  $M$  the number of sellers, each in the market to trade a single unit. Assume that  $N$  and  $M$  are iid  $P(\lambda)$  random variables. This is a reasonable stochastic model of a market with a large number of players, a small fraction of whom are active at any given call, and an active trader is a buyer or a seller independently with probability  $1/2$ . Suppose that the bid prices and the ask prices are uniformly distributed over  $[0, K]$ , where  $K > 0$  is a given constant. Let  $Q$  be the number of units that get traded, and the interval  $[L, U]$  be such that the supply and demand are balanced for all prices  $p \in [L, U]$ . Then Mendelson shows that

$$\begin{aligned} P(Q = m) &= e^{-\lambda K} \frac{(\lambda K)^{2m}}{(2m)!} + e^{-\lambda K} \frac{(\lambda K)^{2m+1}}{(2m+1)!}, \quad m = 0, 1, 2, \dots, \\ E(Q) &= \lambda K/2 - (1/4)(1 - e^{-2\lambda K}), \\ \text{var}(Q) &= \lambda K - 2\lambda K e^{-2\lambda K} + (1/4)e^{-\lambda K} \sinh(\lambda K) - (1/4)e^{-2\lambda K} \sinh^2(\lambda K), \\ E(L) &= K/2 - (1 - e^{-2\lambda K})/(4\lambda), \\ E(U) &= K/2 + (1 - e^{-2\lambda K})/(4\lambda), \\ \text{var}(L) = \text{var}(U) &= \frac{K}{4\lambda} - \frac{1}{16\lambda^2}(1 - e^{-2\lambda K})(3 - e^{-2\lambda K}), \\ E(\Delta) = E(U - L) &= (1 - e^{-2\lambda K})/(2\lambda). \end{aligned}$$

These results imply that, as the number of traders increases (that is,  $\lambda \rightarrow \infty$ ), the market prices become efficient.

**Question:** Extend this analysis to the case where the ask and bid prices have a non-uniform distribution. ■

## Chapter 5

# Quote-Driven Markets: Dealers as Inventory Managers

In this chapter we shall develop and analyze stochastic models of the pricing mechanisms in a continuous trading in a quote driven market where trading occurs strictly by market orders. As described in Chapter 1, a typical quote driven market for a security is operated by a market maker or a dealer, who stands ready to buy at the bid price, and sell at the ask price from his own inventory with any one who wishes to do so. He continuously adjusts these bid and ask prices in response to the inflow of buy and sell orders coming to him from the traders, and also the level of his own inventory. If the inflow of buy orders is large, he ends up selling more, and his inventory depletes. In response he raises the ask price. Similarly, if the inflow of sell orders increases, he ends up buying more, and his inventory increases and he responds by lowering the bid price. He makes money by buying at the bid price, and selling at the ask price, which is higher than the bid price. He tries to manage the bid-ask spread to maximize his profits. If he makes the bid too low, there will be very few buyers. If he makes the ask too high, it will make the sellers scarce.

The earliest paper describing the role of the market maker or the dealer in a continuous trading markets is Bagehot [3]. The bid -ask spread is an indicator of the cost of transacting without delay, the main service provided by the market maker or the dealer. Demsetz [11] is one of the earliest papers presenting a regression model for estimating the cost of transacting as evidenced by the bid-ask spread. In this chapter we shall study models which employ continuous time stochastic processes with discrete state spaces.

## 5.1 A Static Model.

This material is motivated by the model analyzed in Garman [14]. The market maker posts two prices for a stock: the ask price ( $a$ ) at which he is willing to sell, and the bid price ( $b$ ) at which he is willing to buy. We consider the static case: the market maker sets these prices once and never changes them.

He maintains  $a > b$  and makes a profit from the bid-ask spread  $a - b$ . The market buy orders (to buy from the market maker at price  $a$ ) arrive according to a Poisson process with rate  $\mu(a)$ , and the market sell orders (to sell to the market maker at price  $b$ ) arrive according to a Poisson process with rate  $\lambda(b)$ . It makes sense to assume that  $\mu(a)$  is a decreasing function of  $a$  and  $\lambda(b)$  is an increasing function  $b$ . For the sake of simplicity, assume that all orders are of size 1.

Let  $X(t)$  be the number of shares of the stock (called the inventory) held by the market maker at time  $t$ . If  $X(t) > 0$ , the market maker has a long position, and if  $X(t) < 0$  he has a short position. From the Poisson assumption it follows that  $\{X(t), t \geq 0\}$  is a birth and death process on all integers with birth rates  $\lambda_i = \lambda(b)$  and death rates  $\mu_i = \mu(a)$ ,  $-\infty < i < \infty$ . From the theory of birth and death processes we know that  $\{X(t), t \geq 0\}$  is null recurrent if  $\lambda(b) = \mu(a)$  and transient otherwise. Thus it makes sense for the market maker to choose  $b$  and  $a$  so that

$$\lambda(b) = \mu(a).$$

At this equilibrium pricing the expected inventory level stays constant. However, this gives only one equation for the two unknowns  $a$  and  $b$ . We derive another equation by considering the cash reserves in the market maker's account. Under our assumptions, the market maker's cash reserves increase at rate

$$a\mu(a) - b\lambda(b).$$

It seems reasonable that the market maker tries to maximize his cash inflow rate. Thus we can model the market maker's decision problem as the following constrained optimization problem:

$$\begin{aligned} &\text{Maximize} && a\mu(a) - b\lambda(b) \\ &\text{Subject to} && \mu(a) = \lambda(b). \end{aligned}$$

Now let

$$R(\theta) = \theta\mu^{-1}(\theta), \tag{5.1}$$

and

$$C(\theta) = \theta\lambda^{-1}(\theta). \tag{5.2}$$

If the ask price induces the traders to buy at rate  $\theta$ , it will generate cash inflow rate of  $R(\theta)$  for the market maker. Similarly, if the bid price induces the traders to sell at rate  $\theta$ , it will generate



cash outflow rate of  $C(\theta)$  for the market maker. Using

$$\mu(a) = \lambda(b) = \theta$$

we see that the net cash flow rate is given by  $R(\theta) - C(\theta)$ . Hence the above maximization problem can be written as

$$\text{Maximize } f(\theta) = R(\theta) - C(\theta). \quad (5.3)$$

Now suppose  $\mu(\cdot)$  is decreasing convex and  $\lambda(\cdot)$  is increasing concave. Then  $\mu^{-1}(\cdot)$  is decreasing convex, and  $\lambda^{-1}(\cdot)$  is increasing convex. Using these characterizations we can show that  $R(\cdot)$  is concave and  $C(\cdot)$  is convex. Hence  $f(\cdot)$  is concave. Hence we get the following general result:

**Theorem 5.1** *If  $\mu(\cdot)$  is decreasing convex and  $\lambda(\cdot)$  is increasing concave, the market maker's maximization problem has a unique maximum.*

The optimal  $\theta^*$  that maximizes  $f(\theta)$  in Equation 5.3 can be thought of as the buy and sell rate at the optimal ask price  $a^* = \mu^{-1}(\theta^*)$  and the optimal bid price  $b^* = \lambda^{-1}(\theta^*)$ .

**Example 5.1** Consider the following buy and sell functions:

$$\lambda(b) = \alpha + \beta b, \quad (5.4)$$

$$\mu(a) = \gamma - \delta a, \quad (5.5)$$

where

$$\gamma > \alpha > 0, \quad \beta > 0, \quad \delta > 0.$$

Then the above maximization problem can be solved explicitly to yield the optimum  $a^*$  and  $b^*$  as follows:

$$\theta^* = \lambda(b^*) = \mu(a^*) = \frac{\beta\gamma + \alpha\delta}{2(\beta + \delta)}, \quad (5.6)$$

$$a^* = \frac{\gamma - \theta^*}{\delta}, \quad b^* = \frac{\theta^* - \alpha}{\beta}.$$

The optimal bid ask spread is  $a^* - b^*$ . ■

We can explicitly include the transaction cost in the market maker's objective function. Suppose the market maker incurs a positive transaction cost of  $\zeta$  per trade (buy or sell). Then the transaction cost rate is  $2\theta$ , and maximization problem Equation 5.3 changes to

$$\text{Maximize } g(\theta) = f(\theta) - 2\zeta\theta. \quad (5.7)$$

Under the hypothesis of Theorem 5.1,  $g(\cdot)$  is a concave function. Let  $\theta^{**}$  be its maximizer. Concavity of  $f$  implies that

$$\theta^{**} \leq \theta^*.$$

Thus

$$\begin{aligned} a^{**} &= \mu^{-1}(\theta^{**}) \geq \mu^{-1}(\theta^*) = a^* \\ b^{**} &= \lambda^{-1}(\theta^{**}) \leq \lambda^{-1}(\theta^*) = b^* \end{aligned}$$

Thus the presence of the transaction costs increases the bid price, decreases the ask price, lowers the volume, and increases the spread. Clearly, even though the market maker has a complete monopoly in setting the prices, he cannot make infinite profits due to the price sensitivity of the traders.

## 5.2 A Dynamic Model

A clear deficiency of the static model is the assumption that the ask and bid prices do not change with time. This is eminently untrue in the real markets managed by dealers or market makers. In this section we study a dynamic model where the market maker explicitly changes the bid and ask price after every transaction to account for the change in his inventory. The material in this section is based on the work of Amihud and Mendelson [2]. We use the theory of semi-Markov Decision Processes, see Tijms [49].

As in the static model, let  $X(t)$  be the inventory held by the market maker at time  $t$ . We shall assume that  $0 \leq X(t) \leq M$  for a fixed  $M \geq 0$ . (In reality the inventory can be positive or negative. However, since there is no charge to carry the inventory, we can always assume the above, as long as there is an upper and lower bound on the inventory.) The market maker has to decide the ask price  $a_k$ , and the bid price  $b_k$  to be used when the inventory is  $k$ . Let  $\mu(a)$  and  $\lambda(b)$  be as in the previous section.

For a fixed  $[a_1, \dots, a_M]$  and  $[b_0, b_1, \dots, b_{M-1}]$  the inventory process  $\{X(t), t \geq 0\}$  is a birth and death process with birth and death rates given by

$$\lambda_k = \lambda(b_k), \quad \mu_k = \mu(a_k), \quad 0 \leq k \leq M.$$

We assume that  $\mu_0 = \lambda_M = 0$ . Thus the market maker is not allowed to sell if the inventory is 0, or buy if the inventory is  $M$ . Then the net cash flow rate in state  $k$  is given by

$$q_k = R(\mu_k) - C(\lambda_k),$$

where  $R$  and  $C$  are as defined in Equations 5.1 and 5.2. The decision problem of the market maker is thus reduced to finding the bid and ask prices that maximize the long run net cash flow rate. We formulate this as a semi-Markov decision process (SMDP), with average reward criteria.

Since there is a one to one correspondence between  $\lambda_k$  and  $b_k$ , and between  $\mu_k$  and  $a_k$ , we can equivalently treat the birth and death rates as the decision variables for the market maker.

Suppose the inventory is  $k$  at time zero. If the market maker chooses birth rate  $\beta$  and death rate  $\delta$ , he generates a cash flow at rate  $R(\delta) - C(\beta)$  for an  $\exp(\beta + \delta)$  amount of time. At the end of this period the inventory jumps to state  $k + 1$  with probability  $\beta/(\beta + \delta)$  or to state  $k - 1$  with probability  $\delta/(\beta + \delta)$ . The process repeats indefinitely. Clearly the market maker is constrained to use  $\delta = 0$  in state 0 and  $\beta = 0$  in state  $M$  to keep the inventory bounded between 0 and  $M$ . We can now formulate the market maker's decision problem as an SMDP problem as follows. Using the standard results about SMDPs (see Tijms [49]) we get the following version of the optimality equations (called the Dynamic Programming Equations or Bellman Equations)

$$h(0) = \max_{\beta > 0} \left\{ \frac{-C(\beta) - g}{\beta} + h(1) \right\}, \quad (5.8)$$

$$h(k) = \max_{\beta > 0, \delta > 0} \left\{ \frac{R(\delta) - C(\beta) - g}{\beta + \delta} + \frac{\beta}{\beta + \delta} h(k + 1) + \frac{\delta}{\beta + \delta} h(k - 1) \right\}, \quad 1 \leq k \leq M - 1, \quad (5.9)$$

$$h(M) = \max_{\delta > 0} \left\{ \frac{R(\delta) - g}{\delta} + h(M - 1) \right\}. \quad (5.10)$$

These are  $M + 1$  equations in  $M + 2$  unknowns  $\{h(k), 0 \leq k \leq M\}$  and  $g$ . We can reduce this redundancy by setting  $h(0) = 0$ . If there is a solution to the above equations with  $h(0) = 0$ , the total net cash flow over  $[0, t]$  starting from state  $k$  is given by

$$h(k) + tg + o(t),$$

where  $o(t)$  is any function such that  $o(t)/t \rightarrow 0$  as  $t \rightarrow \infty$ . Furthermore, if the minimum on the right hand side of  $h(k)$  is obtained at  $\beta$  and  $\delta$  then it is optimal to use  $\mu_k = \delta$  and  $\lambda_k = \beta$ , and  $g$  represents the long run net cash flow rate of the optimal policy. There are many methods of solving the above optimality equations, such as value iteration and policy iteration. See Tijms [49]. Since this problem has such a special structure, it is possible to develop a specialized numerical procedure to solve it. We first derive the main results in the next theorem, which match with Equations 9 and 11 of Amihud and Medelson [2], which they derive using a different method.

**Theorem 5.2** *Suppose the hypothesis of Theorem 5.1 holds. Then the optimal  $\lambda_k$ 's,  $\mu_k$ 's, and  $g$  are given by the unique solution to*

$$R'(\mu_{k+1}) = C'(\lambda_k), \quad 0 \leq k \leq M - 1, \quad (5.11)$$

$$[R(\mu_k) - \mu_k R'(\mu_k)] - [C(\lambda_k) - \lambda_k C'(\lambda_k)] = g, \quad 0 \leq k \leq M, \quad (5.12)$$

where we have assumed  $\lambda_M = \mu_0 = 0$ .

**Proof:** Let  $[h(0), h(1), \dots, h(M)]$  and  $g$  satisfy the Equations 5.8, 5.9, and 5.10. Suppose the values of  $\beta$  and  $\delta$  that optimizes the right hand side of the  $k$ th Bellman equation are

$$\beta = \lambda_k, \quad \delta = \mu_k, \quad 0 \leq k \leq M.$$

Since these are optimal, the derivatives of the functions that are minimized must be zero at these values. Setting the derivatives with respect to  $\delta$  and  $\beta$  to zero, we get

$$\frac{R'(\mu_k) + h(k-1)}{\lambda_k + \mu_k} = \frac{R(\mu_k) - C(\lambda_k) - g + \lambda_k h(k+1) + \mu_k h(k-1)}{(\lambda_k + \mu_k)^2},$$

$$\frac{-C'(\lambda_k) + h(k+1)}{\lambda_k + \mu_k} = \frac{R(\mu_k) - C(\lambda_k) - g + \lambda_k h(k+1) + \mu_k h(k-1)}{(\lambda_k + \mu_k)^2}.$$

The above equations are valid for  $0 \leq k \leq M$ , assuming  $\lambda_M = \delta_0 = 0$ . From the above equations we can derive

$$C'(\lambda_0) = h(1), \quad R'(\mu_k) + C'(\lambda_k) = h(k+1) - h(k-1), \quad 1 \leq k \leq M-1, \quad R'(\mu_M) = h(M) - h(M-1).$$

Substituting the optimal values of the birth and death rates in the optimality equations, we get

$$\lambda_0 h(0) = -C(\lambda_0) - g + \lambda_0 h(1), \quad (5.13)$$

$$(\lambda_k + \mu_k) h(k) = R(\delta) - C(\beta) - g + \lambda_k h(k+1) + \mu_k h(k-1), \quad 1 \leq k \leq M-1, \quad (5.14)$$

$$\mu_M h(M) = R(\mu_M) - g + \mu_M h(M-1). \quad (5.15)$$

Algebraic manipulations of the above equations yield

$$h(k) = \sum_{i=0}^{k-1} C'(\beta_i), \quad 0 \leq k \leq M.$$

Using this we get Equations 5.11 and 5.12. ■

Once we have  $\lambda_k$ 's and  $\mu_k$ 's, we can easily get the optimal prices  $a_k$ 's and  $b_k$ 's. Note that the Equations 5.11 and 5.12 provide an efficient algorithm to compute the optimal birth and death rates. The next theorem gives important properties of the optimal ask and bid prices.

**Theorem 5.3** *The optimal bid and ask prices satisfy:*

$$b_0 > b_1 > \cdots > b_{M-1},$$

$$a_1 > a_2 > \cdots > a_M,$$

and

$$a_k > b_k, \quad 0 \leq k \leq M.$$

Equivalently,

$$\lambda_0 > \lambda_1 > \cdots > \lambda_{M-1} > 0,$$

$$0 < \mu_1 < \mu_2 < \cdots < \mu_M.$$

**Proof:** See Theorem 3.2 of Amihud and Medelson [2]. ■

Note that the static model essentially solves

$$R'(\theta^*) = C'(\theta^*)$$

and uses  $a^* = \mu^{-1}(\theta^*)$  and  $b^* = \lambda^{-1}(\theta^*)$ . One can also show that the optimum cash flow rate of the market maker using the dynamic policy is in fact worse than that of the market maker in the static model. This seemingly counter-intuitive result is due to the fact that the dynamic market maker is constrained to keep inventory bounded from above and below.

**Example 5.2** Consider the linear buy and sell rate functions given in Equations 5.4 and 5.5. This gives

$$R(\mu) = \frac{\gamma\mu - \mu^2}{\delta},$$

$$C(\lambda) = \frac{\lambda^2 - \lambda\alpha}{\beta}.$$

Equation 5.12 becomes

$$\frac{\mu_k^2}{\delta} + \frac{\lambda_k^2}{\beta} = g, \quad 0 \leq k \leq M$$

and Equation 5.11 reduces to

$$\beta\mu_{k+1} + \delta\lambda_k = \frac{1}{2}(\gamma\beta + \delta\alpha), \quad 0 \leq k \leq M - 1.$$

(Remember  $\lambda_M = \mu_0 = 0$ .) These can be solved numerically in an iterative fashion. Let  $\theta^*$  be the optimal equilibrium rate in the static model as given by Equation 5.6, and  $a^*$  and  $b^*$  be the corresponding optimal ask and bid prices. It can be shown that the optimal bid and ask prices in the dynamic model satisfy:

$$a_1 > a_2 > \cdots > a_M > a^* > b^* > b_0 > b_1 > \cdots > b_{M-1}.$$

We refer the reader to Amihud and Medelson [2] for many more interesting results about this model. As in the static model, it is easy to incorporate transaction costs in the model. What is interesting is that the model suggests that the bid-ask spread is a natural consequence of the profit motive and the monopoly power of the dealer and will exist even if the transaction costs are zero. Note that these models do not take into account the fundamental value of the asset at all. Alternately, the fundamental value is supposed to be reflected in the  $R$  and  $C$  functions, and does not change with time.

## Chapter 6

# Quote-Driven Markets: Dealers as Utility Maximizers

In this chapter we shall continue the study of quote driven markets with market orders. In Chapter 5 we studied a dynamic pricing model of a dealer who sets ask and bid prices in response to his inventory and price sensitive traders to maximize net cash flow rate. This implicitly assumes that the dealer is risk neutral, i.e., has a linear utility function. Also the value of the inventory does not play any role in that model since it is an infinite horizon model and the inventory is never liquidated. In this chapter we shall remove some of these restrictions. Thus we shall study models where the dealer is risk averse and faces finite horizon. We shall find that these models have the flavor of the inventory models and need the tools of Markov Decision Processes (MDPs).

### 6.1 Single Period Model

The material of this section is largely based on Stoll [48]. Consider a dealer with initial inventory  $Q_0$  of a stock, and initial amount  $c_0$  in a cash account. The initial value of the stock is  $V_0$ , and hence the the initial value of the portfolio is

$$W_0 = c_0 + Q_0 V_0.$$

Let  $U(w)$  be the utility to the dealer of the portfolio value  $w$ . Let  $R_0$  be the random rate of return on the stock, i.e., the value of the stock at time 1 is given by  $V_0(1 + R_0)$ . Without loss of generality, assume the rate of return on the cash account is zero. Let  $W_1(q)$  be the value of the portfolio at time 1 if the dealer executes a trade of size  $q$  (buy if  $q > 0$ , sell if  $q < 0$ ). Thus, in the absence of any trades, the value of the port folio at time 1 will be

$$W_1(0) = c_0 + Q_0 V_0(1 + R_0) = W_0 + Q_0 V_0 R_0. \tag{6.1}$$

The dealer deems  $Q_0$  to be optimal, in the sense that he will be perfectly happy to leave it unchanged over one period. In other words,  $c_0$  and  $Q_0$  are such that

$$\mathbf{E}(U(W_1(0))) = \mathbf{E}(U(W_0)). \quad (6.2)$$

Now, the dealer's main job is to provide immediacy, i.e., be willing to buy or sell any demands from the traders. He does this by publishing a schedule  $C(q)$ ,  $-\infty < q < \infty$ , with the following interpretation: he will enter into a transaction of size  $q$  for a fee of  $C(q)$ . Thus if a trade of size  $q$  goes through at time zero, his new portfolio has  $Q_1 = Q_0 + q$  shares of the stock and  $c_1 = c_0 - C(q)$  cash. Its value at time 1 is

$$W_1(q) = c_1 + Q_1 V_0(1 + R_0) = W_1(0) + qV_0(1 + R_0) - C(q). \quad (6.3)$$

We assume that the dealer is willing to enter into any transaction that leaves the expected value of the utility of his portfolio unchanged. That is, he chooses the function  $C(q)$  to ensure

$$\mathbf{E}(U(W_1(q))) = \mathbf{E}(U(W_1(0))). \quad (6.4)$$

Using the first two terms of the Taylor series expansion of  $U(W_1(q))$  around  $w_1(q) = \mathbf{E}(W_1(q))$ , we get

$$U(W_1(q)) = U(w_1(q)) + U'(w_1(q))(W_1(q) - w_1(q)) + \frac{1}{2}U''(w_1(q))(W_1(q) - w_1(q))^2.$$

Taking expectation, we get

$$\mathbf{E}(U(W_1(q))) = U(w_1(q)) + \frac{1}{2}U''(w_1(q))(Q_0 + q)^2 \text{Var}(R_0).$$

Using Equation 6.4 we get

$$U(w_1(q)) + \frac{1}{2}U''(w_1(q))(Q_0 + q)^2 \text{Var}(R_0) = U(w_1(0)) + \frac{1}{2}U''(w_1(0))Q_0^2 \text{Var}(R_0). \quad (6.5)$$

Again using the Taylor expansion of  $U'$  and  $U''$  we get

$$U(w_1(q)) = U(w_1(0)) + U'(w_1(0))(w_1(q) - w_1(0)),$$

(??? Here Stoll mysteriously drops the second derivative term. However, if we include that term, we end up with a quadratic for  $C(q)$ , and the solution does not behave well! ???) and

$$U''(w_1(q)) = U''(w_1(0)).$$

Substituting in Equation 6.5, and simplifying, we get

$$C(q) = qV_0(1 + \mathbf{E}(R_0)) - \frac{1}{2}\rho(w_1(0))((Q_0 + q)^2 - Q_0^2)V_0^2 \text{Var}(R_0), \quad (6.6)$$

where

$$\rho(w) = -\frac{U''(w)}{U'(w)}$$

is the Pratt coefficient of risk aversion at  $w$ . Note that the above equations satisfies  $C(0) = 0$ , that is, the dealer will keep his current portfolio for one more period at no cost.

One can interpret  $C(q)/q$  as the per share ask price at  $q > 0$ ,  $C(-q)/(-q)$  as the bid price per share at  $q < 0$  and  $(C(q) - C(-q))/q$  as the bid-ask spread at trading volume  $q$ . Note that this increases linearly in  $q$  as  $q$  becomes large. It should be noted that in the special case of negative exponential utility function  $U(w) = -\exp(-\rho w)$  with  $\rho > 0$ , and normally distributed  $R_0$ , the result in Equation 6.6 is exact.

Note the differences between the dynamic model of Section 5.2 and this model. In the current model we do not need any model of how the traders behave. We do not need any supply or demand rates at different prices, since the result is conditional: if there is a demand for a trade of size  $q$ , the dealer will be willing to execute it for a fee of  $C(q)$ . The Dealer does not need to know if such a demand will arise or not. He just keeps all contingencies ready to keep the expected utility unchanged.

## 6.2 Multiple Period Model

The extension of the model of Section 6.1 to multiple periods is problematic without invoking a model of what trades actually take place. Stoll [48] does study such an extension. We shall study another model here, which can be thought of as the discrete time version of the model of Amihud and Mendelson [2] with explicit benefit of holding inventory

Suppose the dealer makes decisions at discrete times  $n \geq 0$ . Let  $Q_n$  be the number of shares held by the dealer at time  $n$  and  $R_n$  be the cash dividend on one share of the stock in period  $n$ . Assume that  $\{R_n, n \geq 0\}$  is a sequence of iid random variables with common mean  $r$  and common variance  $\sigma^2$ . At each period  $n$ , the dealer announces an ask price  $a_n$  and a bid price  $b_n$ . Let  $A_n$  be the number of shares he sells at price  $a_n$  in period  $n$ , and  $B_n$  be the number of shares he buys at price  $b_n$  in period  $n$ . The distribution of  $A_n$  depends upon  $a_n$  and that of  $B_n$  depends on  $b_n$ . Let

$$\alpha(a) = \mathbf{E}(A_n | a_n = a), \quad \beta(b) = \mathbf{E}(B_n | b_n = b), \quad n \geq 0.$$

Let  $I_n$  be his income in period  $n$  from these transactions and the dividend. Then we have

$$Q_{n+1} = Q_n + B_n - A_n, \quad I_n = a_n A_n - b_n B_n + R_n Q_n, \quad n \geq 0.$$

Let  $C_n$  be the cash position of the dealer at time  $n$ . Thus

$$C_{n+1} = C_n + I_n.$$



Let  $\delta$  be the discount factor. Then the expected total discounted income (ETDI) of the dealer is given by

$$\mathbf{E}\left(\sum_{n=0}^{\infty} \delta^n I_n\right).$$

The dealer wants to maximize this expected value.

We use the technique of MDPs to solve this problem. Let  $v(q)$  be the optimal ETDI of the dealer if he starts with initial inventory  $q$ . Then the dynamic programming equation is given by

$$v(q) = \sup_{a,b} \{rq + a\alpha(a) - b\beta(b) + \delta\mathbf{E}(v(q + B - A))\}. \quad (6.7)$$

Here  $B - A$  has the same distribution as  $B_n - A_n$  for a given pair  $(a, b)$  for any  $n$ . Now define

$$C(a, b) = \left(a - \frac{r\delta}{1-\delta}\right)\alpha(a) - b\left(1 - \frac{r\delta}{1-\delta}\right)\beta(b). \quad (6.8)$$

Assume that this functions reaches its global maximum at  $(a^*, b^*)$ , and define

$$c^* = C(a^*, b^*). \quad (6.9)$$

The main result is given in the next theorem.

**Theorem 6.1** *The solution to Equation 6.7 is given by*

$$v(q) = \frac{1}{1-\delta}(rq + c^*). \quad (6.10)$$

*It is optimal to use the ask price  $a^*$  and bid price  $b^*$  in each period, regardless of the inventory level.*

**Proof:** Let  $v(\cdot)$  be as given in Equation 6.10. Then

$$\mathbf{E}(v(q + B - A)) = \frac{1}{1-\delta}(rq + \beta(b) - \alpha(a) + c^*).$$

Hence

$$rq + a\alpha(a) - b\beta(b) + \delta\mathbf{E}(v(q + B - A)) = rq + \left(a - \frac{\delta}{1-\delta}\right)\alpha(a) - b\left(1 - \frac{\delta}{1-\delta}\right)\beta(b) + \frac{\delta}{1-\delta}c^*.$$

The global maximum over over all  $a, b$  of the above is given by  $\frac{1}{1-\delta}(rq + c^*)$ . This shows that the value function  $v(q)$  given in Equation 6.10 satisfies the optimality equation 6.7. Hence, from the results of MDPs, we see that it is the unique solution. ■

It is surprising that the optimal ask and bid prices are independent of the inventory level, since the pricing decisions affect the purchases and the sales, which affect the inventory in the next period, which affects the income. This is closer to the result given in Garman [14], than Amihud and Mendelson [2]. This is the consequence of the risk neutrality of the dealer and the fact that there

is no restriction on the inventory.

We can generalize the above model to a dealer with utility function  $U$ , who is interested in maximizing

$$\mathbf{E}\left(\sum_{n=0}^{\infty} \delta^n U(I_n)\right). \quad (6.11)$$

Note that this is different than maximizing

$$\mathbf{E}\left(U\left(\sum_{n=0}^{\infty} \delta^n I_n\right)\right). \quad (6.12)$$

MDPs can be used to maximizing the former function, but not the latter. In the former case, the optimality equation becomes

$$v(q) = \sup_{a,b} \{\mathbf{E}(U(Rq + aA - bB)) + \delta \mathbf{E}(v(q + B - A))\}. \quad (6.13)$$

**Question:** Try to solve the above equation for  $U(w) = -e^{-\rho w}$ ,  $R \sim N(r, \sigma_R^2)$ ,  $A \sim N(\alpha(a), \sigma_A^2)$ , and  $B \sim N(\beta(b), \sigma_B^2)$ . ■

Although the MDPs cannot be used to maximize the objective function in Equation 6.12, they can be used to maximize

$$\mathbf{E}\left(U\left(\sum_{n=0}^N I_n\right)\right)$$

for a fixed  $N$ , that is, maximize the utility of the wealth at time  $N$ . This leads us to a finite horizon problem that is analogous to the the discrete version of the model considered in Ho and Stoll [26]. Let  $N$  be the time horizon, and assume that the dealer can liquidate his entire inventory of the stock at price  $p$  at time  $N$ . Thus the value of his portfolio at time  $N$  is given by  $C_N + pQ_N$ . The dealer sets the bid and set prices over  $\{0, 1, \dots, N-1\}$  to maximize  $\mathbf{E}(U(C_N + pQ_N))$ . Let  $v_n(c, q)$  be the optimal value function if  $C_n = c$  and  $Q_n = q$ . The optimality equations are

$$v_n(c, q) = \sup_{a,b} \{\mathbf{E}(v_{n+1}(c + R_n q + aA_n - bB_n, q + B_n - A_n))\}, \quad 0 \leq n \leq N-1, \quad (6.14)$$

with the terminal value function

$$v_N(c, q) = \mathbf{E}(U(c + pq)). \quad (6.15)$$

In theory, we can solve the Equation 6.14 in a backward fashion, starting from  $v_N$ . Let  $a_n^*(c, q)$  and  $b_n^*(c, q)$  be the values of  $a$  and  $b$  that maximize the right hand side in Equation 6.14. Then the optimal policy for the dealer at time  $n$  with  $C_n = c$  and  $Q_n = q$  is to use the ask price  $a_n^*(c, q)$  and bid price  $b_n^*(c, q)$ . In practice, computing these optimal prices is computationally hard. We treat the problem at time  $N-1$  to gain some insight into the solutions.

We have

$$\begin{aligned}
v_{N-1}(c, q) &= \sup_{a, b} \{ \mathbf{E}(v_N(c + R_{N-1}q + aA_{N-1} - bB_{N-1}, q + B_{N-1} - A_{N-1})) \} \\
&= \sup_{a, b} \{ \mathbf{E}(U(c + R_{N-1}q + aA_{N-1} - bB_{N-1} + p(q + B_{N-1} - A_{N-1}))) \} \\
&= \sup_{a, b} \{ \mathbf{E}(U(c + R_{N-1}q + (a - p)A_{N-1} - (b - p)B_{N-1})) \}.
\end{aligned}$$

Let

$$W = c + R_{N-1}q + (a - p)A_{N-1} - (b - p)B_{N-1},$$

and

$$w = \mathbf{E}(W) = c + rq + (a - p)\alpha(a) - (b - p)\beta(b).$$

Using Taylor expansion of  $U$  around  $w$ , we get

$$U(W) = U(w) + U'(w)(W - w) + \frac{1}{2}U''(w)(W - w)^2.$$

Taking expectations, we get

$$\mathbf{E}(U(W)) = U(w) + \frac{1}{2}U''(w)(q^2\text{Var}(R) + (a - p)^2\text{Var}(A) + (b - p)^2\text{Var}(B)),$$

where we have assumed that  $R_{N-1}$ ,  $A_{N-1}$  and  $B_{N-1}$  are independent random variables. Using

$$U(w) = U(c) + U'(c)(rq + (a - p)\alpha(a) - (b - p)\beta(b))$$

and

$$U''(w) = U''(c)$$

we get

$$\begin{aligned}
\mathbf{E}(U(W)) &= U(c) + U'(c)(rq + (a - p)\alpha(a) - (b - p)\beta(b)) \\
&+ \frac{1}{2}U''(c)(q^2\text{Var}(R) + (a - p)^2\text{Var}(A) + (b - p)^2\text{Var}(B)).
\end{aligned}$$

This can be simplified to get

$$\begin{aligned}
\mathbf{E}(U(W)) &= U(c) + (rq + (a - p)\alpha(a) - (b - p)\beta(b)) \\
&- \frac{1}{2}\rho(c)(q^2\text{Var}(R) + (a - p)^2\text{Var}(A) + (b - p)^2\text{Var}(B)).
\end{aligned}$$

Here  $\rho(c) = -U''(c)/U'(c)$  is the coefficient of risk aversion of the dealer at  $c$ . Maximizing the above function we get  $a_{N-1}^*$  and  $b_{N-1}^*$  as a function of  $c$  and  $q$ . This is the optimal ask bid pair at time  $N - 1$  for a given state  $(c, q)$ .

**Question:** Carry out the above maximization for linear decreasing  $\alpha(a)$  and linear increasing  $\beta(b)$ . Assume that the variances of  $A$  and  $B$  do not depend on  $a$  and  $b$ . ■

**Remark:** A related model with more institutional detail is studied by Ohara and Oldfield [37].



What happens when there are multiple dealers competing for the traders' business? This leads to sequential theoretic models. See Ho and Stoll [25], [27]. We do not go into the details here.

## Chapter 7

# Quote-Driven Markets: Transaction Costs

In this chapter we investigate another source of bid-ask spread: the transaction costs. Typically the dealer incurs a cost  $c$  per trade in the form administrative costs. He recovers these costs by buying low at the bid prices and selling high at the ask prices. We study several models that attempt to explain this aspect of bid-ask spread.

### 7.1 Roll Model

We begin with a simple model called the Roll model, initially studied by Roll [41]. Let  $V_k$  be the fundamental value (or the efficient price) of a security at time  $k$ . We shall assume that  $\{V_k, k \geq 0\}$  is a random walk with zero drift:

$$V_{k+1} = V_k + U_k, \quad k \geq 0$$

where  $\{U_k, k \geq 0\}$  are iid random variables with zero mean and variance  $\sigma^2$ . The bid and ask prices at time  $k$  are then

$$b_k = V_k - c, \quad a_k = V_k + c. \quad (7.1)$$

Let  $Q_k = 1$  if the  $k$ th trade is a sale by the dealer, and  $-1$  if it is a purchase by the dealer.  $Q_k$  is called the direction of the trade. We assume that  $\{Q_k, k \geq 0\}$  are iid random variables with

$$P(Q_k = 1) = P(Q_k = -1) = 1/2.$$

The transaction price  $P_k$  is the price at which the  $k$ th trade occurs. Thus

$$P_k = V_k + Q_k c. \quad (7.2)$$

Note that the  $\{V_k, k \geq 0\}$  is not directly observable, but  $\{P_k, k \geq 0\}$  is. This model of the price movements is called the Roll model, studied in Roll [41]. It is described by two parameters:  $c$  and

$\sigma^2$ . Can we estimate  $c$  and  $\sigma^2$  from the price data? This is done by studying the stochastic process of price differences

$$\Delta_k = P_k - P_{k-1} \quad k \geq 1. \quad (7.3)$$

The main result is given in the next theorem.

**Theorem 7.1**  $\{\Delta_k, k \geq 1\}$  is a stationary process with the auto-covariance function given by

$$\gamma_j = \text{Cov}(\Delta_k, \Delta_{k+j}) = \begin{cases} 2c^2 + \sigma^2 & \text{for } j = 0 \\ -c^2 & \text{for } j = 1 \\ 0 & \text{for } j \geq 2 \end{cases}$$

**Proof:** We have, for  $k \geq 1$ ,

$$\Delta_k = P_k - P_{k-1} = V_k + Q_k c - V_{k-1} - Q_{k-1} c = U_{k-1} + (Q_k - Q_{k-1})c.$$

The result for  $j = 0$  follows from

$$\text{Cov}(\Delta_k, \Delta_k) = \text{Var}(U_{k-1} + (Q_k - Q_{k-1})c)$$

and the fact that  $U$ 's and  $Q$ 's are independent and  $Q_k$  and  $Q_{k-1}$  are independent with variance 1. Similarly, for  $j = 1$  we get

$$\text{Cov}(\Delta_k, \Delta_{k+1}) = \text{Cov}(U_{k-1} + (Q_k - Q_{k-1})c, U_k + (Q_{k+1} - Q_k)c)$$

The result follows from straight forward calculations. Similarly, one can show that

$$\text{Cov}(\Delta_k, \Delta_{k+j}) = 0, \quad \text{for } j \geq 2. \quad \blacksquare$$

One conclusion of this model is that the consecutive price increments are negatively correlated. Thus the prices do not form a random walk (or a martingale) if trading costs are present, even though the efficient prices (which are unobservable) form a martingale. One can also use the above theorem to get

$$c = \sqrt{-\gamma_1}, \quad \sigma^2 = \gamma_0 + 2\gamma_1.$$

Since we can estimate  $\gamma_0$  and  $\gamma_1$  from the price data, we can estimate the parameters  $c$  and  $\sigma^2$ .

## 7.2 Roll Model: Correlated Transactions

The Roll model of Section 7.1 assumes that  $\{Q_k, k \geq 0\}$  are iid random variables. Choi et al [7] consider a slight extension by assuming that  $\{Q_k, k \geq 0\}$  is a stationary Discrete Time Markov Chain (DTMC) on  $\{-1, 1\}$  with transition probability matrix described by a single parameter  $\delta \in [0, 1]$  as follows:

$$\begin{bmatrix} \delta & 1 - \delta \\ 1 - \delta & \delta \end{bmatrix}.$$

One motivation for this model is that the trades on an exchange are observed to be correlated, with buys following buys, and sales following sales, more often than not. This may be an indication of herd mentality or large traders splitting their orders into smaller ones to get a better price. With this model Theorem 7.1 is modified as follows:

**Theorem 7.2**  $\{\Delta_k, k \geq 1\}$  is a stationary process with the auto-covariance function given by

$$\gamma_j = \text{Cov}(\Delta_k, \Delta_{k+j}) = \begin{cases} 4c^2(1-\delta) + \sigma^2 & \text{for } j = 0 \\ -4c^2(1-\delta)^2(2\delta-1)^{j-1} & \text{for } j \geq 1 \end{cases}$$

**Proof:** We have

$$\begin{bmatrix} \delta & 1-\delta \\ 1-\delta & \delta \end{bmatrix}^n = \frac{1}{2} \begin{bmatrix} 1 + (2\delta-1)^n & 1 - (2\delta-1)^n \\ 1 - (2\delta-1)^n & 1 + (2\delta-1)^n \end{bmatrix}.$$

Since  $\{Q_k, k \geq 0\}$  is stationary,

$$\text{P}(Q_k = 1) = \text{P}(Q_k = -1) = 1/2, \quad k \geq 0.$$

It can be shown that

$$\text{Cov}(Q_k, Q_{k+j}) = (2\delta-1)^j, \quad j \geq 0.$$

The theorem follows from this. ■

Thus one step correlation is negative, but further auto covariance may alternate, or stay negative, depending on whether  $\delta > 1/2$  or  $\delta < 1/2$ . Note that this model reduces to Roll model if  $\delta = 1/2$ . Choi et al [7] perform detailed statistical analysis to fit the model using data from NYSE trades.

### 7.3 Roll Model: Trade Influence

In practice if one trader buys ( $Q_k = 1$ ), the other traders perceive it as an indication that the fundamental value has moved up, and vice versa. We follow Chapter 8 of Hasbrouck [22], and model this aspect as follows:

$$V_k = V_{k-1} + W_k, \quad W_k = \lambda Q_k + U_k, \quad k \geq 0, \quad (7.4)$$

where  $\lambda > 0$  is an unknown constant. Thus when trader buys, it gives an upward nudge to the fundamental value, and when he sells, it gives a downward nudge. Assume that  $\{U_k, k \geq 0\}$  and  $\{Q_k, k \geq 0\}$  are as in Section 7.1. If we set  $\lambda = 0$ , we get back the original Roll model of Section 7.1. As before, the ask and bid prices are given by Equation 7.1, and the trade prices are as in Equation 7.2. The model now has three parameters  $\lambda, c, \sigma^2$ . Let  $\Delta_k$  be as in Equation 7.3. Then we have the following theorem, whose proof we omit.

**Theorem 7.3**  $\{\Delta_k, k \geq 1\}$  is a stationary process with the auto-covariance function given by

$$\gamma_j = \text{Cov}(\Delta_k, \Delta_{k+j}) = \begin{cases} c^2 + (c + \lambda)^2 + \sigma^2 & \text{for } j = 0 \\ -c(c + \lambda) & \text{for } j = 1 \\ 0 & \text{for } j \geq 2 \end{cases}$$

Thus we cannot estimate all three parameters from the price data. We can estimate the variance of  $W_k$ , since we have

$$\text{Var}(W_k) = \lambda^2 + \sigma^2 = \gamma_0 + 2\gamma_1,$$

and  $\gamma_0$  and  $\gamma_1$  can be estimated.

## 7.4 Roll Model: Further Extensions

We could now combine the correlated trade model of Section 7.2 with the trade influence Roll model of Section 7.3. In this case we have

$$Z_k = Q_k - \mathbb{E}(Q_k | Q_{k-1}, Q_{k-2}, \dots) = Q_k - (2\delta - 1)Q_{k-1}.$$

The process  $\{Z_k, k \geq 0\}$  is called the trade innovation process. It indicates how much the observed trade deviates from the expected trade. It stands to reason that the fundamental value is affected by the trade innovation process, and not the actual trade process. Hence we modify Equation 7.4 as follows:

$$V_k = V_{k-1} + W_k, \quad W_k = \lambda Z_k + U_k, \quad k \geq 0, \quad (7.5)$$

Let the price process and the price increment process be as in Equations 7.2 and 7.3. The probabilistic structure of the price increment process is given in the next theorem.

**Theorem 7.4**  $\{\Delta_k, k \geq 1\}$  is a stationary process with the auto-covariance function given by

$$\gamma_j = \text{Cov}(\Delta_k, \Delta_{k+j}) = \begin{cases} c^2 + (c + \lambda)^2 - 2(2\delta - 1)(c^2 + (\lambda + c)\lambda(2\delta - 1)) + \sigma^2 & \text{for } j = 0 \\ -c(c + \lambda)(2\delta - 1)^{j-1} + c(\lambda + 2c)(2\delta - 1)^j + c(\lambda - c)(2\delta - 1)^{j+1} - \lambda c(2\delta - 1)^{j+2} & \text{for } j \geq 1 \end{cases}$$

One can estimate the  $\gamma$ 's to estimate the model parameters  $c$ ,  $\lambda$ ,  $\delta$ .

Hasbrouck [22], in Chapter 9, Section 2, does a similar analysis assuming that  $\{Q_k, k \geq 0\}$  is a MA(1) process (Moving Average of order 1). Madhavan et al [33] analyze a similar model assuming that  $\{Q_k, k \geq 0\}$  is an AR(1) process (Auto-regressive of order 1). They also report extensively on the statistical analysis of the data for parameter estimation.



In the models so far we have always assumed that the size of each trade is 1. Glosten and Harris [15] consider a model that includes the size of a trade as a variable. We now define  $Q_k = 1$  if the  $k$ th trade is at the ask price, and  $-1$  if it is at the bid price. Let  $S_k$  be the size (unsigned) of the  $k$ th trade, say the number of shares traded. The Glosten-Harris models can be written as

$$\begin{aligned} V_k &= V_{k-1} + W_k, \\ W_k &= U_k + Q_k(\lambda_0 + \lambda_1 S_k), \\ P_k &= V_k + Q_k(c_0 + c_1 S_k). \end{aligned}$$

Thus the fundamental value is affected by the size and the direction of the trade through parameters  $\lambda_0$  and  $\lambda_1$ , while the trading cost has a fixed component  $c_0$  and a linear variable component  $c_1$ .

Several researchers have extended Roll model to multiple securities using vector time series. We shall not go into that literature here. The parameter  $\lambda$  can be thought as capturing the influence of the trade information, as distinct from the public information captured via  $U_k$ 's. This creates an interesting question about the influence of information on the dealer and trader strategies. There is a large literature in this area. We shall study it in the next chapter.

## 7.5 Multivariate Roll Model

Now consider a simple extension of Roll model to multiple securities. Let  $V_{ik}$  be the fundamental value (or the efficient price) of security  $i$  at time  $k$  ( $i = 1, 2, \dots, m$ ), and  $Q_{ik}$  be the signed trade in security  $i$  at time  $k$ . The multivariate Roll model is given by

$$\begin{aligned} V_{i,k+1} &= V_{ik} + U_{ik}, \quad k \geq 0, \\ P_{ik} &= V_{ik} + cQ_{ik}, \quad k \geq 0. \end{aligned}$$

where  $\{U_{ik}, k \geq 0\}$  are uncorrelated random variables with zero mean and variance  $\sigma_i^2$ . Similarly  $\{Q_{ik}, k \geq 0\}$  are uncorrelated random variables. These two may or may not depend on each other. Notice that the transaction cost  $c$  does not depend on the security. The observations are

$$\Delta_{ik} = P_{ik} - P_{i,k-1} \quad k \geq 1. \tag{7.6}$$

Let  $\Delta_k = [\Delta_{1k}, \Delta_{2k}, \dots, \Delta_{mk}]$  be the vector of price differences at time  $k$ . Thus  $\{\Delta_k, k \geq 1\}$  is a multivariate MA(1) time series as described in Section 3.2.

**Example 7.1** Suppose  $m = 2$ ,  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ,  $E(Q_{ik}) = 0$ ,  $\text{Var}(Q_{ik}) = 1$ ,  $E(Q_{1k}Q_{2k}) = \rho$ . Then we have

$$\Gamma(0) = \begin{bmatrix} 2c^2 + \sigma^2 & 2\rho c^2 \\ 2\rho c^2 & 2c^2 + \sigma^2 \end{bmatrix}, \tag{7.7}$$

$$\Gamma(1) = \Gamma(-1) = \begin{bmatrix} -c^2 + \sigma^2 & -\rho c^2 \\ -\rho c^2 & -c^2 \end{bmatrix}, \quad (7.8)$$

and  $\Gamma(k) = 0$  for  $|k| \geq 2$ . Thus  $\{\Delta_k, k \geq 1\}$  is a multivariate MA(1) time series

$$\Delta_k = \epsilon_k + \Theta \epsilon_{k-1}.$$

We have

$$\Gamma(0) = \Sigma_\epsilon + \Theta \Sigma_\epsilon \Theta',$$

$$\Gamma(1) = \Theta \Sigma_\epsilon.$$

Using Equations 7.7 and 7.8 we can compute the MA(1) parameters  $\Theta$  and  $\Sigma_\epsilon$ .

## Chapter 8

# Quote-Driven Markets: Informed Traders

In the last two chapters we studied two possible explanations for the existence of bid-ask spread. In Chapter 5, we studied models that explained the bid-ask spread as the consequence of the dealer managing his inventory. In Chapter 6, we studied that explained the same phenomena as a consequence of the risk aversion of the dealer. In Chapter 7 we studied models that explained the bid-ask spread as a consequence of the transaction costs. In this chapter we study yet another possible explanation: the existence of informed traders. The informed traders trade with the dealer only when their private information guarantees them profits from the trade. So the dealer always loses to the informed traders. He makes up for this loss by trading with the uninformed traders (also called the liquidity traders). In order for this to generate profits for the dealer, he sets up a bid-ask spread, even if he is risk neutral.

### 8.1 Single-Trade Copeland-Galai Model

We begin with a very simple model of a single trade. It is based primarily on the model studied by Copeland and Galai [10], which is the first formalization of the concept of “informed traders” mentioned by Bagehot [3]. Suppose the initially value of one share the stock is  $V_0$ . Let  $V$  be the value of the stock at the time of arrival of the first trade. Let  $f$  be the density of the random variable  $V$ . For simplicity, we assume the trade is for one share. The arriving trader is an informed trader with probability  $\theta$  and an uninformed trader with probability  $1 - \theta$ . An informed trader knows the actual value of  $V$ , while the uninformed trader and the dealer knows only the density  $f$ . The dealer posts his ask price  $a$  and bid price  $b < a$  after the trade arrives, but before knowing whether it is a buy or sell. The dealer does not know if the trader is informed or not. After the ask and bid prices are declared by the dealer, the trade may or may not take place. We assume that the dealer can liquidate his shares for  $V$  per share after the trade is executed.

We study the informed trader's behavior first. He knows  $V = v$ . If  $v < b$  the informed trader sells one share to the dealer, if  $v > a$ , the trader buys one share from the dealer, and if  $b \leq v \leq a$ , there is no trade. Thus the expected profit of the dealer is

$$P_i = - \int_0^b (b - v)f(v)dv - \int_a^\infty (v - a)f(v)dv.$$

Note that this is always negative. Thus the dealer always loses to an informed trader.

Next we study the behavior of an uninformed trader. Let  $\alpha(a)$  be the probability that the trader will buy at price  $a$ ,  $\beta(b)$  be the probability that the trader will sell at price  $b$ , and  $1 - \alpha(a) - \beta(b)$  be the probability that the uninformed trader does not trade at all. We assume that these functions are known to the dealer. The expected profit of the dealer from this trade is

$$P_u = \mathbf{E}[\alpha(a)(a - V) + \beta(b)(V - b)] = a\alpha(a) - b\beta(b) + \mathbf{E}(V)(\beta(b) - \alpha(a)).$$

The expected net profit of the dealer is thus

$$P(a, b) = \theta P_i + (1 - \theta)P_u.$$

If the dealer has a monopoly, he chooses  $a$  and  $b$  to maximize the expected profits. If the dealer is in a competitive market with many other dealers, he will be forced to use  $a$  and  $b$  where his expected net profits is zero after the transaction. Let  $A$  denote the event that the trade is at the ask price (sale by the dealer), and  $B$  be the event that the trade is at the bid price (purchase by the dealer). Then the ask and bid prices are chosen to satisfy

$$a = \mathbf{E}(V|A) = \frac{\mathbf{E}(V; V > a)\theta + \mathbf{E}(V)\alpha(a)(1 - \theta)}{\theta P(V > a) + (1 - \theta)\alpha(a)},$$

$$b = \mathbf{E}(V|B) = \frac{\mathbf{E}(V; V < b)\theta + \mathbf{E}(V)\beta(b)(1 - \theta)}{\theta P(V < b) + (1 - \theta)\beta(b)}.$$

One interesting case arises if we assume  $\alpha(a) = P(V > a)$ ,  $\beta(b) = P(V < b)$ . Then the above maximization problem reduces to maximizing

$$P(a, b) = -\theta \left[ \int_0^b (b - v)f(v)dv + \int_a^\infty (v - a)f(v)dv \right] \\ + (1 - \theta) [aP(V > a) - bP(V < b) + \mathbf{E}(V)(P(V < b) - P(V > a))].$$

For a monopolist dealer, this is maximized at  $a$  and  $b$  satisfying

$$(1 - \theta)(a - \mathbf{E}(V))f(a) = 1 - F(a), \quad (1 - \theta)(\mathbf{E}(V) - b)f(b) = F(b).$$

For example, if  $V \sim U(0, 1)$ , we get the optimal ask and bid prices as

$$a = \frac{3 - \theta}{2(2 - \theta)}, \quad b = \frac{1 - \theta}{2(2 - \theta)}.$$

The spread is

$$a - b = \frac{1}{2 - \theta}.$$

If every body is an informed trader,  $\theta = 1$ , and the trader sets  $a = 1$ ,  $b = 0$ , i.e., he does not trade! If nobody is informed,  $\theta = 0$ , and the trader sets  $a = 3/4$ , and  $b = 1/4$ , that is, the spread is  $1/2$ .

For a dealer in a competitive market, the zero expected profit conditions yield

$$a = \frac{1}{2 - \theta}, \quad b = \frac{1 - \theta}{2 - \theta}.$$

The spread is

$$a - b = \frac{\theta}{2 - \theta}.$$

If every body is an informed trader,  $\theta = 1$ , and the trader sets  $a = 1$ ,  $b = 0$ , i.e., he does not trade! If nobody is informed,  $\theta = 0$ , and the trader sets  $a = 1/2$ , and  $b = 1/2$ , that is, the spread is zero. The ask price is smaller than that in the monopolistic market, and the bid price is higher, implying the spreads are smaller in the competitive markets. This is to be expected.

The main conclusion of this simple model is that the bid ask spread arises even if the dealer is risk neutral, simply because he has to offset the losses to the informed traders.

**Question:** Can one use the models of “information” developed in Chapter 4, where every trader receives a signal about the true value  $V$ ? The more informed the trader, the less is the variance of this signal. ■

## 8.2 Sequential Glosten-Milgrom Model

Next we extend the model of Section 8.1 to a sequence of  $T \geq 1$  trades. Most of the results in this section are based on Glosten and Milgrom [16]. Let  $V$  be the “true” value of the security. It is a random variable with a known density  $f$ . Each trade involves one unit of the security. The dealer posts an ask price  $a_k$  and bid price  $b_k$  before the  $k$ th trade. If a trade takes place at the ask price (dealer sells) we call it a type  $A$  trade, and if it takes place at the bid, we call it a type  $B$  trade. Let  $Q_k$  be the type of the  $k$ th trade. Let define the two slightly different market information sets (or histories)

$$H_0 = \phi, \quad H_k = \{(a_j, b_j, Q_j) : 0 \leq j \leq k - 1\}, \quad k \geq 1$$

$$\hat{H}_k = H_k \cup \{(a_k, b_k)\}, \quad k \geq 0.$$

Let  $P_k$  and  $E_k$  denote the probability and expectation conditioned on  $H_k$ , and  $\hat{P}_k$  and  $\hat{E}_k$  denote the probability and expectation conditioned on  $\hat{H}_k$ .

The sequence of events is as follows: Just before the  $k$ th trade, the dealer knows the history  $H_k$ . Based on this he computes  $f_k$ , the conditional pdf of  $V$ . Using this and his knowledge of the trader behavior (to be specified later) he computes the ask price  $a_k$  and bid price  $b_k$ . This generates the history  $\hat{H}_k$ , which the traders use to generate the  $k$ th trade, thus determining  $Q_k$ . The history  $H_k$  can now be updated to  $H_{k+1}$ , and the process repeats.

Now we describe the behavior of the traders. The trader trading at time  $k$  is an informed trader with probability  $\theta$  and an uninformed trader with probability  $1 - \theta$ . We shall assume that this does not change with time or the evolution of the market. The dealer does not know if the trader is informed or uninformed. The  $k$ th trader knows the history  $\hat{H}_k$ , i.e., he knows  $a_k$  and  $b_k$ , and all the history the dealer used to set these prices. Let  $v_k^u$  be the uninformed traders' estimate of  $V$  just before the  $k$ th trade, and  $v_k^i$  be the informed traders' prior estimate  $V$  just before the  $k$ th trade. We shall specifically assume that  $v_k^u = \hat{E}_k(V)$ , and  $v_k^i = V$ , the "true" value of the security. Let  $\alpha^u(a, v)$  be the probability that the uninformed trader buys when the ask price is  $a$  and his estimate of  $V$  is  $v$ , and  $\beta^u(b, v)$  be the probability that he sells when the bid price is  $b$  and his estimate of  $V$  is  $v$ . Define  $\alpha^i(a, v)$  and  $\beta^i(b, v)$  in a similar fashion for the informed trader. In our example, we shall assume that the informed trader buys if  $V > \text{ask}$ , and sells if  $V < \text{bid}$ . That is,

$$\beta^i(b, v) = 1_{\{v < b\}}, \quad \alpha^i(a, v) = 1_{\{v > a\}}.$$

For the uninformed trader we shall assume that  $\alpha^u(a, v)$  increases in  $v$  and decreases in  $a$ , while  $\beta^u(b, v)$  decreases in  $v$  and increases in  $b$ . Then we have

$$\hat{P}_k(Q_k = A|V = v) = c_k[\theta 1_{\{v > a_k\}} + (1 - \theta)\alpha^i(a_k, \hat{E}_k(V))] \quad (8.1)$$

$$\hat{P}_k(Q_k = A) = c_k[\theta \hat{P}_k(V > a_k) + (1 - \theta)\alpha^i(a_k, \hat{E}_k(V))] \quad (8.2)$$

$$\hat{P}_k(Q_k = B|V = v) = c_k[\theta 1_{\{v < b_k\}} + (1 - \theta)\beta^u(b_k, \hat{E}_k(V))], \quad (8.3)$$

$$\hat{P}_k(Q_k = B) = c_k[\theta \hat{P}_k(V < b_k) + (1 - \theta)\beta^u(b_k, \hat{E}_k(V))], \quad (8.4)$$

where  $c_k$  is such that

$$\hat{P}_k(Q_k = A) + \hat{P}_k(Q_k = B) = 1.$$

We assume that the dealer sets his ask and bid prices as follows:

$$a_k = \hat{E}_k(V|Q_k = A) = \frac{c_k[\theta \hat{E}_k(V; V > a_k) + (1 - \theta)\hat{E}_k(V)\alpha^i(a_k, \hat{E}_k(V))]}{\hat{P}_k(Q_k = A)}, \quad (8.5)$$

$$b_k = \hat{E}_k(V|Q_k = B) = \frac{c_k[\theta \hat{E}_k(V; V < b_k) + (1 - \theta)\hat{E}_k(V)\beta^i(b_k, \hat{E}_k(V))]}{\hat{P}_k(Q_k = B)}. \quad (8.6)$$

The above equations say that the ask price is what the dealer thinks the stock is worth if a trade occurs at the ask price, and bid price is what the stock is worth if a trade occurs at the bid price. This is the reservation (or regret-free) price of the stock for the dealer for the two trades. The

dealer would like to charge prices better than these, but is prevented from doing so by competition in the market. This implies that if there are multiple  $a_k$ 's that satisfy Equation 8.5, we must choose the smallest, and if there are multiple  $b_k$ 's that satisfy Equation 8.6, we must choose the largest. The bid-ask spread just before the  $k$ th trade is given by

$$a_k - b_k = \hat{E}_k(V|Q_k = A) - \hat{E}_k(V|Q_k = B).$$

We next show how the new pdf  $f_{k+1}$  is computed from  $f_k$  by incorporating the information about  $Q_k$ ,  $a_k$  and  $b_k$ . Using Bayes' theorem, we get

$$f_{k+1}(v) = \begin{cases} f_k(v)\hat{P}_k(Q_k = A|V = v)/\hat{P}_k(Q_k = A) & \text{if } Q_k = A, \\ f_k(v)\hat{P}_k(Q_k = B|V = v)/\hat{P}_k(Q_k = B) & \text{if } Q_k = B. \end{cases} \quad (8.7)$$

This shows that  $\{f_k, k \geq 0\}$  is a Markov chain on the set of all distributions. Using this we can compute the new estimates  $v_{k+1}^i$  and  $v_{k+1}^u$ . In our example, since the informed trader knows the actual value of the security, the new trade provides no additional information to him. Also, the uniformed trader simply compute the new expected value of  $V$  by using the updated pdf  $f_{k+1}$ . Hence

$$v_{k+1}^i = V, \quad v_{k+1}^u = \hat{E}_{k+1}(V).$$

The process now repeats for the  $(k + 1)$ st trade. Note that  $c_k$  is canceled from the numerator and the denominator in Equations 8.5, 8.6 and 8.7. Hence we can safely assume that  $c_k$  is equal to 1 in the above, and do not need to compute it.

One can show that

$$a_k \geq E_k(V) \geq b_k,$$

and that, if  $V = v$ ,

$$\lim_{k \rightarrow \infty} a_k \downarrow v, \quad \lim_{k \rightarrow \infty} b_k \uparrow v. \quad (8.8)$$

Thus as more and more trades take place, the spread reduces to zero, and the ask and bid prices converge to the "true" value of the security. One can also show that

$$\lim_{k \rightarrow \infty} \hat{P}_k(Q_k = A) = \lim_{k \rightarrow \infty} \hat{P}_k(Q_k = B),$$

that is, in the limit the buy and sell orders match, and the drift in the inventory goes to zero.

**Question:** Fill in the details from the relevant theorems in [16]. ■

Now define

$$p_k = \begin{cases} a_k & \text{if } Q_k = A, \\ b_k & \text{if } Q_k = B. \end{cases} \quad (8.9)$$

We can think of  $p_k$  as the market price after the  $k$ th trade. The main result is given in the following theorem

**Theorem 8.1**  $\{p_k, k \geq 0\}$  is a Martingale with respect to  $H_k$ .

**Proof:** Using Equations 8.9, 8.5 and 8.6, we see that

$$p_k = \mathbf{E}(V|H_k).$$

Here  $\mathbf{E}$  is expectation with respect to  $f$ . Since  $H_{k+1}$  includes  $H_k$ , we get

$$\mathbf{E}(p_{k+1}|H_k) = \mathbf{E}(\mathbf{E}(V|H_{k+1})|H_k) = \mathbf{E}(V|H_k) = p_k.$$

This proves the Theorem. ■

From Equation 8.8 it follows that, given  $V = v$ ,

$$\lim_{k \rightarrow \infty} p_k = v.$$

Thus the transaction price eventually converges to the true price.

Another important consequence of this model is that the price increments are un-correlated. This contradicts the results of Roll model of Chapter 7, which implies that the successive price increments are negatively correlated. This is a consequence of the absence of transaction cost. If we include a transaction cost  $c$  in the model, the ask and bid prices are changed to

$$\tilde{a}_k = a_k + c, \quad \tilde{b}_k = b_k - c.$$

Thus the bid-ask spread is now

$$\tilde{a}_k - \tilde{b}_k = a_k - b_k + 2c.$$

Thus the bid-ask spread has two components:  $2c$  is the transactional-cost component, while  $a_k - b_k$  is the asymmetric information (or adverse selection) component. If we define

$$Z_k = \begin{cases} 1 & \text{if } Q_k = A, \\ -1 & \text{if } Q_k = B, \end{cases}$$

we can write the transaction price

$$\tilde{p}_k = p_k + cZ_k = \mathbf{E}_0(V|H_k) + cZ_k.$$

One can show that the consecutive increments of this transaction price process has negative covariance. See Glosten and Milogram [16] for more details.

**Remark:** We have made one important assumption: we have assumed that this is Stackelberg game where the dealer is the leader: he knows the rules used by the traders, and sets his prices accordingly. This allows to assume that all the players in the market base their calculations for the  $k$ th trade on  $H_k$ . However, the more realistic model may be a symmetric Nash game. In such a model the traders set their trading strategy in anticipation of how the dealer will set prices, and vice versa. Such a sequential game theoretic model is too complicated. We have seen some game theoretic models in Chapter 4. We shall return to them in the next chapter.



## Chapter 9

# Order-Driven markets: Dynamic Programming Models

In this chapter we shall study the models of order-driven markets. As described in Chapter 1, an order-driven market for a particular security is described by its limit order book for that security. A limit order book for a security is an ordered listing of all outstanding limit orders. Each order states whether it is a buy or a sell order, the size of the order, the limit price at which it is authorized to trade. They are ordered according to price: from the highest ask price to lowest bid price. The ties are broken by time of arrivals: the earliest order has higher preference. Further ties are broken by size of the order: the bigger orders have a higher preference. The the lowest ask is called the market ask, and the highest bid is called the market bid. The gap between the market ask and market bid is called the bid-ask spread. The composition of the limit order book changes stochastically in response to the order arrivals, executions, and cancelations.

In this chapter we consider several dynamic programming models that describe the behavior of traders in an order-driven market. These models explicitly try to model the main feature of the order-driven markets: a limit order may fail to execute. So the trader has to balance the uncertain execution of a more profitable trade, with a certain execution of a less profitable trade. We begin with a model of a single trader who is attempting to trade in an order driven market.

### 9.1 Single Trader, Single Period

The material of this section is based on Section 12.1 of Hasbrouck [22]. We consider a single trader who wants to trade a single unit of a security in an order driven market. The buyer initially has  $n$  units of a security. At time 1 the security has value  $V$  which is a  $N(\mu, \sigma^2)$  random variable. The utility of the buyer is  $U(w) = -e^{-\rho w}$ . At time 0 the market ask and bid prices are  $a$  and  $b$

respectively, with  $a > b$ . The trader has five options:

1. Place a market buy order at price  $a$ ,
2. Place a limit buy order at a price  $a - \delta$ , for a suitable  $\delta > 0$ ,
3. Do nothing,
4. Place a limit sell order at a price  $b + \delta$  for a suitable  $\delta > 0$ ,
5. Place a market sell order at price  $b$ .

The aim of the trader is to make a decision that maximizes the expected utility of his wealth at time 1.

The initial wealth of the trader is  $W = nV$ . Hence the expected utility of the initial wealth is

$$U_0 = -\mathbb{E}(e^{-\rho nV}) = -\exp(-\rho n\mu + \frac{1}{2}\rho^2 n^2 \sigma^2).$$

Under option 1, the trader chooses to place a market buy order, his wealth at time 1 is  $(n+1)V - a$  and its expected utility is

$$U_1^1 = -\exp(-\rho((n+1)\mu - a) + \frac{1}{2}\rho^2(n+1)^2\sigma^2) = \exp(-\rho(\mu - a) + \frac{1}{2}\rho^2(2n+1)\sigma^2)U_0.$$

Next we analyze option 2. Here we need a model for the execution probability. This is in general a complicated quantity. Here we simply assume that  $f(\delta)$  is the probability that a limit order at price  $a - \delta$  will be executed by time 1. It makes sense to assume that  $f$  is a decreasing function of  $\delta$  for  $\delta > 0$ . For the sake of concreteness, assume that  $f(\delta) = ce^{-\delta\lambda}$  where  $0 < c < 1$  and  $\lambda > 0$  are given constants. Note that  $f(0+) = c < 1$ . This is because there is a strictly positive probability that a limit order may not execute, no matter how close it is to the market ask. Then, the expected utility of the terminal wealth under the second option is given by

$$\begin{aligned} U_1^2(\delta) &= -c \exp(-\lambda\delta - \rho((n+1)\mu - a + \delta) + \frac{1}{2}\rho^2(n+1)^2\sigma^2) + (1 - ce^{-\lambda\delta})U_0 \\ &= ce^{-(\lambda+\rho)\delta}U_1^1 + (1 - ce^{-\lambda\delta})U_0. \end{aligned}$$

Let

$$U_1^2 = \max_{\delta > 0} U_1^2(\delta)$$

be the expected utility under the best option 2. Let  $\delta^*(a)$  be the value of  $\delta$  where the above maximum is achieved. Clearly, the expected utility  $U_1^3$  of the third option is simply  $U_0$ . To analyze option four, we assume, for simplicity, that the probability of execution of a limit buy order at price  $b + \delta$  is also given by  $f(\delta)$  of option two. Then, the expected utility of the terminal wealth under the fourth option is given by

$$U_1^4(\delta) = ce^{-(\lambda+\rho)\delta}U_1^5 + (1 - ce^{-\lambda\delta})U_0,$$

where  $U_1^5$ , the expected terminal utility under option 5, is given by

$$U_1^5 = -\exp(-\rho((n-1)\mu + b) + \frac{1}{2}\rho^2(n-1)^2\sigma^2).$$

Hence the maximum expected utility at time 1 is given by

$$U_1 = \max\{U_1^1, U_1^2, U_1^3, U_1^4, U_1^5\}.$$

One can show that there is a critical ask price  $a^*$  below which it is optimal for the trader to put in a market order. If the ask price is  $a > a^*$ , it is optimal to put in a limit order at bid price  $a - \delta^*(a)$ . It is interesting that  $\delta^*(a^*+)$  is typically strictly positive. Thus as the ask price decreases to  $a^*$ , the optimal bid price of the trader increases to a value strictly less than  $a^*$ , and then the trader suddenly switches to a market order. This effect is called the “gravitational pull” of the market order. Similar effects are present on the sell side.

**Question:** The above model does not include any trading costs. Redo the above analysis by including (1) the cost of a market order, (2) the cost of submitting a limit order, and (3) the cost of executing a limit order. ■

## 9.2 Single Trader, Multiple Periods

We now extend the model of the previous section to  $T > 1$  periods. This material is largely based on Cohen et. al. [8]. A single trader can trade at times  $k = 0, 1, \dots, T-1$ . The trading stops in period  $T$ , and  $V$  is the terminal value of the security at time  $T$ . Let  $a_k$  and  $b_k$  be the market ask and market bid prices at time  $k$ ,  $n_k$  be the number of shares owned by the trader at time  $k$ . Let  $f_k(\delta_k)$  be the probability that a limit order  $\delta_k$  away from the market price will execute in period  $k$ . If a limit order is not executed, it is canceled. The trader has the same five options in each period  $k = 0, 1, \dots, T-1$  as described in Section 9.1.

Let  $\alpha_k$  be the option chosen in period  $k$ , and  $t_k = 1$  if a limit trade from the trader is executed in period  $k$ , and  $t_k = 0$  otherwise. We have  $\alpha_k \in A$ , where

$$A = \{1, (2, \delta), 3, (4, \delta), 5, \delta > 0\}.$$

Suppose, for  $0 \leq k \leq T-1$ , the transition probabilities are given by

$$H_k(x, y|a, b, i) = \mathbf{P}(a_{k+1} \leq x, b_{k+1} \leq y|a_k = a, b_k = b, \alpha_k = i), \quad i = 1, 3, 5,$$

$$H_k(x, y, m|a, b, i, \delta) = \mathbf{P}(a_{k+1} \leq x, b_{k+1} \leq y, t_k = m|a_k = a, b_k = b, \alpha_k = (i, \delta)), \quad i = 2, 4, m = 0, 1, \delta > 0.$$

Suppose that given  $\{(a_i, b_i, \alpha_i), 0 \leq i \leq k\}$ ,  $(a_{k+1}, b_{k+1}, t_k)$  depends only on  $(a_k, b_k, \alpha_k)$ . Let  $v_k(n, c, a, b)$  be the maximum expected utility of the wealth at time  $T$  if at time  $k$  the trader has  $n$

units of the security, and  $c$  amount of cash and  $a_k = a$ ,  $b_k = b$ . Then the trader's decision problem can be solved by using the following backward dynamic programming equations:

$$v_T(n, c, a, b) = \mathbf{E}(U(nV + c)),$$

and for  $k = 0, 1, \dots, T - 1$ ,

$$v_k(n, c, a, b) = \max\{v_k^i(n, c, a, b) : 1 \leq i \leq 5\},$$

where

$$\begin{aligned} v_k^1(n, c, a, b) &= \int_{x,y} v_{k+1}(n+1, c-a, x, y) dH_k(x, y|a, b, 1), \\ v_k^2(n, c, a, b) &= \max_{\delta > 0} \left[ \int_{x,y} v_{k+1}(n+1, c-a+\delta, x, y) dH_k(x, y, 1|a, b, 2, \delta) \right. \\ &\quad \left. + \int_{x,y} v_{k+1}(n, c, x, y) dH_k(x, y, 0|a, b, 2, \delta) \right], \\ v_k^3(n, c, a, b) &= \int_{x,y} v_{k+1}(n, c, x, y) dH_k(x, y|a, b, 3), \\ v_k^4(n, c, a, b) &= \max_{\delta > 0} \left[ \int_{x,y} v_{k+1}(n-1, c+b+\delta, x, y) dH_k(x, y, 1|a, b, 4, \delta) \right. \\ &\quad \left. + \int_{x,y} v_{k+1}(n, c, x, y) dH_k(x, y, 0|a, b, 4, \delta) \right], \\ v_k^5(n, c, a, b) &= \int_{x,y} v_{k+1}(n-1, c+b, x, y) dH_k(x, y|a, b, 5). \end{aligned}$$

These equations can be solved starting from  $k = T$ , and working backward to  $k = 0$ . The optimum utility of the terminal wealth is given by  $v_0$ . We saw in Section 9.1 how to compute  $v_{T-1}$  from  $v_T$  for a specific version of these equations. See Cohen et. al. [8] for more discussion of this model, although they allow the trader to either buy or sell, but not do both.

**Question:** Include the costs of transactions in the above analysis. Suppose  $a_k$  and  $b_k$  do not change over  $0 \leq k \leq T$ , and the probability of successful trade is given by  $f(\delta)$  as in Section 9.1. Study the structural properties of the solution. ■

### 9.3 Parlour Model

One main deficiency of the above model is the specification of the probability of execution. It is taken as an exogenous event, while in practice it is a consequence of the behavior of the other traders on the book. In this section we describe a model inspired by Parlour [38] that attempts to address this deficiency.

As in Section 9.2 we assume the trades can occur at times  $k = 0, 1, \dots, T - 1$ , and the book closes at time  $T$ . The book maker maintains the market ask at  $A$ , and the market bid at  $B < A$ , at all times. That is, the dealer accepts limit sell orders at  $A$  and limit buy orders at  $B$ . When there are no limit orders on the book, the dealer will honor the market orders from his inventory. Let  $(X_k, Y_k)$  be the state of the book as seen by the trader arriving at time  $k$ , where  $X_k$  is the number of limit buy orders at  $B$  and  $Y_k$  is the number of limit sell orders at  $A$ . Let  $V_k$  be the cash-value placed on the trade by the  $k$ th trader. Assume that  $V_k$  is known to the  $k$ th trader, and  $\{V_k, 0 \leq k \leq T - 1\}$  are independent non-negative random variables whose distribution is known to all. As in Section 9.1, a trader arriving at time  $k$  has the following five options:

1. Place (and immediately execute) a market buy order at price  $A$ . The trader's equivalent cash position becomes  $V_k - A$ . The book state changes to

$$(X_{k+1}, Y_{k+1}) = (X_k, (Y_k - 1)^+).$$

2. Place (and wait for uncertain execution) a limit buy order at a price  $B$ . The trader's equivalent cash position becomes  $V_k - B$  if the trade executes, otherwise it does not change. The book state changes to

$$(X_{k+1}, Y_{k+1}) = (X_k + 1, Y_k).$$

3. Do nothing. The cash position and the book position does not change.

4. Place (and wait for uncertain execution) a limit sell order at a price  $A$ . The trader's equivalent cash position becomes  $A - V_k$  if the trade executes, otherwise it does not change. The book state changes to

$$(X_{k+1}, Y_{k+1}) = (X_k, Y_k + 1).$$

5. Place (and immediately execute) a market sell order at price  $B$ . The trader's equivalent cash position becomes  $B - V_k$ . The book state changes to

$$(X_{k+1}, Y_{k+1}) = ((X_k - 1)^+, Y_k).$$

Once an action is chosen, the trader cannot cancel it. Let  $d_k(x, y, v)$  be the option chosen by the  $k$ th trader if  $X_k = x, Y_k = y, V_k = v, 0 \leq k \leq T - 1, x \geq 0, y \geq 0$ . Let  $p_k^s(x, y, z)$  ( $p_k^b(x, y, z)$ ) be the probability that at least  $z$  sales at limit price  $A$  (buys at limit price  $B$ ) will occur over  $\{k, k + 1, \dots, T - 1\}$  if  $X_k = x$  and  $Y_k = y$ . Let  $C_k^i(x, y, v)$  be the expected cash position if the  $d_k(x, y, v) = i$ . We have

$$C_k^1(x, y, v) = v - A, C_k^2(x, y, v) = p_{k+1}^b(x + 1, y, x + 1)(v - B), C_k^3(x, y, v) = 0,$$

$$C_k^4(x, y, v) = p_{k+1}^s(x, y + 1, y + 1)(A - v), C_k^5(x, y, v) = B - v.$$

Now consider the  $k$ th trader who sees the book in state  $(x, y)$ . For brevity, let

$$B' = \frac{B - p_{k+1}^s(x, y+1, y+1)A}{1 - p_{k+1}^s(x, y+1, y+1)}, \quad A' = \frac{A - p_{k+1}^b(x+1, y, x+1)B}{1 - p_{k+1}^b(x+1, y, x+1)},$$

$$C' = \frac{p_{k+1}^s(x, y+1, y+1)A + p_{k+1}^b(x+1, y, x+1)B}{p_{k+1}^s(x, y+1, y+1) + p_{k+1}^b(x+1, y, x+1)}.$$

Note that  $B' \leq B$ ,  $A' \geq A$ , and  $B \leq C' \leq A$ . Then the optimal option for the  $k$ th trader with valuation  $v$  is given by

$$\begin{aligned} v \leq B' &\Rightarrow d_k(x, y, v) = 5, \\ B' \leq v \leq C' &\Rightarrow d_k(x, y, v) = 4, \\ C' \leq v \leq A' &\Rightarrow d_k(x, y, v) = 2, \\ A' \leq v &\Rightarrow d_k(x, y, v) = 1. \end{aligned}$$

The ties can be broken arbitrarily. Note that a trader never has to choose “do nothing” under the optimal policy. If the trader values the stock sufficiently low, he should sell (either market or limit), and if he values it sufficiently high, he should buy (either limit or market). The control limits  $B', C', A'$  depend on  $k$  and the state of the order book through the probabilities  $p_{k+1}^s$  and  $p_{k+1}^b$ . Hence the solution will be complete if we can show how to compute them.

First define

$$\phi_k^i(x, y) = \mathbb{P}(d_k(x, y, V_k) = i | X_k = x, Y_k = y)$$

as the probability that the  $k$ th trader chooses option  $i$  in state  $(x, y)$ . By using the first step analysis in DTMCs we get the following recursive equations

$$p_k^s(x, y, 0) = p_k^b(x, y, 0) = 1, \quad 0 \leq k \leq T.$$

For  $z > 0$ ,

$$\begin{aligned} p_T^s(x, y, z) &= p_T^b(x, y, z) = 0, \quad z > 0. \\ p_k^s(x, y, z) &= \phi_k^1(x, y)p_{k+1}^s(x, (y-1)^+, z-1) + \phi_k^2(x, y)p_{k+1}^s(x+1, y, z) \\ &+ \phi_k^4(x, y)p_{k+1}^s(x, y+1, z) + \phi_k^5(x, y)p_{k+1}^s((x-1)^+, y, z), \quad 0 \leq k \leq T-1, \\ p_k^b(x, y, z) &= \phi_k^1(x, y)p_{k+1}^b(x, (y-1)^+, z) + \phi_k^2(x, y)p_{k+1}^b(x+1, y, z) \\ &+ \phi_k^4(x, y)p_{k+1}^b(x, y+1, z) + \phi_k^5(x, y)p_{k+1}^b((x-1)^+, y, z-1), \quad 0 \leq k \leq T-1. \end{aligned}$$

We can now recursively compute the optimal decisions in backward fashion. Parlour [38] studies the monotonicity properties of the optimal policies with respect to  $k$ .

**Question:** Analyze the optimal policies under the assumption that  $V_k \sim U(0, 1)$ , and  $0 < B < A < 1$ . Numerically experiment with different values of  $T$  and observe its effect on the expected

depth of the order book ( $E(X_k + Y_k)$ ) as a function of  $k$ . Study the monotonicity of  $p_k^b$ ,  $p_k^s$ ,  $A'$ ,  $B'$  and  $C'$  as a function of  $k$  for a fixed  $(x, y)$ .

**Question:** Consider the discounted model as follows: Suppose  $T = \infty$ , and a discount factor of  $\rho \in (0, 1)$ . For example, if a trader places a limit buy order at  $B$  at time  $k$ , and it gets executed at time  $k + K$ , the equivalent cash position of the trader is  $V_k - B\rho^K$ .

## 9.4 Foucault Model

The one limiting feature of the Parlour model of Section 9.3 is that the ask and bid prices are fixed, and the traders are not allowed to choose their own ask and bid prices. It does however allow the queues of the standing limit buy and sell orders as part of the state of the order book. Foucault [13], addresses the first issue, but at the cost of second issue. Thus his model allows the ask and bid prices to vary, but there can be one limit buy and one sell order in the book, or the book can be empty. We describe the model in more detail in this section.

As in the Parlour model, traders arrive at times  $k = 0, 1, 2, \dots$ . Unlike in the parlour model, the market closes at a random time  $K$  with pmf

$$P(K = k) = \rho^k(1 - \rho), \quad k \geq 0.$$

That is, if the market is open at time  $k$ , it will stay open at time  $k + 1$  with probability  $\rho$ . The fundamental value  $V_k$  of the security at time  $k$  follows a simple random walk models

$$V_{k+1} = V_k + \sigma\epsilon_{k+1}, \quad k \geq 0$$

where  $\{\epsilon_k, k \geq 1\}$  is a sequence of iid random variables with zero mean and unit variance.  $\sigma \geq 0$  is a fixed constant, representing the volatility of the security. The value of the security when the market closes is given by  $V_K$ . One can think of  $V_k$  as the expected value of  $V_K$  given the information  $\{\epsilon_i, 1 \leq i \leq k\}$ . The private valuation of the security by the  $k$ th trader is  $V_k + Y_k$ , where  $\{Y_k, k \geq 0\}$  are iid random variables with zero mean.

If the order book is not empty when the  $k$ th trader arrives, it has exactly one limit buy order at  $B_k$  and one limit sell order at  $A_k > B_k$ . If the order book is empty,  $A_k = \infty$  and  $B_k = -\infty$ . The  $k$ th trader observes the state of the order book  $(A_k, B_k)$ , and private evaluation  $V_k + Y_k$ , and chooses one of the following three possible options:

1. If the order book is not empty, buy at price  $A_k$ .
2. If the order book is not empty, sell at price  $B_k$ .

3. Ignore the current state of the book, and place a new limit sell order at  $A_{k+1}$ , and a new limit buy order at  $B_{k+1}$ .

Clearly, if the order book is empty, the only feasible option is the third one. The limit order placed at time  $k$  is either executed at time  $k + 1$ , or is canceled. The expected utility from option 1 is  $V_k + Y_k - A_k$ , from option 2 is  $B_k - V_k - Y_k$ . The utility from option 3 is more complicated. If the market closes at time  $k + 1$ , the limit orders stay unexecuted, and the utility is zero. If the market stays open and the trader at time  $k + 1$  decides to use option 3, the utility is zero. If the market stays open and the trader at time  $k + 1$  decides to use option 1, the expected utility is  $A_k - V_k - Y_k$ , and if the trader uses option 2, it is  $V_k + Y_k - B_k$ .

The geometric nature of the market closing time, and one-period life time of the orders, implies that the optimal decision of a trader who sees  $A_k = a$ ,  $B_k = b$ ,  $V_k = v$ , and  $Y_k = y$ , depends only on  $(a, b, v, y)$  and not on  $k$ . Let  $\phi(a, b, v, y)$  be the maximum utility of the trader who sees the state  $(a, b, v, y)$ . Then the dynamic programming equation is given by:

$$\phi(a, b, v, y) = \max\{v + y - a, b - v - y, \rho\psi(a, b, v, y)\}$$

where

$$\psi(a, b, v, y) = \max_{\{\alpha, \beta\}}\{(\alpha - v - y)\mathbb{P}(d(\alpha, \beta, v + \sigma\epsilon, Y) = 1) + (v + y - \beta)\mathbb{P}(d(\alpha, \beta, v + \sigma\epsilon, Y) = 2)\},$$

where

$$\begin{aligned} d(a, b, v, y) &= 1 \text{ if } \phi(a, b, v, y) = v + y - a, \\ d(a, b, v, y) &= 2 \text{ if } \phi(a, b, v, y) = b - v - y, \\ d(a, b, v, y) &= 3 \text{ if } \phi(a, b, v, y) = \rho\psi(a, b, v, y), \end{aligned}$$

where  $\epsilon$  has the same distribution as  $\epsilon_k$ , and  $Y$  has the same distribution as  $Y_k$ .

Foucault [13] provides a complete solution to the above equation under the assumption that the pmfs of  $\epsilon_k$  and  $Y_k$  are given

$$\begin{aligned} \mathbb{P}(\epsilon_k = 1) &= \mathbb{P}(\epsilon_k = -1) = .5, \\ \mathbb{P}(Y_k = L) &= \mathbb{P}(Y_k = -L) = .5. \end{aligned}$$

We reproduce their solution for the special case  $\sigma = 0$  below. In this case the value of the security does not change, and is given by

$$V_k = V_0, \quad k \geq 0.$$

If the trader decides to place limit orders, the optimal ask and bid values are given by

$$\begin{aligned} \alpha &= \alpha(a, b, v, y) = v + L \frac{2 - \rho}{2 + \rho}, \\ \beta &= \beta(a, b, v, y) = v - L \frac{2 - \rho}{2 + \rho}. \end{aligned}$$



Since these are independent of  $(a, b, y)$ , the state of the book is either  $(\alpha, \beta)$  or  $(\infty, -\infty)$ . Thus we need to only consider the cases  $(\infty, -\infty, V_0, L)$ ,  $(\infty, -\infty, V_0, -L)$ ,  $(\alpha, \beta, V_0, L)$  and  $(\alpha, \beta, V_0, -L)$ . We have

$$\begin{aligned}
d(\infty, -\infty, V_0, L) &= d(\infty, -\infty, V_0, -L) = 3, \\
\phi(\infty, -\infty, V_0, L) &= \phi(\infty, -\infty, V_0, -L) = \rho L \frac{2 - \rho}{2 + \rho}, \\
\psi(\alpha, \beta, V_0, L) &= \psi(\alpha, \beta, V_0, -L) = L \frac{2 - \rho}{2 + \rho}, \\
\phi(\alpha, \beta, V_0, L) &= \phi(\alpha, \beta, V_0, -L) = \rho L \frac{2}{2 + \rho}, \\
d(\alpha, \beta, V_0, L) &= 1, \quad d(\alpha, \beta, V_0, -L) = 2.
\end{aligned}$$

The last equation implies that

$$P(d(\alpha, \beta, V_0, Y) = 1) = P(d(\alpha, \beta, V_0, Y) = 2) = .5.$$

We see that when all customers behave optimally, a customer with  $Y = L$  places a market buy order, and with  $Y = -L$  places a market sell order, if the order book is not empty. If the order book is empty, the customer places a limit buy order at  $\beta$  and a limit sell order at  $\alpha$ . Thus the order book alternates between full and empty.

**Question:** Put the results from Foucault [13] for the general case for  $\sigma > 0$ , in the above framework. You will need to consider two case  $L > \sigma$  and  $L < \sigma$  to do a careful analysis. ■

The Foucault model does not allow queues of the limit orders to develop. In the next chapter we shall study models that explicitly study the queues of limit orders on the order book.

## Chapter 10

# Order-Driven Markets: Simple Queueing Models

In this chapter we study several “queueing” type models of the order book.

### 10.1 The Garman Model

In this section we study a simple model of the order book described in Garman [14]. The limit prices are assumed to be a discrete set of increasing prices  $\{0 < p_1 < p_2 < \cdots < p_n < \infty\}$ , for a fixed  $n \geq 1$ . All order sizes are equal to 1. Let  $X_j(t)$  be the number of orders on the order book with limit price  $p_j$  ( $-\infty < X_j(t) < \infty$ ). If there are  $n_j$  buy orders at time  $t$  with limit price  $p_j$ , we set  $X_j(t) = n_j$ , and if there are  $n_j$  sell orders at time  $t$  with limit price  $p_j$ , we set  $X_j(t) = -n_j$ . It will be useful to define  $p_0 = 0$  and  $p_{n+1} = \infty$ , and  $X_0(t) = \infty$  and  $X_{n+1}(t) = -\infty$ . Thus there are an infinite number of buyers willing to buy at zero price, and an infinite number of sellers willing to sell at infinite price. The state of the book at time  $t$  is given by

$$X(t) = [X_1(t), X_2(t), \dots, X_n(t)].$$

A buy order at the highest bid price is available to satisfy any sell order at that price or lower, and a sell order at the lowest ask price is available to any buy order at that price or higher. This implies that if  $X_j(t) > 0$  for any  $1 \leq j \leq n$ , then we must have  $X_i(t) \geq 0$  for all  $1 \leq i \leq j$ , and if  $X_j(t) < 0$  for any  $1 \leq j \leq n$ , then we must have  $X_i(t) \leq 0$  for all  $j \leq i \leq n$ . Now define

$$a(t) = \min\{j \in \{0, 1, \dots, n+1\} : X_j(t) < 0\}, \quad (10.1)$$

and

$$b(t) = \max\{j \in \{0, 1, \dots, n+1\} : X_j(t) > 0\}. \quad (10.2)$$

Thus we can define the market ask price at time  $t$  as

$$A(t) = p_{a(t)}, \quad (10.3)$$

and the market bid price at time  $t$  as

$$B(t) = p_{b(t)}. \quad (10.4)$$

Clearly  $A(t) > B(t)$  for all  $t \geq 0$ , and the spread at time  $t$  is given by

$$S(t) = A(t) - B(t).$$

Next we describe the process of order arrival. Suppose orders with bid price  $p_j$  (buy orders) arrive according to a Poisson process with rate  $\lambda_b(j)$  and orders with ask price  $p_j$  (sell orders) arrive according to a Poisson process with rate  $\lambda_a(j)$ ,  $1 \leq j \leq n$ . If a buy order arriving at time  $t$  is at bid price less than  $A(t)$ , it will be added to the queue of buy orders at that bid price. On the other hand if its bid price is at or higher than  $A(t)$ , it will be immediately crossed with a sell order at the ask price  $A(t)$ , and will result in increasing  $X_{a(t)}(t)$  by 1 (remember that  $X_{a(t)}(t)$  is negative). The transaction price will be  $a(t)$ . If  $X_{a(t)}$  is zero after the transaction, the  $a(t)$  and  $A(t)$  will increase; else they stay unchanged. If a sell order arriving at time  $t$  is at an ask price more than  $B(t)$ , it will be added to the queue of sell orders at that ask price. On the other hand if its ask price is at or lower than  $B(t)$ , it will be immediately crossed with a buy order at the bid price  $B(t)$ , and will result in reducing  $X_{b(t)}(t)$  by 1. The transaction price will be  $b(t)$ . If  $X_{b(t)}$  is zero after the transaction, the  $b(t)$  and  $B(t)$  will decrease; else they stay unchanged. Finally, we assume that the traders are impatient, and will cancel their orders if they do not get executed for a random amount of time, called the “patience time”. We shall assume that all patience times of the active orders are iid  $\exp(\nu)$  random variables.

It is clear that  $\{X(t), t \geq 0\}$  is a Continuous Time Markov Chain (CTMC). To describe the state space we need the following notation. Let  $x = [x_1, x_2, \dots, x_n]$ , with  $-\infty < x_i < \infty$ ,  $1 \leq i \leq n$ . Define

$$\alpha(x) = \min\{j \in \{0, 1, \dots, n+1\} : x_j < 0\},$$

$$\beta(x) = \max\{j \in \{0, 1, \dots, n+1\} : x_j > 0\}.$$

The state space of the CTMC is given by:

$$S = \{x = [x_1, x_2, \dots, x_n] : -\infty < x_i < \infty, 1 \leq i \leq n, \alpha(x) > \beta(x)\}.$$

Let  $x_0 = \infty$  and  $x_{n+1} = -\infty$ . The non-zero transition rates are

$$q(x, x - e_i) = x_i \nu + 1_{\{\beta(x)=i\}} \sum_{j \leq i} \lambda_a(j), \quad x_i > 0,$$

$$q(x, x + e_i) = -x_i \nu + 1_{\{\alpha(x)=i\}} \sum_{j \geq i} \lambda_b(j), \quad x_i < 0,$$

$$q(x, x + e_i) = 1_{\{\alpha(x) > i\}} \lambda_b(i), \quad x_i \geq 0,$$

$$q(x, x - e_i) = 1_{\{\beta(x) < i\}} \lambda_a(i), \quad x_i \leq 0.$$

We also have

$$q(x) = -q(x, x) = \nu \sum_{j=1}^n |x_j| + \sum_{j=1}^n \lambda_b(j) + \sum_{j=1}^n \lambda_a(j), \quad x \in S.$$

Thus a transition in the CTMC results in a change in exactly one component by  $\pm 1$ . The CTMC is irreducible and positive recurrent as long as  $\nu > 0$ . Let

$$p(x, t) = \mathbf{P}(X(t) = x), \quad x \in S,$$

$$p(x) = \lim_{t \rightarrow \infty} p(x, t), \quad x \in S.$$

The Chapman-Kolmogorov forward equations are given by

$$p'(x, t) = q(x, x)p(x, t) + \sum_{i=1}^n p(x + e_i, t)q(x + e_i, x) + \sum_{i=1}^n p(x - e_i, t)q(x - e_i, x).$$

In general these are tough to solve. The limiting distributions satisfy the balance equations given by

$$q(x)p(x) = \sum_{i=1}^n p(x + e_i)q(x + e_i, x) + \sum_{i=1}^n p(x - e_i)q(x - e_i, x).$$

In general these equations are also tough to solve. We illustrate a special case below.

## 10.2 Special Case: A Clearing Market

Consider the special case of the Garman model with  $n = 1$ . Thus all orders arrive for a single price  $p = p_1$ , and the market simply acts as a clearing house where unmet orders are queued up. The buy orders arrive at rate  $\lambda = \lambda_b(1)$  and sell orders arrive at rate  $\mu = \lambda_a(1)$ . The state of the market is given by  $X(t) = X_1(t)$ . This a birth and death process on all integers with birth and death rates given by

$$\lambda_j = \begin{cases} \lambda & \text{if } j \geq 0 \\ \lambda - \nu j & \text{if } j < 0, \end{cases}$$

$$\mu_j = \begin{cases} \mu & \text{if } j \leq 0 \\ \mu + \nu j & \text{if } j > 0, \end{cases}$$

Now let

$$p_j = \lim_{t \rightarrow \infty} \mathbf{P}(X(t) = j), \quad -\infty < j < \infty.$$

One can easily compute the  $p_j$ 's as follows.

$$p_j = \begin{cases} p_0 \prod_{i=1}^j \frac{\lambda}{\mu + i\nu} & \text{if } j > 0 \\ p_0 \prod_{i=1}^{-j} \frac{\mu}{\lambda + i\nu} & \text{if } j < 0, \end{cases}$$

where  $p_0$  is determined by

$$\sum_{j=-\infty}^{\infty} p_j = 1.$$

Now let  $\Gamma(x, y)$  be the incomplete Gamma function

$$\Gamma(x, y) = \int_0^y e^{-t} x^t dt,$$

and define

$$c^+ = (\lambda/\nu)^{-\mu/\nu} e^{\lambda/\nu} \Gamma(\mu/\nu, \lambda/\nu),$$

$$c^- = (\mu/\nu)^{-\lambda/\nu} e^{\mu/\nu} \Gamma(\lambda/\nu, \mu/\nu).$$

Garman [14] shows that

$$p_0 = (1 + c^+ + c^-)^{-1}.$$

In steady state there are buy orders outstanding with probability

$$p^+ = \lim_{t \rightarrow \infty} \mathbf{P}(X(t) > 0) = c^+ p_0$$

and sell orders outstanding with probability

$$p^- = \lim_{t \rightarrow \infty} \mathbf{P}(X(t) < 0) = c^- p_0.$$

The expected state of the clearing market in steady state is shown to be

$$L = \lim_{t \rightarrow \infty} \mathbf{E}(X(t)) = \frac{\lambda - \mu}{\nu},$$

while the expected number of outstanding orders in steady state is given by

$$M = \lim_{t \rightarrow \infty} \mathbf{E}(|X(t)|) = \frac{1}{\nu} [(\lambda + \mu)p_0 + (\lambda - \mu)p^+ + (\mu - \lambda)p^-].$$

We can derive some further results not given in Garman. The rate at which orders leave due to impatience is given by  $M\nu$ . Since the orders arrive at rate  $\lambda + \mu$ , the probability that an order is successfully executed is given by

$$1 - \frac{M\nu}{\lambda + \mu} = 1 - p_0 - \frac{(\lambda - \mu)p^+ + (\mu - \lambda)p^-}{\lambda + \mu}.$$

The expected buy orders in steady state is given by

$$M_b = \lim_{t \rightarrow \infty} \mathbf{E}(X(t); X(t) > 0) = \frac{1}{\nu} [\lambda p_0 + (\lambda - \mu)p^+],$$

while the expected number of sell orders is

$$M_a = \lim_{t \rightarrow \infty} \mathbf{E}(-X(t); X(t) < 0) = \frac{1}{\nu} [\mu p_0 + (\mu - \lambda)p^-].$$

The rate at which the buy orders leave due to impatience is  $\nu M_b$ , and the rate at which sell orders leave is  $\nu M_a$ . Hence the steady probability that a buy order is executed is

$$1 - \nu M_b / \lambda = 1 - p_0 - (1 - \mu/\lambda)p^+,$$

and the steady probability that a sell order is executed is

$$1 - \nu M_a / \mu = 1 - p_0 - (1 - \lambda / \mu) p^-.$$

The concept of market ask and market bid, and spread do not make any sense in this special case. Let

$$x(t) = \mathbf{E}(X(t)), \quad t \geq 0.$$

**Question:** Show that

$$x'(t) = \lambda - \mu - \nu x(t).$$

This has a solution

$$x(t) = \frac{\lambda - \mu}{\nu} (1 - e^{-\nu t}) + x(0) e^{-\nu t}.$$

Is there a tractable expression for  $\mathbf{E}(|X(t)|)$ ? ■

**Question:** Compute the steady state distribution for the general model with  $n > 1$ . ■

### 10.3 Extensions to the Garman Model

One restrictive assumption of the Garman model is that all orders are limit orders, and there are no market orders. The second restrictive assumption of the Garman model is the assumption that the arrival rates  $\lambda_b(j)$  and  $\lambda_a(j)$  depend only on  $j$  (or on  $p_j$ ). For example it seems reasonable to expect that  $\lambda_b(j)$  is a decreasing function of  $j$ , while  $\lambda_a(j)$  is an increasing function of  $j$ . However, in practice, it is observed that the arrival rate of buy (sell) orders decreases as you move away from the market bid (ask) price. Thus what we need is to make the arrival rates depend on the state of the order book, at the very least on  $(j, a(t))$  or  $(j, b(t))$ . This makes the model even more intractable. See Cont, et al [9].

In this section we present a model studied by Domowitz and Wang [12] that removes these restrictions. Let  $X(t)$ ,  $a(t)$  and  $b(t)$  be as defined in Section 10.1. Let  $\lambda_a(i, j)$  be the Poisson arrival rate of a sell order at price  $p_i$  if the market bid is  $p_j$ . Clearly  $\lambda_a(i, j) = 0$  if  $i < j$ , since no seller will offer to sell at prices less than the market bid. We can think of  $\lambda_a(j, j)$  as the arrival rate for market sell orders when the bid price is  $p_j$ . Similarly, let  $\lambda_b(i, j)$  be the Poisson arrival rate of buy orders with limit price  $p_i$  if the market ask price is  $p_j$ . We have  $\lambda_b(i, j) = 0$  if  $i > j$ , and  $\lambda_b(j, j)$  is the arrival rate of market buy orders if the market ask price is  $p_j$ . The model also allows additional market order that depend on both market ask and market bid. Let  $\gamma_a(i, j)$  ( $\gamma_b(i, j)$ ) be the arrival rate of the market sell (buy) orders if the market bid is  $p_i$  and ask is  $p_j$ . The model allows the outstanding orders to be canceled. A sell order with ask price  $p_i$  can cancel at rate  $\nu_a(i, j)$  if the market bid is  $p_j$ , and a buy order with limit price  $p_i$  can cancel at rate  $\nu_b(i, j)$  if the market ask

price is  $p_j$ .

Unfortunately, the results presented in Domowitz and Wang [12] are approximate, and hence we shall go into more details of them here.

The third restrictive assumption is that all orders are of unit size. Can this be extended to orders of random sizes?

Finally, the Garman model assumes that the traders are non-strategic. In practice they are trying to optimize their returns taking advantage of the information provided by the order book. Exploring this aspect will take us into game theoretic models.

## 10.4 Luckock Model

Luckock [31] develops a stochastic model of an order book that is closely related to the Garman model. He studies the distribution of the bid and ask prices in steady state. We shall present some of his results below.

The potential buyers (sellers) arrive at the market according to a Poisson process and place iid random limit orders. Let  $\lambda_b(x)$  be the rate at which limit buy order arrive with prices  $x$  or more, and  $\lambda_a(x)$  be the rate at which limit sell orders arrive with prices  $x$  or less. Thus  $\lambda_a(\cdot)$  is an increasing function, and  $\lambda_b(\cdot)$  is a decreasing function. We assume that both functions are continuous over  $(0, \infty)$ .

Let  $a(t)$  be the market ask price and  $b(t)$  be the market bid price at time  $t$ . Clearly, if there are no sell orders on the book at time  $t$ ,  $a(t) = \infty$ , and if there are no buy orders on the book at time  $t$ , the  $b(t) = 0$ . Note that  $a$  or  $b$  are not Markov processes, since their evolution depends on the actual ask and bid prices of each of the individual orders on the order book. Assume that they are stationary. In particular, define

$$A(x) = \mathbf{P}(a(t) \leq x),$$

$$B(x) = \mathbf{P}(b(t) \geq x),$$

as the stationary distributions of the market ask and bid. Let

$$x_{min} = \inf\{x \in [0, \infty] : B(x) < 1\}, \quad x_{max} = \sup\{x \in [0, \infty] : A(x) < 1\}.$$

Suppose the order book is in steady state at time 0, and  $a$  is the market ask price and  $b$  the market bid price. Then

$$x_{min} < a \leq x_{max}, \quad x_{min} \leq b < x_{max},$$

with probability 1. The next theorem yields an important property of  $A$  and  $B$ .

**Theorem 10.1** *There exists a constant  $\kappa$  such that for  $x \in (x_{min}, x_{max})$ ,*

$$(1 - A(x))\lambda_b(x) + (1 - B(x))\lambda_a(x) = \kappa. \quad (10.5)$$

**Proof:** Let  $x \in (x_{min}, x_{max})$ . Then the rate at which a buy order at bid price  $x$  gets added to the order book is  $(1 - A(x))d\lambda_b(x)$ , and the rate at which a buy order a bid price  $x$  is deleted from the book is  $\lambda_a(x)dB(x)$ . Since the order book is in steady state, these two rates must be equal. Similar analysis can be done of the sell orders at ask price  $x$ . Hence we get

$$(1 - A(x))d\lambda_b(x) - \lambda_a(x)dB(x) = 0,$$

$$(1 - B(x))d\lambda_a(x) - \lambda_b(x)dA(x) = 0.$$

Adding these two equations and integrating by parts yields Equation 10.5. ■

Note that  $(1 - A(x))\lambda_b(x)$  can be interpreted as the steady state frequency of transactions at prices  $x$  or more, and  $(1 - B(x))\lambda_a(x)$  can be interpreted as the steady state frequency of transactions at prices  $x$  or less. The above theorem says that their sum is constant on  $(x_{min}, x_{max})$ . One can show that

$$\kappa \in (\lambda_e, 2\lambda_e],$$

where

$$\lambda_e = \inf_x \max(\lambda_a(x), \lambda_b(x)).$$

The next theorem gives the main result.

**Theorem 10.2** *Assume that  $\lambda_a$  and  $\lambda_b$  are strictly positive continuous functions. The quantities  $\kappa, x_{min}, x_{max}, A$ , and  $B$  are given by the unique solution to*

$$x_{min} = \inf\{x : \lambda_b(x) < \kappa\}, \quad (10.6)$$

$$x_{max} = \sup\{x : \lambda_a(x) < \kappa\}, \quad (10.7)$$

$$\frac{1}{\kappa} \left[ \frac{1}{\lambda_a(x_{min})} + \frac{1}{\lambda_b(x_{max})} \right] = \frac{1}{\lambda_a(x_{max})} \frac{1}{\lambda_b(x_{max})} - \int_{x_{min}}^{x_{max}} \frac{1}{\lambda_b(x)} d\left(\frac{1}{\lambda_a(x)}\right), \quad (10.8)$$

$$A(x) = 1 - \frac{\lambda_a(x)}{\lambda_a(x_{min})} - \kappa\lambda_a(x) \int_{x_{min}}^x \frac{1}{\lambda_b(x)} d\left(\frac{1}{\lambda_a(x)}\right), \quad (10.9)$$

$$B(x) = 1 - \frac{\lambda_b(x)}{\lambda_b(x_{max})} + \kappa\lambda_b(x) \int_x^{x_{max}} \frac{1}{\lambda_a(x)} d\left(\frac{1}{\lambda_b(x)}\right). \quad (10.10)$$

**Proof:** See Luckock [31]. ■

Luckock [31] further analyzes two interesting special cases. We provide the main results below.



### 10.4.1 Separable Markets.

Let

$$\begin{aligned}\mu_b &= \lambda_b(\infty), \quad \mu_a = \lambda_a(0), \\ x_b &= \inf\{x \geq 0 : \lambda_b(x) = \mu_b\}, \\ x_a &= \sup\{x \geq 0 : \lambda_a(x) = \mu_a\}.\end{aligned}$$

Thus  $\lambda_b(x) = \mu_b$  for  $x > x_b$ , and  $\lambda_a(x) = \mu_a$  for  $x \leq x_a$ . The market is called *separable* if  $x_b < x_a$ . Thus the limit orders are always settled against the market orders, and never against other limit orders. Thus the order book can be treated as two independent one-sided order books. (See Section 11.1.) This makes the analysis tractable. The final results are given below:

$$\begin{aligned}x_{max} &= \sup\{x \geq 0 : \lambda_a(x) < \mu_a + \mu_b\}, \\ x_{min} &= \inf\{x \geq 0 : \lambda_b(x) < \mu_a + \mu_b\}, \\ \kappa &= \mu_a + \mu_b, \\ A(x) &= \begin{cases} 0 & \text{if } x \in (0, x_a) \\ \frac{\lambda_a(x) - \mu_a}{\mu_b} & \text{if } x \in [x_a, x_{max}] \\ 1 & \text{if } x \in [x_{max}, \infty) \end{cases} \\ B(x) &= \begin{cases} 1 & \text{if } x \in (0, x_{min}] \\ \frac{\lambda_b(x) - \mu_b}{\mu_a} & \text{if } x \in (x_{min}, x_b] \\ 0 & \text{if } x \in (x_b, \infty). \end{cases}\end{aligned}$$

### 10.4.2 Constant Elasticities

Consider the functions

$$\lambda_a(x) = \lambda_e x^{\eta_a}, \quad \lambda_b(x) = \lambda_e x^{-\eta_b},$$

where  $\lambda_e > 0$ ,  $\eta_a > 0$  and  $\eta_b > 0$  are given constants. Thus the buy and sell orders have constant elasticities. We get the following results if  $\eta_a \neq \eta_b$ :

$$\begin{aligned}\kappa &= \lambda_e \left( \frac{\eta_a}{\eta_b} \right)^{\frac{\eta_a \eta_b}{\eta_a^2 - \eta_b^2}}, \\ x_{min} &= \left( \frac{\eta_a}{\eta_b} \right)^{-\frac{\eta_a}{\eta_a^2 - \eta_b^2}}, \\ x_{max} &= \left( \frac{\eta_a}{\eta_b} \right)^{\frac{\eta_b}{\eta_a^2 - \eta_b^2}}, \\ A(x) &= 1 - \frac{1}{\eta_a - \eta_b} [\eta_a (x/x_{min})^{\eta_b} - \eta_b (x/x_{min})^{\eta_a}], \quad x_{min} \leq x \leq x_{max}, \\ B(x) &= 1 - \frac{1}{\eta_a - \eta_b} [\eta_a (x/x_{max})^{-\eta_b} - \eta_b (x/x_{max})^{-\eta_a}], \quad x_{min} \leq x \leq x_{max}.\end{aligned}$$

In case  $\eta_a = \eta_b = \eta$ , the above results reduce to

$$\kappa = \lambda e^{1/2}, \quad x_{min} = e^{-(1/2)\eta}, \quad x_{max} = e^{(1/2)\eta},$$

$$A(x) = 1 - e^{1/2} x^\eta (1/2 - \eta \ln x),$$

$$B(x) = 1 - e^{1/2} x^{-\eta} (1/2 + \eta \ln x).$$

Luckock [31] also considers optimal trading strategies. We shall present these results in the next chapter.

**Question:** Luckock does not present any results for market depth (number of outstanding orders in the book at various prices in steady state, except in the separable case. Explore this problem. ■

**Question:** Explore the joint distribution of the market ask and bid in steady state. Note that we can get the expected bid-ask spread from Luckock's results, but not its distribution. ■

**Question:** Luckock model does not allow order cancellation. Explore a model where every limit buy (sell) order at price  $x$  leaves without execution with rate  $\theta_b(x)$  ( $\theta_a(x)$ ). ■

**Question:** Explore the steady state distributions in Luckock's model with the following assumption: the order book is limited to at most  $M$  out-standing sell orders and  $N$  outstanding buy orders. When the limit order book is full the newly arriving orders leave without execution. Even  $M = N = 1$  presents a non-trivial case. ■

## Chapter 11

# One-Sided Order Books: Strategic Traders

In this chapter we shall study the models of one sided order-books, i.e., order books where either the sellers or the buyers submit only market orders. We shall consider the sell side order book: the buyers submit market orders, while the sellers submit limit orders. The book thus contains a ranked list (or a queue) of the sell orders, and when a new buy order arrives it is matched with the sell order with the least ask price. Such a situation arises if the sellers are patient, while the buyers are impatient. To simplify the model, we shall assume that all order sizes equal one. The sellers are strategic, in the sense that each trader tries to optimize his own objective function. The models differ on what is assumed to be known to the sellers when they place an order, and how much freedom they have after they place an order. We shall begin with the simplest model in the next section.

### 11.1 Closed Sell Side Book: No information

In this section we consider a closed sell side order book. Thus the sellers have no information about the state of the order book. The buyers arrive according to  $PP(\lambda_b)$  and the sellers according to  $PP(\lambda_a)$ . Let  $A_n$  be the ask price selected by the  $n$ th trader. We shall assume that  $\{A_n, n \geq 0\}$  are iid random variables with support in  $[0, \infty)$  and cdf  $A(\cdot)$ . The sellers know these parameters.

Let  $X(t)$  be the number of sellers on the order book at time  $t$ .  $\{X(t), t \geq 0\}$  is the queue length process of an  $M/M/1$  queue with traffic intensity

$$\rho = \lambda_a / \lambda_b.$$

However, the service discipline is not first come first served: it is “least ask price first”. When a new trader arrives, he is served in this order. Since the arrival process is Poisson, we do not have to

worry about multiple arrivals at the same time. When a buyer arrives he buys at the ask price of the trader at the head of the line. If there are no sell orders, the buyer leaves without a transaction. (Alternatively, he buys from the dealer at price  $u$ .)

Let  $W_n$  be the time the  $n$ th sell order spends on the book, that is time between the arrival of the  $n$ th order and its execution. Using queueing theory terminology, we shall call  $W_n$  as the waiting time of the  $n$ th customer. We are interested in computing

$$T(a) = \lim_{n \rightarrow \infty} \mathbb{E}(W_n | A_n = a).$$

Thus  $T(a)$  is the expected waiting time, in steady state, of a customer with ask price  $a$ . We do not need to assume that  $\rho < 1$  to compute this quantity. The next theorem gives the main result.

**Theorem 11.1** *If the ask prices are continuous random variables, we have for  $a \geq 0$ :*

$$T(a) = \begin{cases} \frac{1}{\lambda_b(1-\rho A(a))^2} & \text{if } \rho A(a) < 1, \\ \infty & \text{if } \rho A(a) \geq 1. \end{cases} \quad (11.1)$$

*If the ask prices are discrete random variables taking values  $0 \leq a_1 < a_2 < \dots$ , we have for  $j \geq 1$ :*

$$T(a_j) = \begin{cases} \frac{1}{\lambda_b(1-\rho A(a_{j-1}))(1-\rho A(a_j))} & \text{if } \rho A(a_j) < 1, \\ \infty & \text{if } \rho A(a_j) \geq 1. \end{cases} \quad (11.2)$$

Here we assume  $A(a_0) = 0$ .

**Proof:** Consider a tagged trader with ask price  $a$ , arriving to a system in equilibrium at time zero. His expected waiting time consists of two components: waiting for the orders already on the book with ask prices below  $a$ , and the those who come during his waiting time with ask prices below  $a$ . Let  $0 \leq y < a$ . The expected number of traders who have arrived before him, and have ask prices in  $(y, y + dy)$  is given by  $\lambda_a T(y) dA(y)$ , while the expected number of traders who arrive after him with ask prices in  $(y, y + dy)$  is given by  $\lambda_a T(a) dA(y)$ . The tagged trader has to wait an expected time  $1/\lambda_b$  for each of these traders and an additional  $1/\lambda_b$  for himself. Hence, we get

$$T(a) = \frac{1}{\lambda_b} \left[ 1 + \int_0^a \lambda_a T(y) dA(y) + \int_0^a \lambda_a T(a) dA(y) \right].$$

This gives

$$(1 - \rho A(a))T(a) = \frac{1}{\lambda_b} + \rho \int_0^a T(y) dA(y).$$

Using the boundary condition  $T(0) = 1/\lambda_b$ , and assuming ask prices are continuous, this equation can be solved to get Equation 11.1. The discrete case follows by a similar argument. ■

The above result appears in various forms in three papers: Kleinrock [28], Phipps [39], and White and Christie [50]. Note that  $T(a)$  is always finite if  $\rho < 1$ . If  $\rho \geq 1$ ,  $T(a) < \infty$  for

$a < A^{-1}(1/\rho)$ , and  $T(a) = \infty$  if  $a \geq A^{-1}(1/\rho)$ , i.e., if the asking price is too large, then it will take an infinite time to execute the order on the average.

Next we consider the question of choosing a reasonable ask price cdf  $A(\cdot)$ . This analysis is primarily motivated by Kleinrock [28]. Let  $H_n$  be the cost incurred by the  $n$ th seller while waiting for one unit of time for the trade to execute. We call it the impatience factor of the  $n$ th trader. The higher the  $H_n$ , the more impatient is the  $n$ th trader. We shall assume that  $\{H_n, n \geq 0\}$  is a sequence of iid continuous random variables with support  $[0, \infty)$  and cdf  $F(\cdot)$ . A trader with impatience factor  $h$  and ask price  $a$  gets an expected net benefit of  $a - hT(a)$  when his sell order is executed. We assume that a trader knows his own impatience factor  $h$ , and chooses an ask price  $a$  based on it. Now, it makes sense that the net benefit to all the traders must be the same, say a fixed constant  $p$ , else the traders with smaller benefits would try to use a different  $a$ . Hence we assume that the traders choose

$$a = hT(a) + p.$$

Hence the cdf of the ask price satisfies

$$A(a) = F((a - p)/T(a)) = F(\lambda_b(a - p)(1 - \rho A(a))^2). \quad (11.3)$$

The next question is: how do the traders choose  $p$ ? Clearly, the traders would like to choose as high a  $p$  as possible. Of course a higher  $p$  will result in a lower  $\lambda_b$ . So the appropriate model is to assume that a given  $p$  induces a buyer arrival rate  $\lambda_b(p)$ , which is a decreasing function of  $p$ . Then the traders will choose a  $p$  that will maximize the expected net benefit of a typical trader. This will happen by choosing a  $p$  such that

$$\lambda_b(p) = \lambda_a, \quad (11.4)$$

or  $\rho(p) = \lambda_a/\lambda_b(p) = 1$ . This ensures that all orders are eventually executed. Any  $p$  higher than this will cause some of the orders to stay unexecuted forever.

**Example 11.1** Suppose the patience factors  $H_n$  are  $U(0, 1)$ . Then  $F(x) = x$ , and hence Equation 11.3 yields a quadratic in  $A(x)$ , that can be solved to get

$$A(a) = \frac{1}{2\rho^2} \left[ 2\rho - \frac{1}{\lambda_b(a - p)} \left( \sqrt{1 + 4\rho\lambda_b(a - p)} - 1 \right) \right]. \quad (11.5)$$

The range of the above cdf is  $[p, p + 1/\lambda_b(1 - \rho)^2]$  if  $\rho < 1$ , and  $[p, \infty)$  if  $\rho \geq 1$ . Next suppose

$$\lambda_b(p) = \alpha - \beta p$$

where  $\alpha > \lambda_a$  and  $\beta > 0$ . Hence, Equation 11.4 gives  $p = (\alpha - \lambda_a)/\beta$ . For this choice of  $p$ , Equation 11.5 reduces to

$$A(a) = 1 - \frac{1}{2\lambda_a(a - p)} \left( \sqrt{1 + 4\lambda_a(a - p)} - 1 \right), \quad a \geq p.$$

Substituting in Equation 11.1 we get

$$T(a) = \frac{4\lambda_a(a-p)^2}{(\sqrt{1+4\lambda_a(a-p)}-1)^2}, \quad a \geq p.$$

**Question:** Do the analysis of the previous example assuming that the patience factor is discrete:  $h_1$  with probability  $\alpha$  and  $h_2 > h_1$  with probability  $1 - \alpha$ . ■

**Question:** Is the cdf  $A$  of Equation 11.3 a symmetric Nash equilibrium? That is, suppose all sellers use this  $A$  for placing orders. Now consider a tagged customer with random patience factor arrives at such a queue in steady state. Will it be optimal for him to choose his ask price using the cdf  $A$ ? ■

## 11.2 Closed Sell Side Book: Partial Information

Let  $A(t)$  be the market ask price at time  $t$ . It is the least of the ask prices of all the sellers on the book. If the book is empty at time  $t$ , we set  $A(t) = \infty$ . We now consider the case where the sellers do not know the state of the book in detail, however, they know the current market ask price. The sellers take this information into account when placing an order. Assume that all sellers are iid and use an ask price distribution  $A_x(\cdot)$  if the market ask price is  $x$  at the time of their arrival. We need to consider two further cases: in the first case, the seller's decision is irrevocable. His order stays on the book until execution. In the second case, the seller has an option of canceling his existing order and placing a new order whenever the market ask price changes.

**Question:** Analyze these two models. I do not see any results about it in the literature.

## 11.3 Open Sell Side Book: Seppi Model

In this section we study a slightly modified version of the model considered by Seppi [45]. Consider an open sell side book maintained by a dealer for a given security. We consider a single period mode, and assume without a loss of generality that the fundamental value of the security is 0. The state of the book (the number of active sell orders and their respective ask prices) is open to everyone. The sellers place limit orders at prices predetermined set of prices  $0 < p_1 < p_2 < \dots < p_n < \infty$ . We also define  $p_0 = 0$ .

Initially patient sellers place limit sell orders, with  $S_j$  being the number of limit orders placed at price  $p_j$ . The state of the order book is thus  $S = [S_1, S_2, \dots, S_n]$ . Let

$$Q_j = \sum_{i=1}^j S_i$$

be the cumulative limit sell orders with limit price  $p_j$  or less. Then a single (aggregate) market buy order arrives for  $B$  shares.  $B$  is a random variable with a distribution known to all, and it is independent of  $S$ . Based on  $S$  and  $B$ , the dealer decides to intervene. He can intervene at any price  $p_k$  and sell  $(B - Q_k)^+$  shares and make a profit of

$$\pi_k = p_k(B - Q_k)^+. \quad (11.6)$$

This reflects the order execution priority that all limit orders at  $p_j \leq p_k$  are satisfied at their respective limit prices before the dealer can make any sales. The dealer chooses the price  $p^*(S, B)$  to be the  $p_k$  that maximizes his expected profit. Equation 11.6 implies that there are thresholds  $\hat{Q}_j = \hat{Q}_j(S)$  such that

$$0 = \hat{Q}_0 \leq \hat{Q}_1 \leq \hat{Q}_2 \leq \dots \leq \hat{Q}_n \leq \hat{Q}_{n+1} = \infty$$

and

$$B \in [\hat{Q}_j, \hat{Q}_{j+1}) \Rightarrow p^*(B) = p_j, \quad 0 \leq j \leq n.$$

Note that some the above intervals might be empty. This completely describes the optimal policy for the dealer given  $S$  and  $B$ .

Next we describe the optimal policy for the limit order sellers. The limit order sellers are expected to pay a per share cost of  $c$  for submitting the order. Thus the expected profit from a limit order placed at  $p_j$  is

$$p_j \mathbf{P}(B > \hat{Q}_j) - c.$$

They know how the dealer behaves, and behave in a competitive manner. That is, they choose  $S$  so that the expected profit from any order is zero. (If the profit were positive, more limit sellers will join the order book.) The main result is given in the next theorem. We use the following notation:

$$H_j = \inf\{x \geq 0 : \mathbf{P}(B > x) \geq c/p_j\}, \quad j \geq 1.$$

If  $c/p_j \geq 1$ , we define  $H_j = 0$ .

$$\gamma_j = \frac{p_j - p_{j-1}}{p_j}, \quad 1 \leq j \leq n.$$

**Theorem 11.2** *The equilibrium state of the order book is given by*

$$Q_1 = H_1,$$

$$Q_j = \gamma_j H_j + (1 - \gamma_j) H_{j-1}, \quad 2 \leq j \leq n.$$

**Proof:** See Seppi [45]. ■

The state of the order book  $S$  can be computed from the  $Q$ 's given above. We illustrate with an example.

**Example 11.2** Suppose  $B \sim \exp(\theta)$ , and  $p_j = j\Delta$ , where  $\Delta$  is the tick size. Then we can write the equilibrium state of the order book as

$$S_1 = \theta \ln(\Delta/c), \quad S_j = \frac{\theta}{j} \left( \ln(j) - \frac{1}{j-1} \sum_{i=1}^{j-1} \ln(i) \right), \quad 2 \leq j \leq n. \quad \blacksquare$$

## 11.4 Open Sell Side Book: Rosu Model

In this section we study a slightly modified version of the model considered by Rosu [42]. Consider an open sell side book maintained by a dealer. The state of the book (the number of active sell orders and their respective ask prices) is open to everyone. The sellers are patient and place limit orders for one unit each, while the buyers are impatient, and buy one unit each at the lowest ask price on the book. The dealer stands ready to buy any amount at  $B$  and sell any amount at  $A > B$ . One can think of  $B$  as the reservation price of the sellers, below which no seller wants to sell. The dealer has complete control over  $A$ . Clearly, the higher the  $A$ , more seller will be willing to put in a limit sell order on the dealer's book. The dealer chooses  $A$  so that there are at most  $K$  sell orders on the book at any time. We shall describe in detail how this is managed.

As in Section 11.1, we assume that the buyers with market orders arrive according to  $PP(\lambda_b)$  and the sellers with limit orders according to  $PP(\lambda_a)$ . Let  $X(t)$  be the number of active sell orders on the book at time  $t$ . Clearly  $\{X(t), t \geq 0\}$  is the queue length process of an  $M/M/1/K$  queue with arrival rate  $\lambda_a$  and departure rate  $\lambda_b$ . Let  $a(t) = [a_1, a_2, \dots, a_{X(t)}]$  be the vector of ask prices in decreasing order. All prices are in the interval  $[B, A]$ .

If  $X(t) = k > 0$  when a buyer arrives, he is matched with a sell order at price  $a_k$  immediately, and the number of sell orders reduces by one. At this point each of the remaining  $k - 1$  sellers may decide to either cancel his order or change his ask price. If  $X(t) = 0$ , the buy order is lost (or satisfied at  $A$  by the dealer). When a sell order arrives at time  $t$  and sees  $X(t-) = k < K$  orders already on the book, the seller decides to choose his ask price based on  $X(t-)$  and  $a(t-)$ . This may also result in other sellers canceling or altering their orders. The main aim of this model is to analyze how the sellers behave in response to the dynamics of the order book.

We use the impatience model from Section 11.1, and assume that the impatience factor for all traders is a fixed positive constant  $h$ . If a seller has an order on the book at ask price  $a$ , and he has to wait  $T$  time units before his order gets executed, his expected net benefit is

$$a - hE(T).$$

When a seller arrives at an empty order book, he will clearly place an order at  $A$ . Now, the next event may be an arrival of a buyer or another seller. In the former case the seller sells at  $A$ . In the



latter case, the new seller places an order at  $a$ , in response to which the first seller changes his ask price. Clearly the new prices of the two sellers must be such that their expected net benefits are the same. This gives us the general concept of “(non-cooperative) equilibrium”.

**Definition 11.1** *The prices on the order book are said to be in (non-cooperative) equilibrium if the expected net benefit is the same to all the traders on the book.*

The term “non-cooperative” brings out the fact that the traders are not allowed to collude in setting the prices. We shall drop it from now on. If the book is in equilibrium, no trader will change his order by himself, since that will bring on changes by others.

Since the buyer always buys from the least ask price, the expected net benefit depends only on the least ask price. Let  $f_k$  be the expected net benefit of the traders when there are  $k$  orders on the order book. The next theorem gives the main result.

**Theorem 11.3** *Suppose  $K, B, \lambda_a, \lambda_b$  and  $h$  are fixed. Define*

$$\lambda = \lambda_a + \lambda_b, \quad p = \lambda_a/\lambda, \quad q = \lambda_b/\lambda.$$

*Then  $A$  is chosen so that the following equations hold:*

$$f_0 = A \tag{11.7}$$

$$f_k = pf_{k+1} + qf_{k-1} - h/\lambda, \quad 1 \leq k < K, \tag{11.8}$$

$$f_K = f_{K-1} - h/\lambda_a, \tag{11.9}$$

$$f_K = B. \tag{11.10}$$

**Proof:** Follows from the first step analysis. ■

The next theorem gives the solution to the above equations. First define

$$\beta_k = \begin{cases} \frac{1-\rho^k}{1-\rho} & \text{if } \rho \neq 1, \\ k & \text{if } \rho = 1. \end{cases}$$

**Theorem 11.4** *The Equations 11.7-11.10 have the following solution if  $\rho \neq 1$ :*

$$A = B + \frac{h}{\lambda_a} \left[ \beta_K \left( \beta_{K-1} + \frac{2-\rho}{1-\rho} \right) - \frac{K}{1-\rho} \right]$$

$$f_0 = A,$$

$$f_1 = A - \frac{h}{\lambda_a} (1 + \beta_{K-1})$$

$$f_k = A(1 - \beta_k) + f_1 \beta_k + \frac{h}{\lambda_a} \frac{K - \beta_K}{1 - \rho}, \quad 2 \leq k \leq K - 1$$

$$f_K = B.$$

If  $\rho = 1$ , the solution is given by

$$\begin{aligned} A &= B + \frac{h}{\lambda_a} K^2 \\ f_k &= B + \frac{h}{\lambda_a} K(K - k), \quad 0 \leq k \leq K. \end{aligned}$$

**Question:** Verify the above solution. ■

With the solution in the above theorem we are ready to identify one equilibrium for this order book, given in the following theorem.

**Theorem 11.5** *If an arriving seller sees  $k$  sell orders on the book (not including himself), it is optimal for him to place a sell order at ask price  $f_k$  if  $0 \leq k \leq K - 1$ , and never change the order until its execution. If he sees  $K$  orders on the book, it is optimal for him not to enter an order.*

There are many other equilibria possible for this order book. However, they all produce the same expected net benefit for the sellers. The model we have considered here is a slightly modified version of the model considered by Rosu [42]. He assumes that  $A$  and  $B$  are fixed. This necessitates the use of mixed strategies by some of the traders. We assume that  $B$  is fixed but  $A$  is under the dealer's control. This avoids mixed strategies and provides a cleaner solution.

Let  $Q(t) = (X(t) - 1)^+$ . Then we see that the market ask price at time  $t$  is given by  $A(t) = f_{Q(t)}$ , and the bid-ask spread is  $S(t) = A(t) - B$ . Since the stochastic nature of the  $Q$  process is known, one can compute the stochastic nature of the  $A$  and  $S$  processes.

**Question:** Compute the mean and variance of the bid-ask spread in steady state. The variance can be thought of as the volatility of the market. Study its dependence on the model parameters  $\lambda_a$ ,  $\lambda_b$ ,  $K$ ,  $B$  and  $h$ . ■

**Question:** Compute the auto covariance function of the bid-ask spread assuming that the system starts in steady state. ■

**Question:** Extend the model to the case where the sellers have different patience factors. ■

**Question:** Extend the model to the case where the market buy orders arrive according to a compound Poisson process. ■

**Question:** Extend the model to the case where the sell orders arrive according to a compound Poisson process. ■

**Question:** Extend the model to the case where the arrival rate of the buyers depends on the market ask price. Since  $A(t)$  depends on  $X(t)$  in a deterministic fashion, this is same as assuming that the buyer arrival rate is  $\lambda_b(k)$  when there are  $k$  orders on the order book. ■

Rosu [42] also studies the general two sided book. The analysis gets more complicated.

## Chapter 12

# Optimal Execution Strategies

Suppose a large institutional investor wants to buy a million shares in a particular stock over three days. How should it do it? One method is to put a market order for a million shares with a dealer. This will typically immediately wipe out the sell side of the order book, and the buyer will end up paying a very high price as the news of the order spreads and the dealer jacks up the ask price. The other method is to put in a limit buy order at a low bid price and slowly change the price to attract more and more sell orders. Even this will cause the buyer to pay high price over all. Another method is to split the order into smaller limit or market orders and execute them one by one. In this chapter we study the optimal policies to execute such a large trade. The model needs to take into account the price impact of the trading policy, and the tradeoff between market and limit buy orders.

### 12.1 Optimal Splitting: Bertsimas Model

In this section we study the models from Bertsimas and Lo [4], where they assume that the buyer uses only market orders. Suppose the buyer has the opportunity to buy at time  $1, 2, \dots, T$ , where  $T$  is a given fixed positive integer. He plans to buy a total of  $S$  shares using those  $T$  opportunities. Let  $S_t$  be the number of shares the buyer needs to buy over  $t, t + 1, \dots, T$ . Based on the state  $(P_{t-1}, S_t)$  at time  $t$ , the buyer places a market order of size  $B_t$  at time  $t$ . The buyer pays a price of  $P_t$  per share, where  $P_t$  is the new price reflecting the impact of his order  $B_t$ , as well as the new information that becomes available at time  $t$ . The general dynamics of the system state is given by

$$S_1 = S, \quad S_{t+1} = S_t - B_t, \quad 1 \leq t \leq T, \quad S_{T+1} = 0,$$

$$P_t = f(P_{t-1}, X_t, B_t, \epsilon_t), \quad X_t = g(X_{t-1}, \eta_t), \quad 1 \leq t \leq T,$$

where  $\{\epsilon_t, 1 \leq t \leq T\}$  and  $\{\eta_t, 1 \leq t \leq T\}$  are two independent sequences of iid zero mean random variables. The  $f$  function describes how the price evolution depends on the size of the buy order (the price impact), and also the dependence on the history through the explanatory variable  $X_t$ .

The cost at time  $t$  is  $P_t B_t$ .

Thus we can model  $\{(P_{t-1}, X_{t-1}, S_t), B_t), t \geq 0\}$  as a Markov decision process. The objective is to minimize the expected total cost over  $\{1, 2, \dots, T\}$ . Define

$$V_t(p, x, s) = \inf \mathbf{E} \left( \sum_{k=t}^T P_k B_k \mid P_{t-1} = p, X_{t-1} = x, S_t = s \right)$$

where the infimum is taken over all possible admissible policies over  $\{t, t+1, \dots, T\}$ , starting in state  $(p, x, s)$ . The Bellman equation is given by

$$V_{T+1}(p, x, 0) = 0, \quad V_{T+1}(p, x, s) = \infty \text{ if } s \neq 0,$$

$$V_t(p, x, s) = \min_{0 \leq b \leq s} \{ \mathbf{E}(f(p, g(x, \eta_t), b, \epsilon_t) b) + \mathbf{E}(V_{t+1}(f(p, g(x, \eta_t), b, \epsilon_t), g(x, \eta_t), s - b)) \}, \quad 1 \leq t \leq T.$$

The  $b$  that achieves the minimum in the  $t$ -th equation above gives the optimal decision  $B_t$  in state  $(p, x, s)$ , and  $V_1(p_0, x_0, S)$  gives the optimal purchase cost over  $\{1, 2, \dots, T\}$ . We treat two special cases below.

### 12.1.1 Additive Permanent Price Impact

Suppose the price evolution is given by

$$P_0 = p_0, \quad P_t = P_{t-1} + \theta B_t + \epsilon_t, \quad 1 \leq t \leq T,$$

where  $\theta > 0$  is a given constant. Thus there is no explanatory variable  $X_t$ , and the impact of the market order size is linear and permanent. In this case the the state is given by  $(p, s)$ , and the Bellman Equation reduces to

$$V_{T+1}(p, 0) = 0, \quad V_{T+1}(p, s) = \infty \text{ if } s \neq 0,$$

$$V_t(p, s) = \min_{0 \leq b \leq s} \{ b(p + \theta b) + \mathbf{E}(V_{t+1}(p + \theta b + \epsilon_t, s - b)) \}, \quad 1 \leq t \leq T.$$

The next theorem gives a complete solution to these equations.

**Theorem 12.1** *The optimal order size in period  $t$  with  $(P_{t-1}, S_t) = (p, s)$  is given by*

$$B_t(p, s) = \frac{s}{T - t + 1}, \quad 1 \leq t \leq T,$$

and the optimal cost is given by

$$V_t(p, s) = sp + \frac{\theta s^2}{2} \frac{T - t + 2}{T - t + 1}, \quad 1 \leq t \leq T.$$

**Proof:** Follows by direct verification. See Bertsimas and Lo [4]. ■

Thus starting in state  $(p_0, S)$ , the optimal strategy is to split the order evenly in  $T$  batches and place a market order of size  $S/T$  in each of the  $T$  periods. The total price paid under this strategy is

$$V_1(p_0, S) = Sp_0 + \frac{\theta S^2 T + 1}{2T}.$$

Thus, as  $T \rightarrow \infty$ , we see that

$$V_1(p_0, S) \rightarrow Sp_0 + \frac{\theta S^2}{2} \geq Sp_0.$$

Thus even if we have unlimited number of opportunities to buy the total stock  $S$ , we still pay more than  $Sp_0$ , and the price impact cannot be avoided. This is the consequence of permanent price impact.

### 12.1.2 Additive Permanent Price Impact with Information

Next we consider a price dynamics given by

$$P_0 = p_0, \quad X_0 = x_0,$$

$$P_t = P_{t-1} + \theta B_t + \gamma X_t + \epsilon_t, \quad 1 \leq t \leq T,$$

$$X_t = \rho X_{t-1} + \eta_t, \quad 1 \leq t \leq T,$$

where  $-\infty < \gamma < \infty$ ,  $\theta > 0$ , and  $-1 < \rho < 1$  are given parameters. As before  $\{\epsilon_t, 1 \leq t \leq T\}$  is a sequence of iid random variables with mean zero and variance  $\sigma_\epsilon^2$ , and  $\{\eta_t, 1 \leq t \leq T\}$  is an independent sequence of iid random variables with mean zero and variance  $\sigma_\eta^2$ . One can think of  $X_t$  as the relevant market information that arises in period  $t$  that affects the price in period  $t$ .

The Bellman equations reduce to

$$V_{T+1}(p, x, 0) = 0, \quad V_{T+1}(p, x, s) = \infty \text{ if } s \neq 0,$$

$$V_t(p, x, s) = \min_{0 \leq b \leq s} \{b(p + \theta b + \gamma \rho x) + \mathbb{E}(V_{t+1}(p + \theta b + \gamma \rho x + \gamma \eta_t + \epsilon_t, \rho x + \eta_t, s - b))\}, \quad 1 \leq t \leq T.$$

The solution is given in the next theorem. First some notation:

$$a_0 = \theta, \quad a_k = \frac{\theta(k+2)}{2(k+1)},$$

$$b_0 = \gamma, \quad b_k = \gamma + \frac{\theta \rho b_{k-1}}{2a_{k-1}},$$

$$c_0 = 0, \quad c_k = \rho^2 c_{k-1} - \frac{\rho^2 b_{k-1}^2}{4a_{k-1}},$$

$$d_0 = 0, \quad d_k = d_{k-1} + c_{k-1} \sigma_\eta^2,$$

$$e_0 = 1, \quad e_k = \frac{1}{k+1},$$

$$f_0 = 0, \quad f_k = \frac{\rho b_{k-1}}{2a_{k-1}}.$$

Note that  $\theta > 0$  implies that all  $a$ 's are positive, and all  $c$ 's and  $d$ 's are negative.  $b$ 's can vary, but are positive if  $\gamma > 0$  and  $\rho > 0$ .

**Theorem 12.2** *The optimal order size in period  $t$  with  $(P_{t-1}, X_{t-1}, S_t) = (p, x, s)$  is given by*

$$B_t(p, x, s) = e_{T-t}s + f_{T-t}x, \quad 1 \leq t \leq T,$$

and the optimal cost is given by

$$V_t(p, x, s) = sp + a_{T-t}s^2 + b_{T-t}xs + c_{T-t}x^2 + d_{T-t}, \quad 1 \leq t \leq T.$$

Thus the optimal strategy asks the buyer to always divide the remaining stock demand into equal parts, and then adjust it according to the current information. See Bertsimas and Lo [4] for various implications of the above result.

### 12.1.3 Multiplicative Temporary Price Impact

Next we consider a price dynamic given by

$$\begin{aligned} \tilde{P}_0 &= p_0, \quad X_0 = x_0, \\ P_t &= (1 + \theta B_t + \gamma X_t)\tilde{P}_t, \quad 1 \leq t \leq T, \\ \tilde{P}_t &= \tilde{P}_{t-1} \exp(\epsilon_t), \quad 1 \leq t \leq T, \\ X_t &= \rho X_{t-1} + \eta_t, \quad 1 \leq t \leq T, \end{aligned}$$

where  $-\infty < \gamma < \infty$ ,  $\theta > 0$ , and  $-1 < \rho < 1$  are given parameters. Here  $\{\epsilon_t, 1 \leq t \leq T\}$  is a sequence of iid random variables with mean  $\mu_\epsilon$ , variance  $\sigma_\epsilon^2$  and  $q = \mathbf{E}(\exp(\epsilon_t))$ , and  $\{\eta_t, 1 \leq t \leq T\}$  is an independent sequence of iid random variables with mean zero and variance  $\sigma_\eta^2$ . One can think of  $X_t$  as the relevant market information that arises in period  $t$  that affects the price in period  $t$ . Note that the impact of  $B_t$  and  $X_t$  is felt only during period  $t$ , and not during any later periods. Hence we call this ‘‘temporary impact’’ model. The multiplicative model for  $\tilde{P}_t$  is inspired by the geometric Brownian motion models of asset prices.

The state at time  $t$  is now given by  $(\tilde{P}_{t-1}, X_{t-1}, S_t)$ . The Bellman equations reduce to

$$V_{T+1}(p, x, 0) = 0, \quad V_{T+1}(p, x, s) = \infty \text{ if } s \neq 0,$$

$$V_t(p, x, s) = \min_{0 \leq b \leq s} \{b(1 + \theta b + \gamma \rho x)pq + \mathbf{E}(V_{t+1}(p \exp(\epsilon_t), \rho x + \eta_t, s - b))\}, \quad 1 \leq t \leq T.$$

The solution is qualitatively similar to the previous model with additive permanent impact with information, and we refer the students to Bertsimas and Lo [4] for complete results.

## 12.2 Optimal Splitting and Timing: OW model

The Bertsimas model of the previous section considered optimal splitting of a large order, but the times of placement were constrained to be  $1, 2, \dots, T$ . In this section we consider a model studied by Obizhaeva and Wang [36] that allows the size of the split orders as well as the timing of order placement as decision variables. It also incorporates a more detailed model of the limit order book dynamics. Also see Almgren [1] for another relevant paper.

As before, we consider a decision problem faced by a buyer who wants to purchase a fixed amount  $B$  of stock by optimally placing market orders with a dealer. Let  $A_t$  be the ask price in the dealer's book at time  $t$ . Assume that the buyer wants to conclude his business in  $N$  trades over the time interval  $[0, T]$ . Let  $B_n$  be the size of the market buy order placed by the buyer at time  $t_n$ . The strategy of the buyer is given by the two vectors  $(t_1, t_2, \dots, t_N)$  and  $(B_1, B_2, \dots, B_N)$ . Of course we must have

$$0 \leq t_1 < t_2 < \dots < t_N \leq T,$$

and

$$B_i \geq 0, \quad B_1 + B_2 + \dots + B_N = B.$$

Next we assume that number of limit sell orders in the book with limit prices between  $A_t$  and  $A_t + \delta$  ( $\delta > 0$ ) is  $q\delta$ , where  $q > 0$  is a given constant, reflecting the depth of the order book. Thus an order of  $B_n$  at time  $t_n$  instantaneously removes  $B_n$  orders that occupy the price range  $[A_{t_n}, A_{t_n} + B_n/q]$ . Thus the total cost of this buy order is  $B_n(A_{t_n} + \frac{B_n}{2q})$ . The limit sell orders are always added or removed at  $A_t$  thus creating a change in  $A_t$ . We assume that  $A_t$  is given by

$$A_t = F_t + \sum_{n=1}^N (\lambda(1 - e^{-\rho(t-t_n)}) + \frac{1}{q}e^{-\rho(t-t_n)})B_n 1_{\{t_n \geq t\}},$$

where  $\{F_t, t \geq 0\}$  is an exogenously specified price process, representing the ask price evolution in the absence of the buy orders from our decision maker. Here  $\lambda$  and  $\rho > 0$  are fixed parameters modeling the price impact of the buy order. Thus an order of  $B_n$  at time  $t_n$  immediately raises the ask price by  $B_n/q$ , but it asymptotically approaches  $\lambda B_n$  with decay rate  $\rho$ . Thus the price impact becomes more fleeting as  $\rho$  increases.

The optimization problem can now be written as

$$\text{Minimize } \mathbb{E} \left( \sum_{n=1}^N (A_{t_n} + B_n/q)B_n \right),$$

where the minimization is done over all strategies that make decisions based only on the information available at the point of decision making.



### 12.2.1 Special Case: Discrete Trading Times

In this section we consider a special case where the trading is constrained to occur at times  $t_n = n\tau$  ( $0 \leq n \leq N$ ) where  $\tau = T/N$ . We formulate and solve this problem as a Markov Decision Process. The decision at time  $n$  is  $B_n$ , the amount of stock to buy at time  $n$ . Let

$$S_0 = B, \quad S_n = B - \sum_{k=0}^{n-1} B_k, \quad 1 \leq n \leq N$$

be the amount of stock that has to be purchased over time  $n, n+1, \dots, N$ . We must satisfy the constraint  $S_{N+1} = 0$ . The random walk component of the price is given by

$$F_0 = f_0, \quad F_{n+1} = F_n + \epsilon_{n+1}, \quad n \geq 0$$

where  $\{\epsilon_n, n \geq 0\}$  is a sequence of iid random variables with mean zero and variance  $\sigma^2\tau$ . The ask price is given by

$$A_0 = f_0; \quad A_n = F_n + \sum_{k=0}^{n-1} (\lambda(1 - e^{-\rho\tau(n-k)}) + \frac{1}{q}e^{-\rho\tau(n-k)})B_k, \quad n \geq 1.$$

Let  $\kappa = 1/q - \lambda$ , and define

$$D_0 = 0, \quad D_n = \kappa \sum_{k=0}^{n-1} B_k e^{-\rho\tau(n-k)}, \quad n \geq 1.$$

Then, we can write

$$\begin{aligned} D_{n+1} &= e^{-\rho\tau}(\kappa B_n + D_n), \\ A_n &= F_n + \lambda(B - S_n) + D_n. \end{aligned}$$

We see that the state of the system at time  $n$  is completely described by  $(F_n, S_n, D_n)$ . If the state of the system is  $(f, s, d)$  at time  $n$ , and an order is placed for  $B_n = b$  units, the next state is given by

$$(F_{n+1}, S_{n+1}, D_{n+1}) = (f + \epsilon_{n+1}, s - b, e^{-\rho\tau}(\kappa b + d)).$$

The cost of this purchase is

$$c(f, s, d; b) = b(A_n + b/(2q)) = b(f + \lambda(B - s) + d + b/(2q)).$$

Now let  $V_n(f, s, d)$  be the purchase cost of the optimal policy over  $n, n+1, \dots, N$  if the state at time  $n$  is  $(f, s, d)$ . The DP equations become

$$\begin{aligned} V_{N+1}(f, 0, d) &= 0, \quad V_{N+1}(f, s, d) = \infty, \quad \text{if } s \neq 0, \\ V_n(f, s, d) &= \min_{0 \leq b \leq s} \{c(f, s, d; b) + \mathbf{E}(V_{n+1}(f + \epsilon_{n+1}, s - b, e^{-\rho\tau}(\kappa b + d)))\}, \quad 0 \leq n \leq N. \end{aligned}$$

These equations can be solved recursively in a backward fashion to get the solution given in the following theorem. First some notation:

$$\begin{aligned}\alpha_N &= \frac{1}{2q} - \lambda, \quad \beta_N = 1, \quad \gamma_N = 0, \\ \delta_n &= \left[ \frac{1}{2q} + \alpha_n - \beta_n \kappa e^{-\rho\tau} + \gamma_n \kappa^2 e^{-2\rho\tau} \right]^{-1}, \quad 0 \leq n \leq N, \\ \alpha_n &= \alpha_{n+1} - \frac{1}{4} \delta_{n+1} (\lambda + 2\alpha_{n+1} - \beta_{n+1} \kappa e^{-\rho\tau})^2, \quad 0 \leq n \leq N-1, \\ \beta_n &= \beta_{n+1} e^{-\rho\tau} + \frac{1}{2} \delta_{n+1} (1 - \beta_{n+1} \kappa e^{-\rho\tau} + 2\kappa \gamma_{n+1} e^{-2\rho\tau}) (\lambda + 2\alpha_{n+1} - \beta_{n+1} \kappa e^{-\rho\tau}), \quad 0 \leq n \leq N-1, \\ \gamma_n &= \gamma_{n+1} e^{-2\rho\tau} - \frac{1}{4} \delta_{n+1} (1 - \beta_{n+1} \kappa e^{-\rho\tau} + 2\kappa \gamma_{n+1} e^{-2\rho\tau})^2, \quad 0 \leq n \leq N-1.\end{aligned}$$

**Theorem 12.3** *The optimal decision in state  $(F_n, S_n, D_n) = (f, s, d)$  is given by*

$$B_n(f, s, d) = -\frac{1}{2} \delta_{n+1} [d(1 - \beta_{n+1} \kappa e^{-\rho\tau} + 2\kappa \gamma_{n+1} e^{-2\rho\tau}) - s(\lambda + 2\alpha_{n+1} - \beta_{n+1} \kappa e^{-\rho\tau})], \quad 0 \leq n < N,$$

and

$$B_N(f, s, d) = s.$$

The optimal expected cost over  $n, n+1, \dots, N$  is

$$V_n(f, s, d) = (f + \lambda B)s + \alpha_n s^2 + \beta_n ds + \gamma_n d^2, \quad 0 \leq n \leq N.$$

**Proof:** Verified by direct substitution. See Obizhaeva and Wang [36]. ■

It is interesting to study the limiting behavior of the above policy as  $N \rightarrow \infty$  for a fixed  $T$ . This is equivalent to allowing trading times as decision variables. We show in the following theorem that, in the limit, the optimal policy is to buy an amount  $B_0$  at time 0,  $B_T$  at time  $T$ , and buy the securities continuously at rate  $b$  over  $(0, T)$ . The main result is given in the next Theorem.

**Theorem 12.4** *The parameters of the optimal policy in continuous is given by*

$$\begin{aligned}B_0 &= B_T = \frac{B}{\rho T + 2}, \\ b &= \frac{\rho B}{\rho T + 2}.\end{aligned}$$

The optimal expected cost over  $[t, T]$  is

$$V_t(f, s, d) = fs + \lambda Bs + \alpha_t s^2 + \beta_t sd + \gamma_t d^2, \quad 0 \leq t \leq T$$

where

$$\alpha_t = \frac{\kappa}{\rho(T-t) + 2} - \frac{\lambda}{2}, \quad \beta_t = \frac{2}{\rho(T-t) + 2}, \quad \gamma_t = \frac{\rho(T-t)}{2\kappa(\rho(T-t) + 2)}.$$

**Proof:** Follows from taking limits and identifying

$$b = \lim_{N \rightarrow \infty} \frac{B_n}{T/N}.$$

See Obizhaeva and Wang [36]. ■

One can obtain the same solution by directly working with a continuous time model and assuming that  $\{F_t, t \geq 0\}$  is a Brownian motion.

### 12.3 Optimal Order Placements: Harris Model

In this section we consider a model based on Section 15.2 of Hasbrouck [22], which in turn is based on Harris [20]. Let  $A(t)$  be the ask price at time  $t$ . Suppose that  $\{A(t), t \geq 0\}$  is a Brownian motion with drift  $\mu$  and variance parameter  $\sigma$ . Decisions have to be made at a predetermined discrete set of time points  $0 \leq t_0 < t_1 < \dots < t_{T-1} < t_T \leq T$ . We shall consider the special case where  $T$  is a positive integer and  $t_k = k$  for  $k = 0, 1, \dots, T$ .

At each time  $k$ , the trader can place a market order and pay  $A(k)$ , in which case the problem terminates. Otherwise, the buyer can place a limit order at price  $B_k < A(k)$ . The limit order gets executed by time  $t + 1$  if there is a  $\tau \in [k, k + 1)$  such that  $A(\tau) = B_k$ , in which case it is executed and the problem terminates, else it is canceled, and the trader faces the same options again at time  $k + 1$ . We develop DP equations for this problem below.

We first collect the following facts from Brownian motion. Let  $\Phi$  and  $\phi$  be the cdf and pdf of a standard normal random variable. Let  $H$  be the event

$$H(b) = \{A(\tau) = b \text{ for some } \tau \in [0, 1)\}.$$

For  $b < a$  let

$$\alpha(a, b) = \mathbb{P}(H(b) | A(0) = a)$$

be the probability that a limit buy order at price  $b$  placed at time zero will execute before time 1. We can show that

$$\alpha(a, b) = 1 - \Phi\left(\frac{a - b + \mu}{\sigma}\right) - \exp\left(\frac{2\mu(b - a)}{\sigma^2}\right) \Phi\left(\frac{b + a - \mu}{\sigma}\right), \quad b < a.$$

Next let  $f(x|a, b)/(1 - \alpha(a, b))$  be the conditional density of  $A(1)$  given  $A(0) = a$  and that the event  $H(b)$  does not occur. We can show that

$$f(x|a, b) = \left[ \phi\left(\frac{\mu + a - x}{\sigma}\right) - \exp\left(\frac{2\mu(b - a)}{\sigma^2}\right) \phi\left(\frac{\mu + (b - x) + (b - a)}{\sigma}\right) \right] / \sigma.$$

The expectation of the above conditional density is

$$\beta(a, b) = \mu + 2(a - b) \exp\left(\frac{2\mu(b - a)}{\sigma^2}\right) \Phi\left(\frac{b - a + \mu}{\sigma}\right) / (1 - \alpha(a, b)).$$

Let  $V_k(a)$  be the minimum expected cost of purchasing one unit of security by time  $T$  starting with  $A(k) = a$ . With the above notation we can write the DP equations as

$$V_T(a) = a,$$

$$V_k(a) = \min\{a, G_k(a)\}, \quad 0 \leq k \leq T - 1,$$

where

$$G_k(a) = \min_{b < a} \{b\alpha(a, b) + \int_b^\infty V_{k+1}(x) f(x|a, b) dx\}, \quad 0 \leq k \leq T - 1.$$

# Bibliography

- [1] Almgren, R. (2003). Optimal Execution with Nonlinear Impact Functions and Trading-Enhanced Risk. *Applied Mathematical Finance*, Vol. 10, 118.
- [2] Amihud, Y. and H. Mendelson. (1980). Dealership Markets: Market Making with Inventory. *Journal of Financial Economics*, Vol 8, 31-53.
- [3] Bagehot, W. (1971). The Only Game in Town. *Financial Analysts Journal*, Vol. 27,12-14+22.
- [4] Bertsimas, D. and A. Lo. (1998). Optimal control of execution costs. *Journal of Financial Markets*, Vol. 1, 1-50.
- [5] Brockwell, P., and R. Davis. (1996). *Introduction to Time Series and Forecasting*. Springer, New York.
- [6] Chatterjee, K. and W. Samueleson. (1983). Bargaining under Incomplete Information. *Operations Research*, Vol. 31, 835-851.
- [7] Choi, J. Y., S. Salandro, and K. Shastri. (1988). On the Estimation of Bid-Ask Spreads: Theory and Evidence. *The Journal of Financial and Quantitative Analysis*, Vol. 23, 219 -230.
- [8] Cohen, K. J., S. F. Maier, R. A. Schwartz and D. K. Whitcomb. (1981). Transaction Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread. *The Journal of Political Economy*, Vol. 89, 287-305.
- [9] Cont, R., S. Stoikov, R. Talreja. (2008). A stochastic model for order book dynamics. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1273160#1136118](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1273160#1136118)
- [10] Copeland, T.E. and D. Galai. (1983). Information Effects on the Bid-Ask Spread. *The Journal of Finance*, Vol. 38, 1457-1469.
- [11] Demsetz, H. (1968). The Cost of Transacting. *The Quarterly Journal of Economics*, Vol. 82, 33-53.
- [12] Domowitz, I. and J. Wang. (1994). Auctions as Algorithms: Computerized Trade Execution and Price Discovery. *Journal of Economic Dynamics and Control*, Vol. 18, 29-60.

- [13] Foucault, T. (1999). Order Flow Composition and Trading Costs in a Dynamic Limit Order Market. *Journal of Financial Markets*, Vol. 2, 99-134.
- [14] Garman, M. B. (1976). Market Microstructure. *Journal of Financial Economics*, Vol. 3, 257-275.
- [15] Glosten, L. R. and L. E. Harris. (1988). Estimating the Components of the Bid-Ask Spread. *Journal of Financial Economics*, Vol. 21, 123-142.
- [16] Glosten, L. R. and P. R. Milgrom. (1985). Bid and Ask Prices in Specialist Markets in Heterogeneously Informed Traders. *Journal of Financial Economics*, Vol. 14, 71-100.
- [17] Grossman, S. J. (1976). On the efficiency of competitive stock markets when traders have diverse information. *Journal of Finance* Vol. 31, 573-85.
- [18] Grossman, S. J. and J. E. Stiglitz. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, Vol. 70, 393-408.
- [19] Harsanyi, J. (1967-68). Games with Incomplete Information Played by Bayesian Players, Parts I, II, and III. *Management Science*, Vol. 14, 159-182, 320-334, 486-502.
- [20] Harris, L. (1998). Optimal dynamic order submission strategies in some stylized trading problems. *Financial Markets, Institutions and Instruments*, Vol. 7, 1-76.
- [21] Harris, L. (2003). *Trading and Exchanges: Market Microstructure for Practitioners*, Oxford University Press, NY.
- [22] Hasbrouck, J. (2007). *Empirical Market Microstructure*. Oxford University Press.
- [23] Ho, T. and R. Macris. (1984). Dealer Bid-Ask Quotes and Transaction Prices: An Empirical Study of Some AMEX Options. *The Journal of Finance*, Vol. 39, 23-45.
- [24] Ho, T., R. Schwartz, and D. Whitcomb. (1985). The trading Decision and Market Clearing under Transaction Price Uncertainty. *Journal of Finance*, Vol. 40, 21-42.
- [25] Ho, T. and R. Stoll. (1980). On Dealer Markets Under Competition. *The Journal of Finance*, Vol. 35, 259-267.
- [26] Ho, T. and R. Stoll. (1981). Optimal Dealer Pricing under Transactions and Returns Uncertainty. *Journal of Financial Economics*, Vol. 9, 47-73.
- [27] Ho, T. and R. Stoll. (1983). The Dynamics of Dealer Markets Under Competition. *The Journal of Finance*, Vol. 38, 1053-1074.
- [28] Kleinrock, L. (1967). Optimum Bribing for Queue Position. *Operations Research*, Vol. 15, 304-318.

- [29] Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica*, Vol. 53, 1315-1335.
- [30] Kyle, A. S. (1989). Informed Speculation with Imperfect Competition. *The Review of Economic Studies*, Vol. 56, 317-355.
- [31] Luckock, H. (2003). A steady-state model of the continuous double auction, *Quantitative Finance*, Vol. 3, 385-404.
- [32] Madhavan, A. and V. Panchapagesan. (2000). Price discovery in auction markets: a look inside the black box. *Review of Financial Studies*, Vol. 13, 627-658.
- [33] Madhavan, A., M. Richardson and M. Roomans. (1997) Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks. *The Review of Financial Studies*, Vol. 10, 1035-1064.
- [34] Mendelson, H. (1982). Market Behavior in a Clearing House. *Econometrica*, Vol. 50, 1505-1524.
- [35] Nicholson, W. (1998). *Microeconomic Theory: Basic Principles and Extensions*, Dryden press, NY.
- [36] Obizhaeva, A., and Wang, J. (2006). Optimal trading strategy and supply demand dynamics. Working Paper, Sloan School, MIT. (<http://www.nber.org/papers/w11444>)
- [37] O'hara, M. and G. S. Oldfield. (1986). The Microeconomics of Market Making. *The Journal of Financial and Quantitative Analysis*, Vol. 21, 361-376.
- [38] Parlour, C. A. (1998). Price dynamics In Limit Order markets. *Review of Financial Studies*, Vol. 11, 781-816.
- [39] Phipps, T. E. (1956). Machine Repair as a Priority Waiting-Line Problem. *Operations research*, Vol. 4, 76-85.
- [40] Pratt, J. W. (1964). Risk Aversion in the Small and in the Large, *Econometrica*, 122-136.
- [41] Roll, R. (1984). A simple measure of the effective bid/ask spread in an efficient market, *Journal of Finance*, Vol. 39, 1127-1139.
- [42] Rosu, I. (2009) Dynamic Model of the Limit Order book. *Review of Financial Studies*, Vol. 22, 4601-4641.
- [43] Rustichini, A., M. A. Satterthwaite, and S. R. Williams. (1994). Convergence to Efficiency in a Simple Market with Incomplete *Econometrica*, Vol. 62, 1041-1063.
- [44] Satterthwaite, M. A., and S. R. Williams. (1989). Bilateral Trade with the Sealed Bid k-Double Auction: Existence and Efficiency. *Journal of Economic Theory*, 48, 107-133.

- [45] Seppi, D. J. (1997). Liquidity provision with limit orders and a strategic specialist. *Rev Financial Studies*, Vol. 10, 103-150.
- [46] Shiller, R. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *The American Economic Review*, Vol. 71, 421-436.
- [47] Smith, E., J. D. Farmer, G. Lszl and S. Krishnamurthy. (2003). Statistical theory of the continuous double auction, *Quantitative Finance*, Vol. 3, 481 514.
- [48] Stoll, R. (1978). The Supply of Dealer Services in Securities Markets. *The Journal of Finance*, Vol. 33, 1133-1151.
- [49] Tijms, H. C. (1994). *Stochastic Models: An Algorithmic Approach*, John Wiley and Sons, NY.
- [50] White, H. and L. S. Christie. (1954). Queues with Pre-emptive Priorities or with Breakdowns. *Operations Research*, Vol. 2, 341-434.
- [51] Wilson, R. (1991). Strategic Analysis of Auctions. *The Handbook of Game Theory*, Ed. Aumann, Hart, Elsevier/North-Holland, Amsterdam.
- [52] Wohl,A., S. Kandel (1995). Implications of an Index-contingent Trading Mechanism. *Journal of Business*, Vol. 70, 471-488.