

Yet Another Trading Simulation: The Nonimmediacy Model

Nagaratnam Jayasreedharan, School of Economics and Finance, University of Tasmania, Private Bag 85, Hobart, TAS 7001
e-mail: nj.sreedharan@utas.edu.au

Abstract

Market participants who want to trade “quickly” demand transactional liquidity. Market participants who demand to trade “immediately” demand transactional immediacy. Traders supply immediacy when the trading returns are high and demand immediacy when the trading returns are low. Traders supply immediacy by being non-immediate and traders demand immediacy by being immediate. As these time-based expectations are functions of the underlying distributions of the clearing prices, one is able to define exact analytical functions for time-differing non-immediacy traders. As non-immediacy is heterogenous, a range of non-immediacy traders with differing non-immediacies can compete in the marketplace and give rise to a “zone” of limit order functions. The bid-ask expectations are also assumed to be rank-dependent to maintain transitivity and monotonicity. The limit orders placed by non-immediacy traders can then be “picked” off by immediacy traders as they arrive at the exchange. A simulation of a simple trading model is undertaken to illustrate the mechanism of price formation under rank-dependent non-immediacy trading. In doing so, we show that immediacy is a consequence of non-immediacy trading and is provided freely in any organised asset markets when participants transmit their demand propensities to the market in the form of limit orders. While certain participants such as dealers and specialists do indeed (at a price) provide immediacy, they need not be the only providers.

Key Words: Immediacy; Nonimmediacy trading; Rank-dependent.

JEL Codes: D–Microeconomics; D01-Microeconomic Behaviour: Underlying Principles; D53-Financial Markets; D84-Expectations; Speculations.

1 Introduction

Market participants adopt “*more than one time horizon in the course of their lives, but they may also have multiple time horizons at a given moment*” (see Dembo and Freeman (1998)). They might have some assets to which they attribute to a very long horizon and other assets which they attribute a greater *immediacy*. Alternatively, market participants who want to trade “quickly” demand transactional liquidity.¹ Market participants supply immediacy when the trading returns are high and demand immediacy when the trading returns are low. Market participants supply immediacy by being non-immediate and traders demand immediacy by being immediate.

Patient traders obtain superior prices to impatient traders because they are willing to search longer and harder to arrange their trades at favourable terms. Impatient traders pay for the privilege of trading when they want to trade (see Harris (2002)). Handa, Schwartz and Tiwari (1998) conclude that “patient traders are rewarded for placing limit orders, and that eager traders are best advised to place market orders”. Schwartz (2000) is of the opinion that when it comes to trading “*the critical difference may actually be between participants who wish to trade with immediacy and those who are willing to trade patiently [with non-immediacy]*”.

A trade is achieved at any time during the trading day when two counterpart orders cross². Counterpart orders cross (see Schwartz and Whitcomb (1988)) when:

- (i) a public trader submits a limit order, and another public trader submits a market order which executes against the limit order;
- (ii) an exchange specialist or over-the-counter market maker sets the quote and public market order executes against the specialist’s or dealer’s quote; or
- (iii) two or more public traders negotiate a trade. The negotiation may take place on the floor of the exchange, or in the upstairs market.

¹ Liquidity has many dimensions one of which is “immediacy”.

² Say one limit order and one market order. Two limit orders cannot cross by definition.

In the first two cases, one party to a trade can be viewed as an *immediacy* trader, and the other party can be viewed as a *non-immediacy* trader.³

The participant who is seeking to trade without delay is an immediacy trader; the immediacy trader generally incurs an execution cost. Immediacy traders are the public market orders (items (i) and (ii) above) and the traders who initiate the negotiation process (item (iii)). Immediacy trading involves the placement of market orders, active searching for the contra party or wanting to trade. The more anxious an investor is to acquire or to dispose of shares at current market prices, the larger the price concession he or she would make to trade immediately. Transaction certainty can be obtained by placing market orders. Non-immediacy trading involves the placement of limit orders and passive waiting for the arrival of a counter party to the trade. Non-immediacy traders are limit order traders (item (i)), the specialist or dealer (item (ii)), and the trader who does not initiate the negotiation process (item (iii)).

All limit orders implicitly have a lower probability of execution relative to market orders. These probabilities are determined by the volatility of the stock being traded and the investor's trading horizon. Further, if one assumes heterogenous trading horizons and rank-dependent probabilities, it is possible for the different limit orders to be indexed and a series of indifference curves to be identified.

Immediacy, the counter-party to non-immediacy, is a necessary by-product of trading and is provided freely in any organised asset markets when participants transmit their demand propensities to the market in the form of limit orders (see Cohen, Maier, Schwartz and Whitcomb (1981)). Hence, while certain participants such as dealers and specialists do indeed (at a price) provide immediacy, they are not the only providers.

In this paper we develop a model of non-immediacy trading and illustrate some of the distributional characteristics via a simple simulation. Our model is in line with

³ Schwartz and Whitcomb [1988] use the terms "active" and "passive" traders, but in this paper the terms "immediacy" and "non-immediacy" traders are used to stress the time-preference implicit in active and passive trading.

the observed fact, in common with Cohen, Maier, Schwartz and Whitcomb (1981), that “a market is composed of many ordinary traders who may periodically transmit either limit or market orders (or else nothing) depending on the resolution of their own optimal order placement strategies”. Consequently, the limit orders of some traders can be said to explicitly establish the market for a security and provide immediacy to other traders who trade via market orders.

The rest of the paper is as follows. In Section 2, buy/sell and bid/ask expectations, immediacy and non-immediacy are defined and discussed. Section 3 discusses equilibrium and non-equilibrium pricing in the context of immediacy and non-immediacy. Immediacy and non-immediacy curves are defined and illustrated in Section 4 followed by a simulated immediacy and non-immediacy trading in Section 5. In Section 6, conclusions with possible implications for further research are presented. Section 7 is the appendix where the various order types are explained and is followed by the reference list in Section 8.

2 Buy/Sell and Bid/Ask Expectations

There are in each act of exchange, two mutually exclusive decisions being made by each participant, one to buy and the other to sell. For the buyer, the estimate of the expectation of the change in the price for the stock is positive, while the seller’s estimate of the same quantity is negative. However, the bid prices represent prices at which a buyer is willing to acquire a stock (and pay cash) and the ask positions represent prices at which a seller is willing to dispose of a stock, and hence receive cash. Consequently, the expectations of outcomes when investing and when trading are mutually opposite to each other.

Let $\{x_t\}$ denote a sequence of random price changes and let $\Omega_t = \{x_t, x_{t-1}, \dots\}$ denote a set of conditioning information or information set based on the past history of x_t . In addition, let $E[\cdot]$ be the expectation operator and $E_\eta[\cdot]$ be the conditional expectation operator dependent on the parameter, η , where $1 \leq \eta \leq \infty$. The buy and sell expectations are then defined as (see Jeyasreedharan (2004)):

$$(1.1) \quad E^{buy} [x_t | \Omega_{t-1}] \geq 0$$

$$(1.2) \quad E^{sell} [x_t | \Omega_{t-1}] \leq 0$$

where E^{buy} is the buy-side expectation of the *investment* price changes, E^{sell} is the sell-side expectation of the *investment* price changes and Ω_{t-1} is the information set at time $\{t-1\}$.

However, the quoted prices represent decisions at which a buyer is willing to acquire a stock (and sell money) and a seller is willing to dispose of a stock, and hence receive money. The corresponding bid-ask expectations (or quote prices) are not equal to but are *opposite* to the ex-ante buy-sell expectations and are defined as:

$$(1.3) \quad E_{\eta \geq 1}^{bid} [x_t | \Omega_{t-1}] \leq 0$$

$$(1.4) \quad E_{\eta \geq 1}^{ask} [x_t | \Omega_{t-1}] \geq 0$$

where $E_{\eta \geq 1}^{bid}$ is the bid-expectation of the *trading* outcomes and $E_{\eta \geq 1}^{ask}$ is the ask-expectation of the *trading* outcomes conditional on the non-immediacy parameter $1 \leq \eta \leq \infty$.

The trader, knowing there may be *eager* buyers and sellers, will try to maximise the trading-utility (gains) by obtaining the best price possible away from the current market price. This maximisation means getting *higher* prices for ask-quotes and *lower* prices for bid-quotes. This is achieved by issuing limit orders and waiting for the transaction to occur. Implicitly this means increasing the weighting of low probability outcomes and is equivalent to the maximisation of probabilities conditional on waiting or non-immediacy times and the decision to buy or sell. These conditional maximisations must be rank-dependent to ensure transitivity and monotonicity in the order functions. Consequently, the conditional bid-ask expectations can be defined more explicitly as:

$$(1.5) \quad E_{\eta \geq 1}^{bid} [x_t | \Omega_{t-1}] = \int_X u(x_{t-1}) \bullet d(g^{bid} \circ F)(x_{t-1}) | \Omega_{t-1} \leq 0$$

$$(1.6) \quad E_{\eta \geq 1}^{ask} [x_t | \Omega_{t-1}] = \int_X u(x_{t-1}) \bullet d(g^{ask} \circ F)(x_{t-1}) | \Omega_{t-1} \geq 0$$

where $g^{bid} \circ F = 1 - [1 - F]^\eta$ and $g^{ask} \circ F = F^\eta$ and $1 \leq \eta \leq \infty$, is the non-immediacy parameter. The transformation functions $(g^{bid} \circ F) = 1 - [1 - F]^\eta$ and $(g^{ask} \circ F) = F^\eta$ are rank-dependent by definition (see Quiggin (1982, Quiggin (1985, Quiggin (1993) for transitivity, monotonicity and rank-dependency). These bid and ask expectations are analogous to the “indexed”⁴ limit orders of Black (1995), meaning an inherent preference ordering is maintained. The bid-ask rank-dependent price changes can then be determined as follows:

$$(1.7) \quad \begin{aligned} F^{-1}(F^\eta(x_{t-1}^c)) &= x_t^a \\ F^{-1}(1 - [1 - F(x_{t-1}^c)]^\eta) &= x_t^b \\ \text{and } x_t^a &\geq x_{t-1}^c \geq x_t^b \end{aligned}$$

where F^η is the rank-dependent transformation, $F^{-1}(\cdot)$ is the inverse of the cumulative distribution function, $F(\cdot)$, x_{t-1}^c is the most recent relative change⁵ in the clearing price, x_t^a is the relative change in the ask price expected and x_t^b is the relative change in the bid price expected and η is the non-immediacy preference parameter. When traders are indifferent to their time-to-execution, they can be assumed to trade via market orders with $\eta = 1$ (immediacy=non-immediacy) and consequently:

$$(1.8) \quad \begin{aligned} F^{-1}(F^{\eta=1}(x_{t-1}^c)) &= x_t^a \\ F^{-1}(1 - [1 - F(x_{t-1}^c)]^{\eta=1}) &= x_t^b \\ \text{and } x_t^a &= x_{t-1}^c = x_t^b \end{aligned}$$

⁴ The term indexed limit orders was defined and used by Black [1995].

⁵ Relative change means any change relative to the previous period's price.

where the bid and ask price changes and the clearing price changes are equal.⁶

The following relationships are then inferred from the preceding arguments:

$$(1.9) \quad \begin{aligned} & F^{buy/sell} [x_t | \Omega_{t-1}] \neq F_{\eta}^{bid/ask} [x_t | \Omega_{t-1}] \\ & \neq F \{ x_t | F_{\eta}^{bid} [x_t | \Omega_{t-1}], F_{\eta}^{ask} [x_t | \Omega_{t-1}] \} \neq F[x_t | \Omega_{t-1}] \end{aligned}$$

where $F^{buy/sell} [x_t | \Omega_{t-1}]$ are the cumulative distributions conditioned on beliefs and information and $F_{\eta}^{bid/ask} [x_t | \Omega_{t-1}]$ are the cumulative distributions conditioned on preferences and information. Beliefs are reflected in the buy-sell choices conditioned on historical price information. Immediacy/non-immediacy preferences are reflected in the limit orders conditioned on historical price information.

In the light of the discussion above, a new classification of traders is warranted: *immediate* and *non-immediate* traders. Immediate traders demand immediacy. Non-immediate traders (dealers, specialists and ordinary traders) provide immediacy. Immediate traders place market orders and non-immediate traders place limit orders. This unique approach allows all market participants to be simply classified and grouped into demanders and providers of immediacy.⁷

2.1 Immediacy

Keynes (1936) defined *animal spirits* as “a spontaneous urge to action rather than inaction, and not the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities”. This spontaneous urge for action rather than

⁶ This equality is characteristic of rational expectation models where traders are assumed to be indifferent to their time-to-execution.

⁷ This approach differs from Cohen, Maier, et al. [1981], who separated the demanders from the providers of immediacy by using groups of ordinary traders and market makers.

inaction is analogous to the concept of *immediacy*⁸. Handa and Schwartz (1996) define immediacy as *immediate liquidity*, and not liquidity per se.

Immediacy traders use market orders to trade. For a buy-side market order this can be expressed as:

$$(1.10) \quad E_{\eta \leq 1}^{buy}[x_t | \Omega_{t-1}] \geq E_{\eta > 1}^{ask}[x_t | \Omega_{t-1}]$$

and for a sell-side market order and

$$(1.11) \quad E_{\eta > 1}^{bid}[x_t | \Omega_{t-1}] \leq E_{\eta \leq 1}^{sell}[x_t | \Omega_{t-1}]$$

where x_t is the change in price at time t , $\{E_{\eta \leq 1}^{buy}[x_t | \Omega_{t-1}]; E_{\eta > 1}^{bid}[x_t | \Omega_{t-1}]\}$ are the buy/bid-expectations operators and $\{E_{\eta \leq 1}^{sell}[x_t | \Omega_{t-1}]; E_{\eta > 1}^{ask}[x_t | \Omega_{t-1}]\}$ are the sell/ask-expectations operators both conditional on the common information set Ω_{t-1} with the non-immediacy parameter, $0 < \eta \leq 1$ ⁹, for the buy-sell expectations and with the non-immediacy parameter, $1 < \eta < \infty$, for the bid-ask expectations. Immediacy traders are generally willing to pay the prices expected by non-immediacy traders and often trade at inferior prices¹⁰.

2.2 Non-immediacy

Non-immediacy¹¹ is defined here as “a non-spontaneous urge to inaction rather than action, and the outcome of a rank-dependent weighted average of quantitative

⁸ The literal meaning of immediacy is “short latency of response; promptness; spending more time with another; spending time with another when they choose”.

⁹ When $0 < \eta \leq 1$, the ensuring PDFs opposite to the bid-ask expectations.

¹⁰ Inferior prices are buy orders that trade at high prices and sell orders that trade at low prices. See Harris [2002] for details.

¹¹ The literal meaning of non-immediacy is “long latency of response; delinquent about being on time; spending little time with another; often glances at watch/clock”.

benefits multiplied by quantitative (subjective) probabilities”. A rational trader, who would generally be less animal-like in nature and hence less spontaneous, will be expected to take into careful consideration all possibilities and probabilities of outcomes when faced with a choice.

An investor who makes only necessary portfolio adjustments or who does not react to news would be more willing to risk missing a particular trade and, accordingly, such an investor can be deemed to be a non-immediate trader. Furthermore, a non-immediate seller ideally will expect the highest possible price. The highest possible price, however, is bounded by the trading horizon or the maximum time-to-execution of the trader. The highest possible price over one day is quite different from that possible over one month. The highest possible price is also bounded by the probability of the clearing prices. The distribution of the quotes will be such that all of the rank-dependent quotes will at least have some finite likelihood of execution. Non-immediacy traders place limit orders and are willing to wait with $\eta > 1$. This can be expressed as:

$$(1.12) \quad E_{\eta > 1}^{bid}[x_t | \Omega_{t-1}] < E_{\eta > 1}^{ask}[x_t | \Omega_{t-1}];$$

where $E_{\eta > 1}^{bid}[x_t | \Omega_{t-1}]$ is the bid-expectations operator and $E_{\eta > 1}^{ask}[x_t | \Omega_{t-1}]$ is the ask-expectations operator conditional on the common information set Ω_{t-1} with the non-immediacy parameter, $\eta > 1$. By waiting, non-immediacy traders often trade at superior prices¹².

¹² Superior prices are buy orders that trade at low prices and sell orders that trade at high prices.

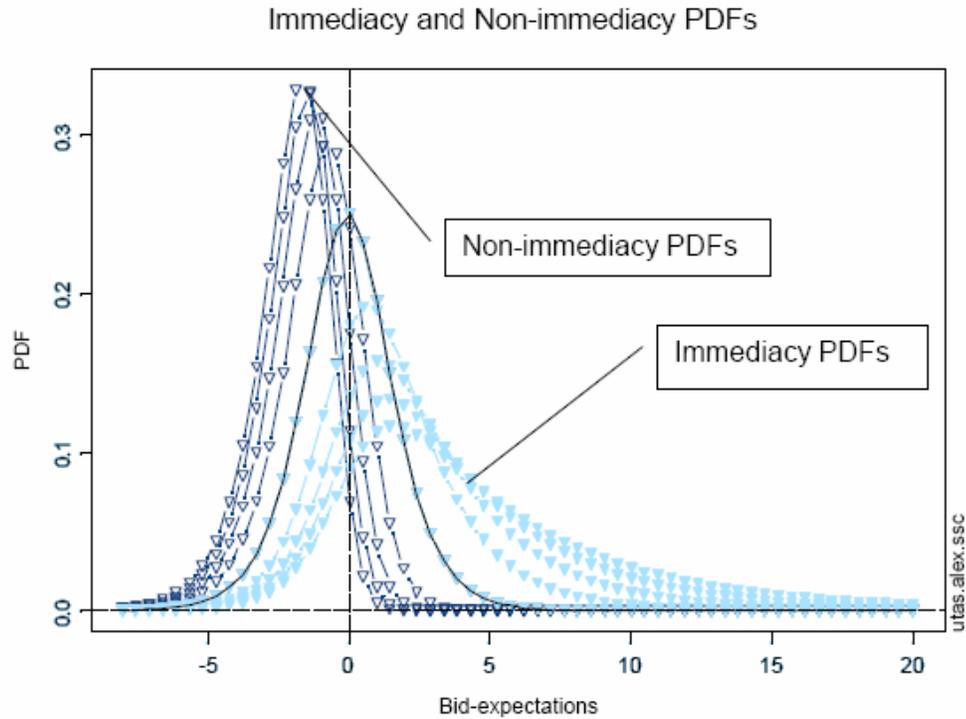


Figure 2-1: Immediacy and Non-immediacy/PDFs

Notes: The buy-side immediacy PDFs are shown with filled down-arrows and the buy-side non-immediacy PDFs are shown with hollowed up-arrows.

In Figure 2-1 the buy-side immediacy and non-immediacy probability density functions (PDFs) are illustrated. The buy-side immediacy PDFs are shown with filled down-arrows and the buy-side non-immediacy PDFs are shown with hollowed down-arrows. Note that the buy-side immediacy traders reveal their beliefs¹³ and preferences, but buy-side non-immediacy traders only reveal their preferences¹⁴ and not their beliefs. Immediacy traders, by issuing market orders, reveal their immediacy preferences and bullish-bearish beliefs. Non-immediacy traders may believe the market to be bearish but, when non-immediate, will quote limit orders are higher than the current clearing prices; thus revealing their non-immediacy

¹³ This does not imply that immediacy traders are irrational. Given their preferences for immediacy, their actions can be deemed rational as long as their trades are at prices within their reservation prices.

¹⁴ Non-immediacy traders stand ready to trade at prices well away from their reservation prices.

preferences but not their bearish beliefs. Trading expectations conditional on non-immediacy preferences, however, can be objectively determined, but trading expectations conditional on immediacy preferences are not measurable. Though the immediacy PDFs are illustrated in Figure 2-1, their actual distributional form or shape would be hard to determine as immediacy traders place market orders and thus there is no way of inferring the order of their immediacies.

2.3 Conditional Immediacy

Stop orders are *conditional unlimited orders* triggered by the market price reaching a predetermined threshold. A stop-loss (stop-gain) order issues an unlimited sell order to the exchange once the asset price falls below (rise above) a certain set price. Stop orders are used as a protection against unwanted losses (when owning a stock) or against unexpected rises (when planning to buy a stock).

Hence a stop order converts to an immediate order once the asset price goes past a certain price. A stop-gain order converts to a market order to sell when the stock price crosses the stop price from below. A stop-loss order converts to a market order to sell when the stock price crosses the stop price from above.

Hence a stop order is given as:

$$(1.13) \quad E_{\eta > 1}^{bid}[x_t | \Omega_{t-1}] \geq E_{\eta \leq 1}^{sell}[x_t | \Omega_{t-1}]; \text{ for } x_t \leq K$$

for a stop-loss order and

$$(1.14) \quad E_{\eta > 1}^{bid}[x_t | \Omega_{t-1}] \geq E_{\eta \leq 1}^{sell}[x_t | \Omega_{t-1}]; \text{ for } x_t \geq K$$

for a stop-gain order where K is the stop price.

A stop-limit order is given as:

$$(1.15) \quad E_{\eta > 1}^{bid}[x_t | \Omega_{t-1}] < E_{\eta > 1}^{ask}[x_t | \Omega_{t-1}]; \text{ for } x_t \leq K$$

for a stop-limit loss order and

$$(1.16) \quad E_{\eta>1}^{bid}[x_t | \Omega_{t-1}] < E_{\eta>1}^{ask}[x_t | \Omega_{t-1}]; \text{ for } x_t \geq K$$

for a stop-limit gain order where K is again the stop price. Analogous conditions can be defined for start orders and start-limit orders on the buy-side.

3 Types of Pricing

3.1 Equilibrium Pricing

Ideally, an exchange market is assumed to bring together information on listed stocks in a manner that, in principle, “makes for perfectly informed behaviour by both buyers and sellers” (see Else and Curwen (1990)). A single offer to buy or sell a stock is assumed not to affect the price as set by the market. Further, the buying and selling is also assumed to be instantaneous.

Under these conditions, the demand and supply curves for a stock can be deemed to intersect at an equilibrium price as shown in Figure 3-1. The demand curve is annotated as $D/n=1$ which means a “demand curve with an immediacy of $n = 1$ ”. The supply curve is annotated as $S/n=1$ which means a “supply curve with an immediacy of $n = 1$ ”.

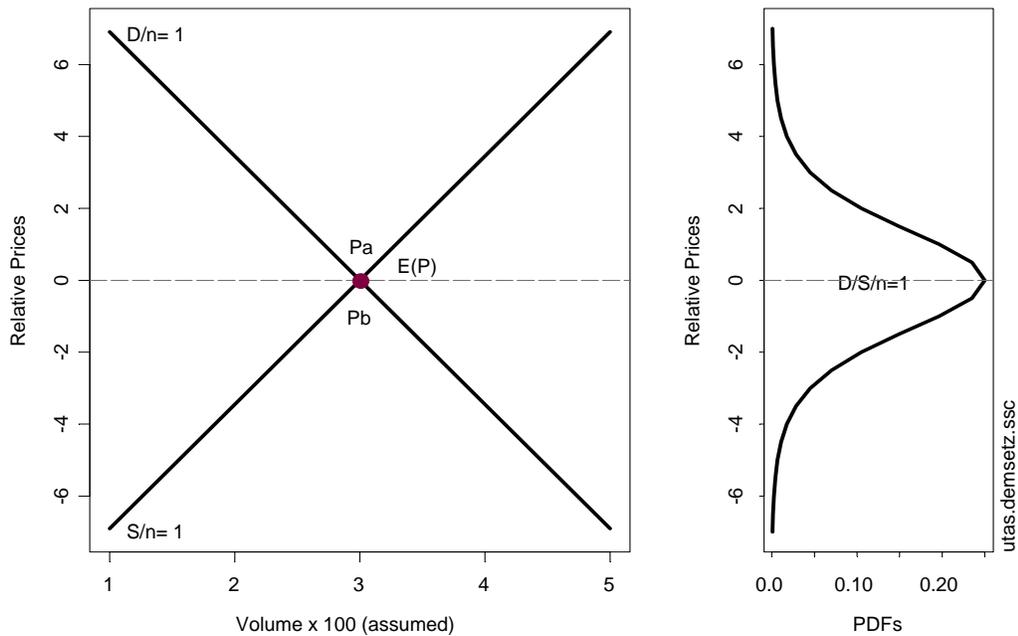


Figure 3-1: Demand and Supply Curves/PDFs

Notes: D=demand curve; S=Supply curve; n=immediacy=1; Pa=ask price; Pb=bid price; E(P)=equilibrium or expected price.

Further, as shown in the left-panel of Figure 3-1, the equilibrium price, $E(P)$, the bid price, P_b , and the ask price, P_a , are the same at any one instant¹⁵. The right-panel depicts the distribution of the equilibrium prices over a period of time. Note that the distribution of the equilibrium prices imply that the trade and quote prices are identical and symmetric (right-panel in Figure 3-1 has only one PDF depicted). As the distribution of the bid-quotes and ask-quotes are also assumed to be identical, the bid-ask spread is zero.

3.2 Immediacy Pricing

While, over time, the number of sellers might equal the number of buyers, at any particular point in time such an outcome is not guaranteed. Traders, in practice,

¹⁵ The equilibrium price is the price at which the aggregate buy-quote order function equals the aggregate sell-quote order function.

transmit their orders sporadically during trading sessions. Demsetz (1968) examined the behaviour of prices under such conditions and considered the possibility of two equilibrium points based on the immediacy of trades.

On the demand side, there is one demand arising from agents who want to buy immediately, and another demand representing the traders who want to buy but do not feel the need to do so immediately. The supply side is defined analogously. If there is an imbalance of traders wanting to buy now, then either some buyers have to wait for sellers to arrive, or they can offer a higher price to induce waiting sellers to transact now.

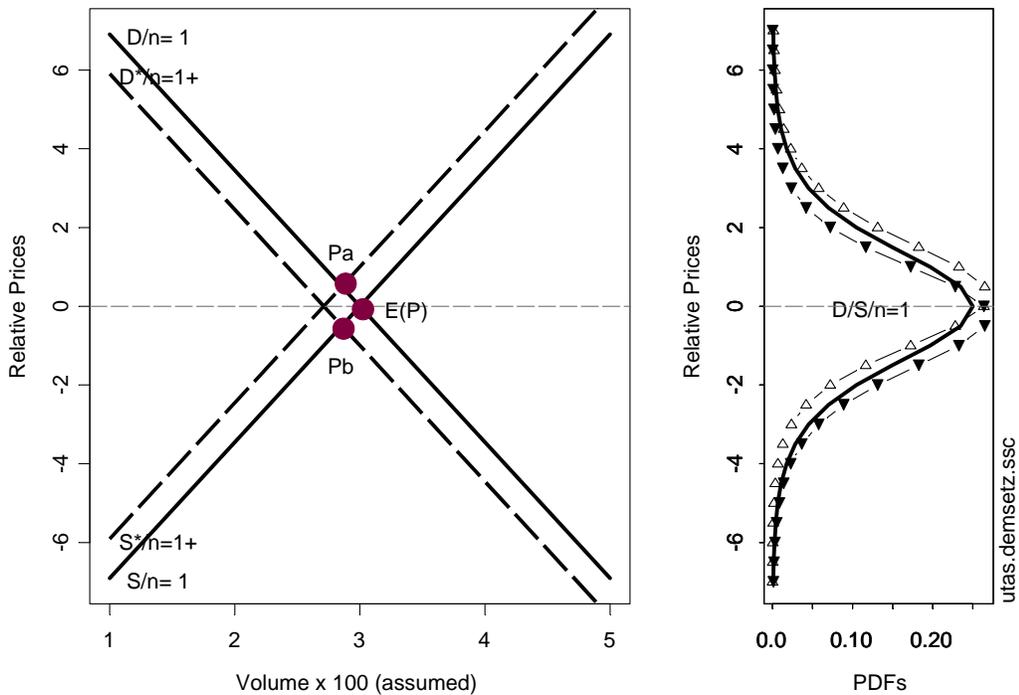


Figure 3-2: Immediacy Costs/Inside Spread ($P_a - P_b$)

Notes: Immediacy costs refer to inventory costs, waiting costs, commissions, taxes, communication expenses, and so forth etc. D/n =Demand/immediacy; D^*/n =Supply/immediacy.

Similarly, if there is an imbalance of sellers wanting to trade now, a lower price must be bid to induce more buyers to trade now (see Demsetz (1968)). At any point in time there are two possible representations of supply and demand in the market. As

shown in Figure 3-2, the two sources of supply and demand results in two, not one, equilibrium prices. These two supply curves and demand curves reflect the two time frames of the trading process; trading now or trading later. The “inside” spread reflects the immediacy costs and is defined as $\{Pa - Pb\}$.

3.3 Non-immediacy Pricing

Trading, however, involves two types of uncertainties (see Schwartz and Whitcomb (1988)):

- (i) Price uncertainty: the opportunity cost of trading at inferior prices (buying at prices that are abnormally high or selling at prices that are abnormally low); and
- (ii) Transaction uncertainty: the opportunity cost of missing a trade because a buy order was priced too low, or because a sell order was priced too high.

In the presence of price and transaction uncertainty, the clearing price will differ from the equilibrium price thus impairing the informational accuracy of the market price. In a call market, where all orders are batched together, a desired clearing price could be achieved under the following conditions (see Schwartz and Whitcomb (1988)):

- (i) investors have *unbiased* expectations concerning the clearing price, and
- (ii) the order functions of buyers and sellers are distributed *symmetrically* around the expected clearing price.

If expectations, however, are biased and if the intensity with which some investors wish to buy is not exactly matched by the intensity with which other investors wish to sell, the order functions of buyers and sellers will be distributed asymmetrically around the expected clearing price. Consequently, the market clearing price will be perturbed from the expected clearing price (see Ho, Schwartz and Whitcomb (1985)).

If the prices established in the call market can be perturbed by individuals reacting to the transaction price uncertainty, so too are the prices established in the continuous

market where trades are made between individual pairs of buyers and sellers, as orders arrive sequentially.

A trader can trade strategically by “waiting” or being patient and place a limit order to obtain better prices. The trader who submits a market order or is immediate must take into account both the immediacy and non-immediacy costs. Demsetz (1968)‘s “immediacy” argument is complemented in this paper with the concept of “non-immediacy”. The non-immediate demand/supply curves (for $\eta = 5$) are shown in Figure 3-3 as the light dashed lines $D'/\eta = 5$ and $S'/\eta = 5$ passing through the points P_A and P_B (the ask and bid-prices of non-immediacy).

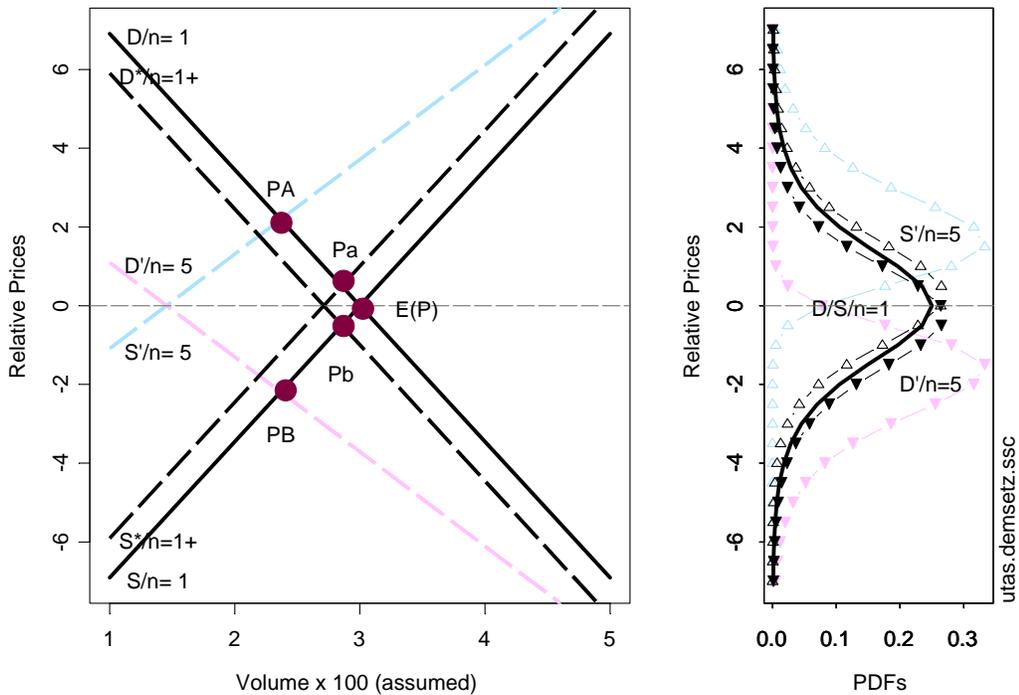


Figure 3-3: Non-immediacy Demand/Supply Curves/PDFs

Notes: PA =price of non-immediacy when D'/n =Demand/non-immediacy=5; PB =price of non-immediacy when S'/n =Supply/non-immediacy=5; All other symbols as explained under Figure 3-1 Notes. PA minus PB is the outside spread.

The “outside” spread reflects the non-immediacy costs and is defined as $\{PA - PB\}$ and is shown in Figure 3-3. The PDFs on the right panel in Figure 3-3 illustrates the

distinct separation of the bid and ask distributions. As market participants use trading strategies based on non-immediacy pricing when they seek to trade, the “inside” spread is not the only spread traders need to price into their order functions.

Now, not all traders have the same non-immediacy preferences. Some traders may be willing to wait longer than others. Each trader’s non-immediacy preference is subjectively determined and as such there will be a range of non-immediacy demand/supply curves implicit in the order functions.

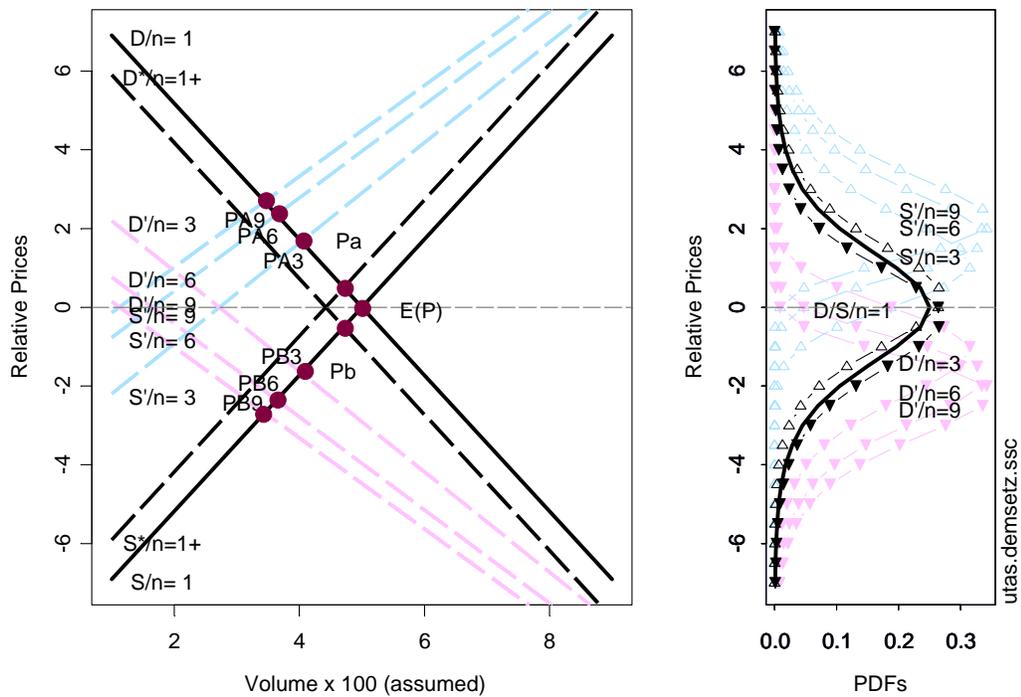


Figure 3-4: Multiple Non-Immediacy Demand/Supply Curves/PDFs

Notes: PA3, PA6 and PA9 are the clearing prices for S''/n =Supply/non-immediacy=3,6 and 9. PB3, PB6 and PB9 are the clearing prices for D''/n =Demand/non-immediacy=3,6 and 9. The legend is the same for the PDF plot. All other symbols are as indicated under Figure 3-1 Notes.

In Figure 3-4 a range of demand/supply curves for varying non-immediacy values of η is depicted. The longer a trader is willing to wait, the higher will be the ask-prices and lower the bid-prices and the demand/supply curves shift to the left along the x-

axis in the left-panel. Correspondingly, based on rank-dependent transformations, the PDFs shift outwards to the top and bottom along the y-axis in the right-panel.

The clearing price is the price at which individual quotes match counterparty quotes whereas the equilibrium price is the price at which the aggregate buy-quote order function equals the aggregate sell-quote order function. Clearing or transaction prices (as the result of non-synchronous trading) can be different from equilibrium prices. Consequently, the market clearing prices are not necessarily the equilibrium prices.

4 Immediacy and Non-immediacy Curves

If traders are not indifferent to their time-to-execution then one needs a way of depicting their time-preferences. One way of doing this is via “quantile-quantile” plots.

A quantile-quantile (QQ) plot is a graphical technique that can be used to determine whether two data sets come from populations with a common distribution. A QQ-plot is a plot of the quantiles of the first data set against the quantiles of the second data set. A 45-degree baseline is generally depicted in the plots. If the two sets come from a population with the same distribution, the points should fall approximately along this 45-degree baseline. The greater the departure from this baseline, the greater the evidence for the conclusion that the two data sets originate from populations with different distributions. A QQ-plot of quote log-returns versus trade log-returns (with the trades on the x-axis and the quotes on the y-axis) results in the data points forming a curved line placed well away from the 45-degree baseline. The curved line depicts the relative change in quote prices expected by traders for a specific non-immediacy (time-preference) given the current relative change in trade prices. The shift in relative changes can also be depicted as a PDF of relative changes. These formulations can be illustrated using QQ-plots.

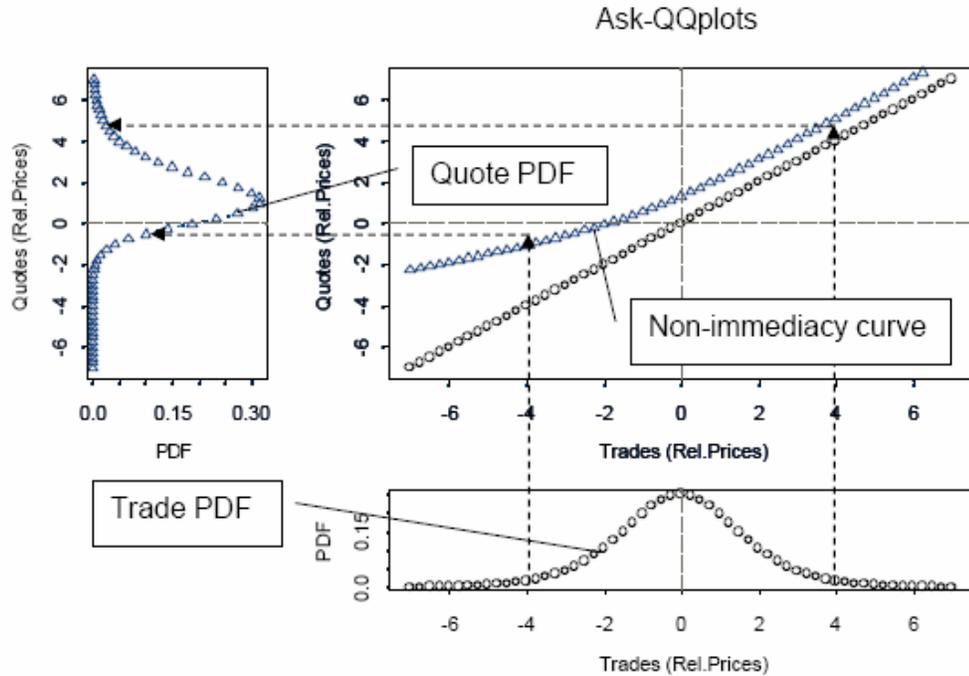


Figure 4-1: QQ-plots and PDFs

Notes: The bottom-panel depicts the PDF of the trade price relative changes. In the middle-panel, the “concave” curve above the diagonal line is the sell (ask) non-immediacy curve with $n=3$. The left-panel depicts the PDF for non-immediacy quotes with $n=3$.

Figure 4-1 shows how the distribution of the relative changes in the trade prices (in the bottom panel) is translated into a distribution for the relative changes in the ask-quote prices for $\eta=3$ (in the left panel). Note that the quote distribution in the left-panel is “fat” upper tail and a “thin” lower tail. This asymmetrical distortion in the probabilities is due to the nonlinear transformation of the probabilities using rank-dependency as depicted by the non-linear non-immediacy curve in Figure 4-1.

The i -th quantile for the ask-quote price is simply the appropriate rank-dependent quantile function evaluated at the i -th quantile for the trade price. The non-linearity effect depends on the order of non-immediacy, η , of the trader and the range of trade prices over the trading horizon.

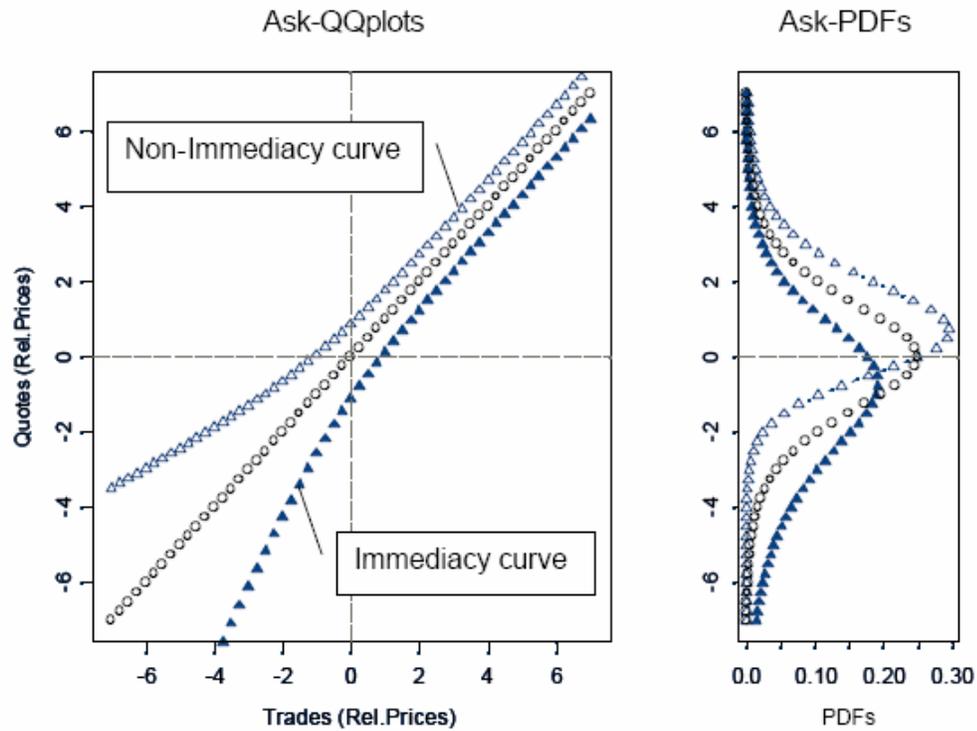


Figure 4-2: Ask-QQ-plots

Notes: In the left-panel, the “concave” curve above the diagonal line is the sell non-immediacy line with $n=2$ and the “convex” curve below the diagonal line is the sell non-immediacy line with $n=1/2$.

In the left-panel of Figure 4-2 are two QQ-plot curves; a non-immediacy curve and an immediacy curve. The seller’s (ask) non-immediacy curve is depicted by hollowed up-arrows and the seller’s (ask) immediacy curve is depicted by filled up-arrows. A seller who is non-immediate will quote or only accept a price higher than the last clearing price. A seller who is immediate will quote or accept a price lower than or equal to the last clearing price and will have a higher likelihood of execution. The only difference between these two curves is their degree of non-immediacy. The non-immediacy curve in Figure 4-2 is for $\eta = 2$ and the immediacy curve is for $\eta = 1/2$.

The 45-degree diagonally dotted line depicts the quote and trades prices when $n=1$ and when traders are immediacy/non-immediacy indifferent. The right-panel traces the PDFs for the corresponding QQ-plot curves and the 45-degree diagonal line as done in Figure 4-1.

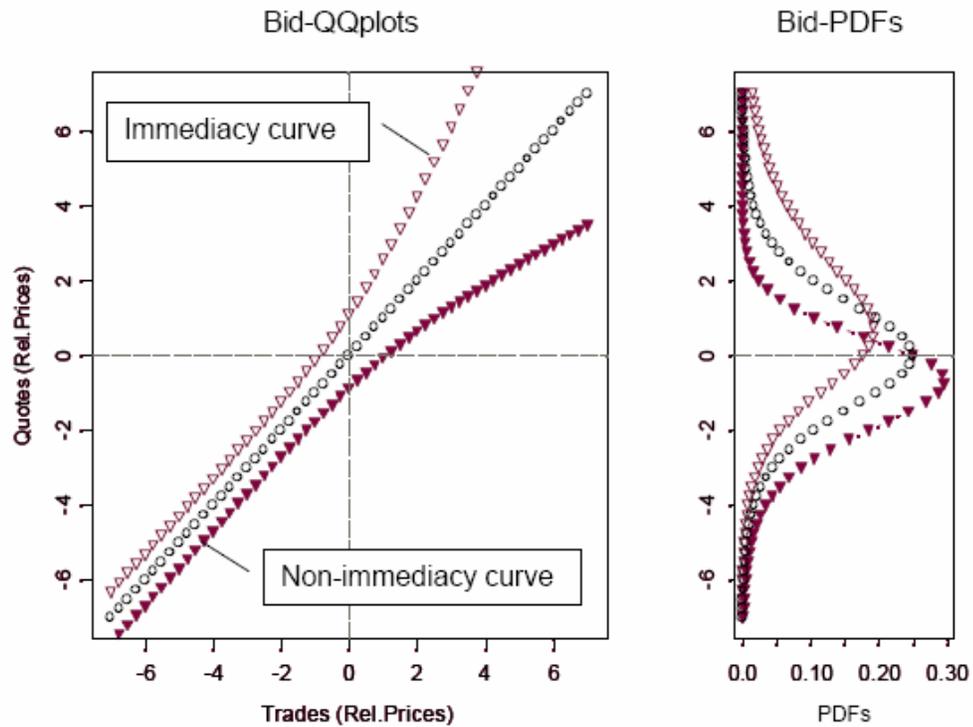


Figure 4-3: Bid-QQ-plots

Notes: In the left-panel, the “concave” curve above the diagonal line is the buy non-immediacy line with $n=1/2$ and the “convex” curve below the diagonal line is the buy non-immediacy line with $n=2$.

In Figure 4-3 the buy-side non-immediacy and immediacy curves are illustrated. The buy-non-immediacy curve is depicted by filled down-arrows and the buy-immediacy curve is depicted by hollowed down-arrows. An immediacy indifferent buyer with $\eta = 1$, will on average expect to trade at the current prices. A more immediate buyer of average will expect to trade at a higher price. A less immediate buyer will expect to obtain a lower price by waiting for a better price. The right-panel illustrates the PDFs for the corresponding QQ-plot curves and the 45-degree diagonal line.

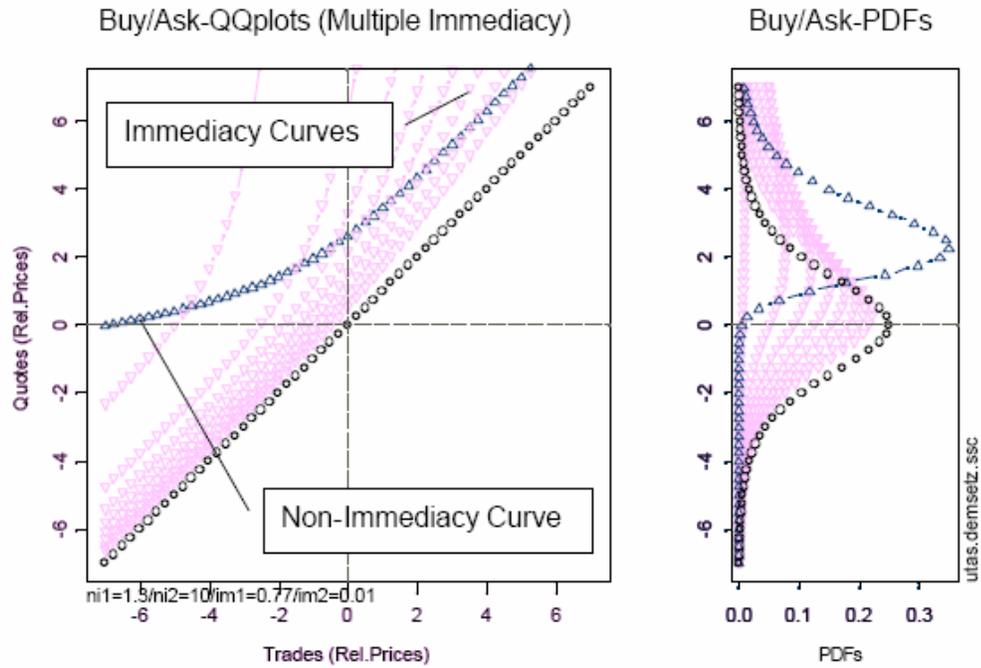


Figure 4-4: Multiple Immediacy Curves/PDFs

Notes: In the left panel, the non-immediacy curve ($n=10$) is the black up-warded curve and the multiple immediacy curves are the grey down-warded curves. The right panel shows the PDFs for the time-preference curves in the right panel.

The left-panel in Figure 4-4 shows a number of buy-immediacy curves (grey down-arrows) together with one sell-non-immediacy curve (black up-arrows) having a time-preference of order $\eta=10$. A trade occurs if an immediacy buyer is willing to transact at the prices quoted by the non-immediacy trader as represented by the black up-arrows. A non-immediacy curve can be analytically determined conditional on the trader's non-immediacy-preferences using Equation (1.7). The immediacy curves, however, are not analytically tractable, meaning that they can only be inferred from the actual clearing prices but cannot be determined uniquely. This indeterminacy is because the market orders are bound by the prevailing non-immediacy prices. Consequently, the prices set by the non-immediacy traders determine the clearing prices. Alternatively, the clearing prices are determined in immediacy/non-immediacy space where non-immediacy traders stand ready to transact with

immediacy traders. Accordingly, immediacy traders are “price-takers” and non-immediacy traders are “price-setters”.¹⁶

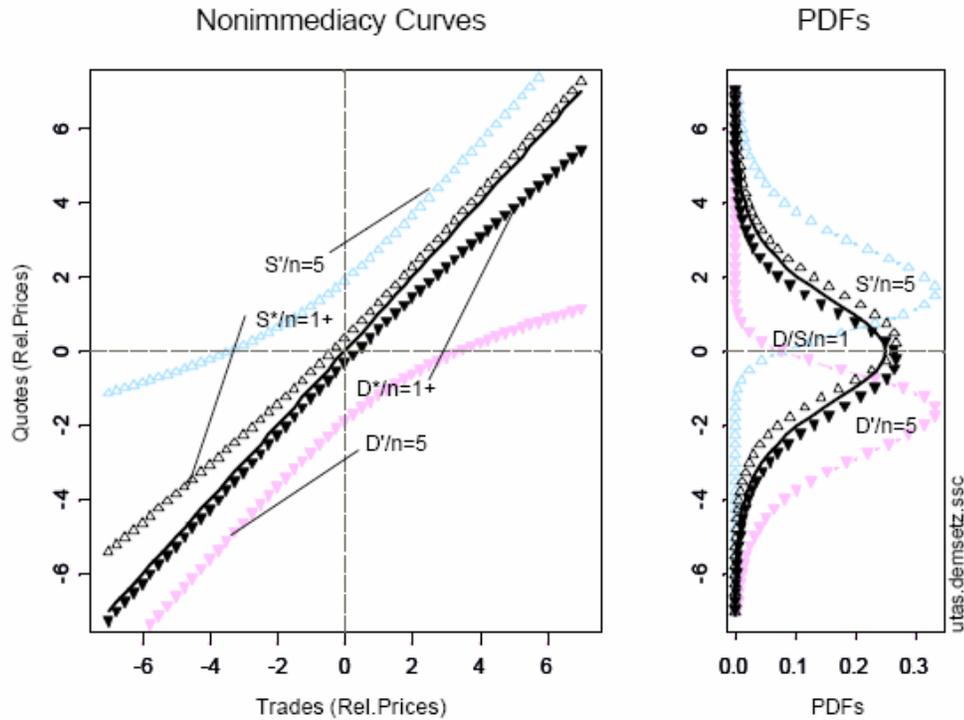


Figure 4-5: Non-immediacy Curves and PDFs

Notes: The up-arrows are the ask non-immediacy curves and ask-quote PDFs and the down-arrows are the bid non-immediacy curves and bid-quote PDFs. The continuous line depicts the immediacy neutral or indifferent curves and the trade PDF.

The Figure 4-5 illustrates together the sell-side and buy-side non-immediacy curves and their associated PDFs for $\eta=1.3$ and $\eta = 5$. The higher the order of the non-immediacy curves, the more curved the corresponding quantile function. For low orders of non-immediacies, the trade PDF can be seen as a mixture of rank-dependent quote PDFs but for higher orders of non-immediacies, the quote PDFs are quite distinct and distant from the trade PDF.

¹⁶ By this definition market-makers are also price-setters but price-setters do not have to be market-makers.

4.1 Inside and Outside Spreads

The inside spread is generally regarded as the implicit cost of immediacy (see Demsetz (1968) and Cohen, Maier, Schwartz and Whitcomb (1979)). In contrast, Treynor (1987) identified the “outside-spread”¹⁷ and not the “inside-spread”¹⁸ as the true cost of immediacy. The outside spread is defined as the difference between the highest ask-price and the lowest bid-price over a particular trading horizon.

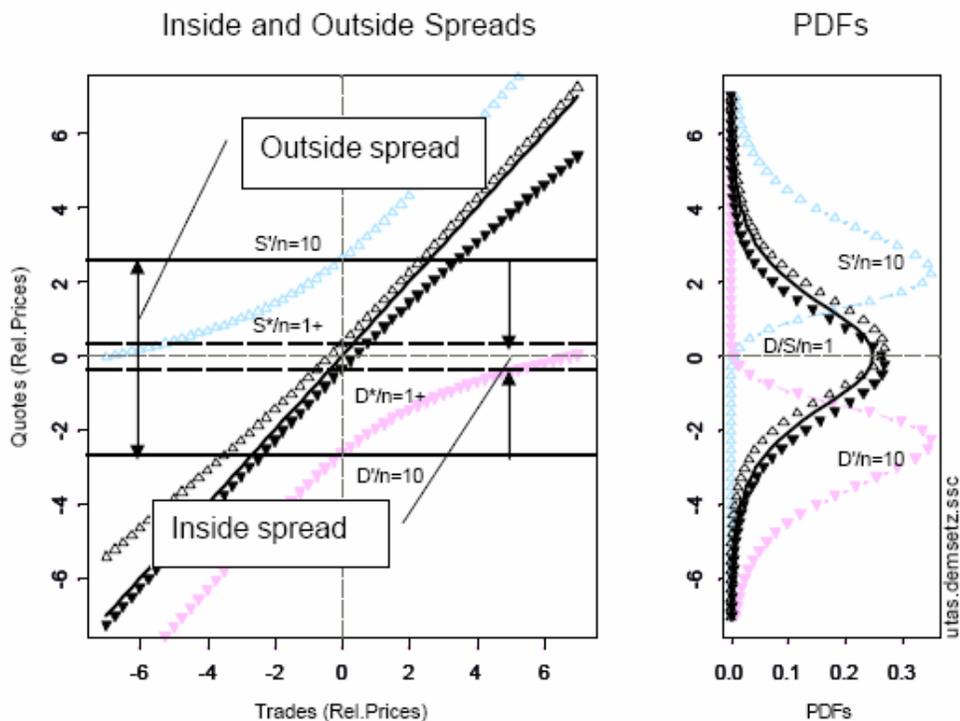


Figure 4-6: Inside and Outside Spreads

Notes: The inside spread is depicted by dashed horizontal lines in the left-panel and the outside spread is delineated by bold horizontal lines in the left panel.

In Figure 4-6 the dashed horizontal lines in the left-panel define the bounds of the inside spread or market spread. The continuous horizontal lines define the bounds of

¹⁷ A spread larger than the minimum tick is referred to as a non-trivial spread. An outside spread, by definition, is a non-trivial spread. An inside spread can be either a trivial or non-trivial spread.

¹⁸ The inside spread is also known as the market spread.

the outside spread. The outside spread is many times larger than the inside spread and thus the true cost of immediacy might be many times more than that assumed by considering only the inside spread.

5 Immediacy and Non-Immediacy Trades

Given the assumptions of immediacy and non-immediacy, one can formulate a very simple trading model to illustrate the dynamics involved in trading. Assume that only non-immediacy traders with $\eta=6$ are present in the market. Figure 5-1 depicts the buy order non-immediacy curve (down arrowed) and the sell order non-immediacy curve (up arrowed) with $\eta=6$.

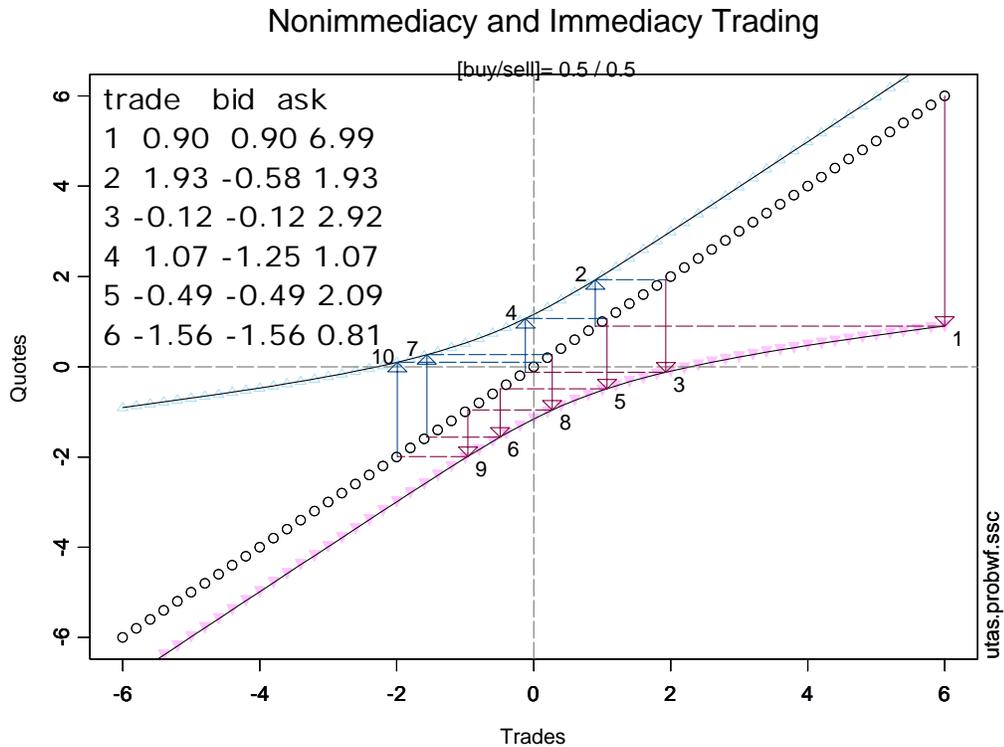


Figure 5-1: Non-immediacy/ Immediacy Trading [n=6]

Notes: The non-immediacy curves are of order $n=6$. The first 10 trades are arrowed in and the first 6 trades are listed together with the bid-ask quotes. The probability of a bid or a ask trade is 1/2 or equal.

If the last relative¹⁹ trade price was at 6.00 (all trades are referenced back to the diagonal) with a particular likelihood, then the new relative bid/ask prices become 0.90/6.99 as quoted by non-immediate traders with a non-immediacy of 6 and with the same likelihood. Say an immediate buyer then decides to trade at the asking-price of 0.90. This price then becomes the current clearing price. The new non-immediacy quotes with the same likelihood of 0.90 are now -0.58/1.93 (bid/ask). Another immediate seller then transacts at 1.93 giving rise to a new current clearing price. Continuing with this process of price generation, trades can be deemed to occur at the prices 0.90, 1.93, -0.12, 1.07 -0.49 and so on until the trade numbered 10 as shown in Figure 5-1 above.

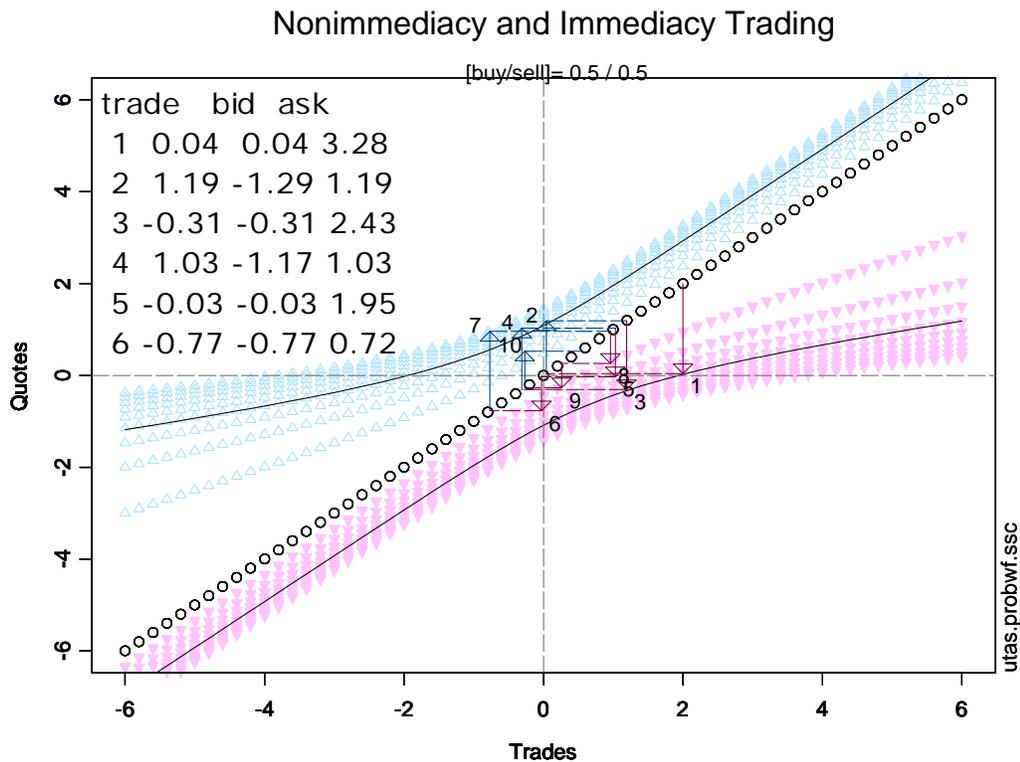


Figure 5-2: Non-immediacy/ Immediacy Trading [n=2-10]
 Notes: The non-immediacy curves are of orders n=2 to 10. The first 10 trades are arrowed in and the first 6 trades are listed together with the bid-ask quotes. The probability of a bid or a ask trade is assumed to be 1/2 (or equal).

¹⁹ The prices are relative, as they refer to the last clearing price of 0 (zero).

As non-immediacy is assumed to be heterogenous, there will be differing non-immediacy traders who will compete in the marketplace and as such assume now there is “zone” of order functions as shown in Figure 5-2. The quotes placed by non-immediacy traders will then be “picked” off by immediacy traders as they arrive at the market. The net result of such behaviour can be simulated and is depicted in Figure 5-3 below.

The top panel in Figure 5-3 shows the “tick-data” and the lower panel shows the “block-data” as simulated using the simple trading model described in Figure 5-2 above. Tick-data lists every trade and inside quotes. Block-data list only trades occurring at the close of each “block” and the maximum and minimum quotes in that block are listed.

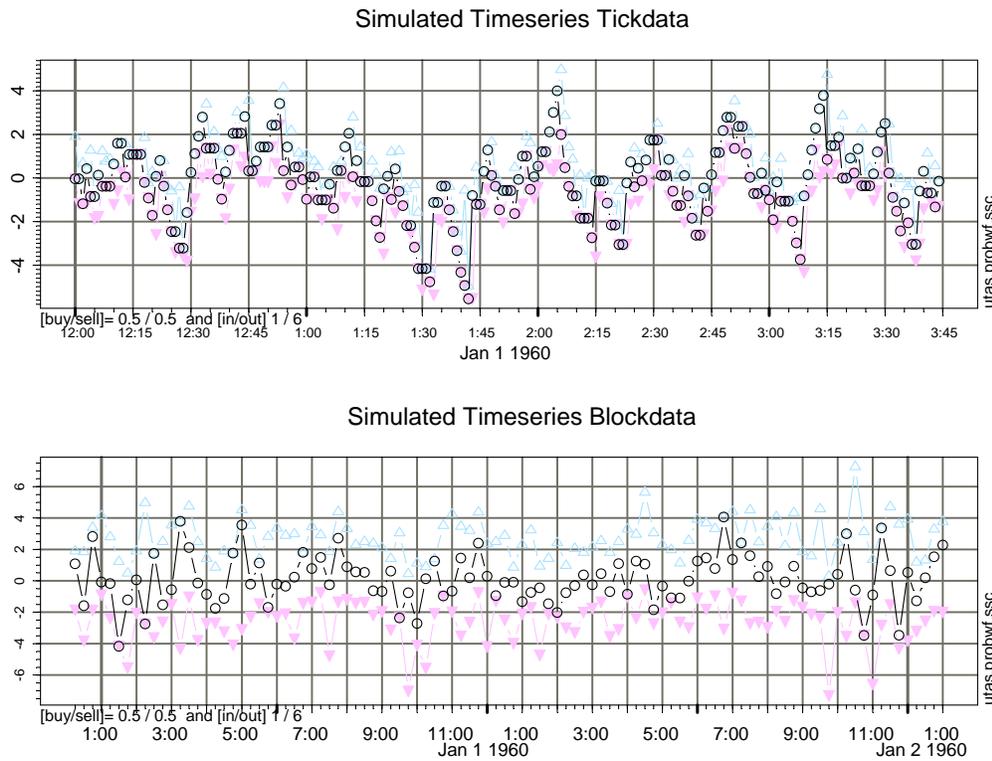


Figure 5-3: Tick-data and Block-data

Notes: Top panel shows tick-by-tick data. Bottom panel shows block-by-block data. The ask-quotes are in up-arrows and the bid-quotes are the down-arrows. The trades are the dots.

The tick-data plot in the top panel distinctively shows that the generally trades occur at the inside quotes. The block-data in the lower panel, however, shows that the trades do sometimes occur at the outside quotes. Only a small percentage of the total

trades occur at the outside quotes. This phenomenon is as expected, as only some of the close prices of the block-by-block data occur as the highest or lowest prices.

Figure 5-4 illustrates the inside-quote PDFs and outside-quote PDFs for quote prices. The trade returns are depicted by the vertically lined PDF. The inside quote PDFs are depicted by the stepped PDFs.

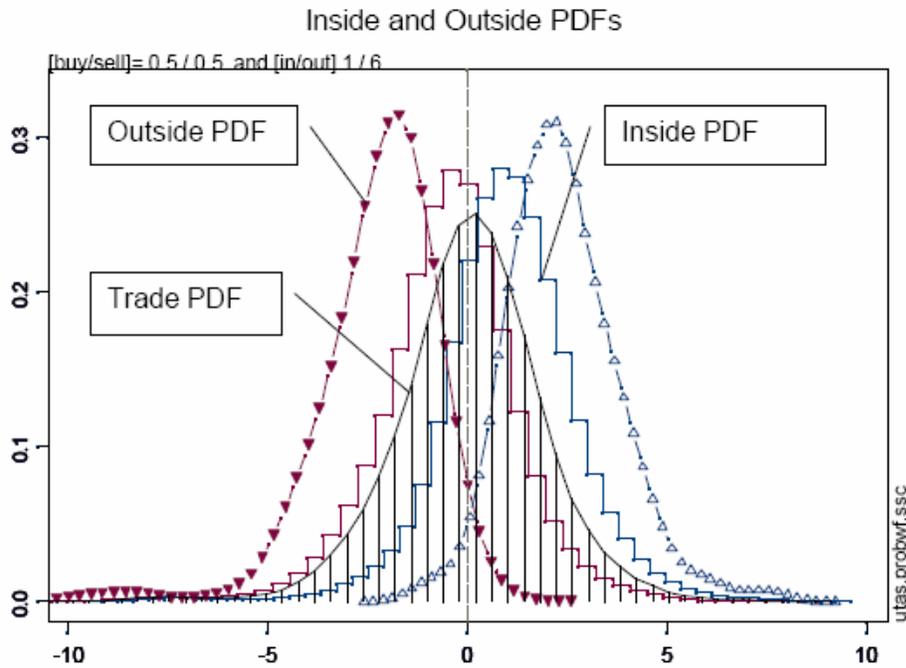


Figure 5-4: Inside and Outside PDFs

Notes: The trade returns are depicted by the vertically lined PDF. The inside quote PDFs are depicted by the stepped PDFs. The arrowed PDFs indicate the outside quotes.

The arrowed PDFs indicate the outside quotes. A large overlap occurs between the inside quote PDFs, whereas the outside quote PDFs have minimal overlap. The inside- and the outside- cumulative distribution functions (CDFs) are illustrated in Figure 5-5. If one considers the distributions of the quotes and trades over the trading horizon, one notes that the CDFs are ordered: the bid CDF is on the left, the trade CDF is in the middle and the offer or ask CDF is on the right in each panel.

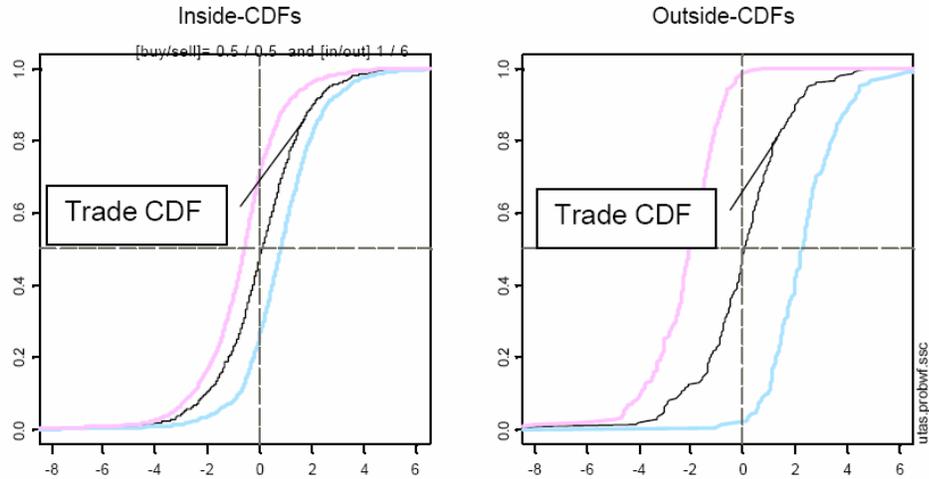


Figure 5-5: Inside and Outside CDFs

Notes: The left-panel shows the CDFs for the inside-quotes with the trade CDF in the middle. The right-panel shows the CDFs for outside-quotes with the trade CDF in the middle.

From a CDF perspective, the right-panel shows the trade-CDF as a possible “convolution” of the outside-CDFs and the left-panel shows the trade PDF as a possible “mixture” of the inside CDFs (see Jeyasreedharan (2004)). These plots are identical to the “S-curves” depicted by Treynor (1987) for defining and describing the outside spreads.

6 Conclusions

In this paper the concept of non-immediacy was defined and shown to be analytically tractable. Immediacy traders place market orders, whereas non-immediacy traders place limit orders. Immediacy and non-immediacy trading gives rise to clearing prices that are not necessarily the equilibrium prices.

The immediacy and non-immediacy concepts introduced enabled the representation of conditional trading expectations based on immediacy/non-immediacy preferences. As these conditional expectations are derived as exact functions of the underlying objective distributions of the clearing prices, one is able to define the exact analytical functions for different groups of non-immediacy traders. The analytical functions are shown to be rank-dependent to maintain conditions of monotonicity and transitivity.

As non-immediacy is heterogenous, a range of non-immediacy traders will compete in the marketplace and as such there will be a “zone” of order functions. The quotes placed by non-immediacy traders will then be “picked” off by immediacy traders as they arrive. Consequently, the distribution of the trade prices can be modelled as a mixture of inside-quote distributions.

It follows that immediacy is a consequence of non-immediacy trading and is provided freely in any organised asset markets when participants transmit their demand propensities to the market in the form of limit orders. While certain participants such as dealers and specialists do indeed (at a price) provide immediacy, they need not be the only providers.

7 Appendix: Order Types

Market participants reveal their trading preferences, in regards to immediacy, by the types of orders they place. Traders who are relatively anxious to transact at a given clearing price submit buy orders that are relatively high or sell orders that are priced relatively low, all else equal. Conversely, traders who are not eager to transact at a given clearing price submit buy orders that are relatively low, or sell orders that are priced relatively high, or may not seek to trade at all. A trader can issue various types of orders to execute his or her trades. These orders can either be market orders or limit orders and/or stop orders and are discussed below.

7.1 Market Orders

A market order is also called an unlimited order. When a market order is issued, an asset is bought or sold at the prevailing market prices. A market order guarantees a transaction (provided of course counterparty quotes are available). A market order does not carry any other additional specifications. The distribution of market order prices will be similar to distribution of clearing prices as every transaction requires at least one market order to consummate a trade (not taking into account any negotiated trades).

7.2 Limit Orders

Limit orders are executed only when the market price is above (for a sell order) or below (for a buy order) a certain threshold set by the investor. A limited-sell order guarantees a minimal price (highest offer) for the sale of an asset. A limited-buy order, vice versa, guarantees a maximal price (lowest bid) for the purchase of the assets. A limit order however, does not guarantee a transaction.

In principle, an investor's desire to trade at any given price can be measured if the investor is able to state a limit order price at which he or she is willing to trade (see Schwartz (1988)). The limit order price is the maximum price for a buy order or the minimum price for a sell order. At the limit order price an investor would be indifferent between transacting and not trading at all. Implicit in this statement is the notion that the distribution of limit order prices will be analogous to the distribution of quote prices as all quote prices are in reality limit order prices.

7.3 Stop Orders

Stop orders are *conditional unlimited orders* triggered by the market price reaching a predetermined threshold. A stop-loss (stop-buy) order issues an unlimited or limited-sell (-buy) order to the exchange once the asset price falls below (rise above) a certain set price. Stop orders are used as a protection against unwanted losses (when owning a stock) or against unexpected rises (when buying a stock). The distribution of stop order prices is harder to establish. If one assumes, however, that traders who use stop orders are aware of non-immediacy trading, then these distributions should also be similar to the limit order price distributions.

8 References

- Black, F, 1995, Equilibrium exchanges, *Financial Analysts Journal* 51, 23-30.
- Cohen, J K, S F Maier, R A Schwartz, and D K Whitcomb, 1981, Transaction costs, order placement strategy, and existence of the bid-ask spread, *Journal of Political Economy* 89, 287-305.
- Cohen, K J, S F Maier, R A Schwartz, and D K Whitcomb, 1979, Market makers and the market spread: A review of recent literature, *Journal of Financial and Quantitative Analysis* XIV, 813-868.
- Dembo, R S, and A Freeman, 1998. *Seeing tomorrow: Rewriting the rules of risk* (John Wiley and Sons).
- Demsetz, H, 1968, The cost of transacting, *Quarterly Journal of Economics* 82, 33-53.
- Else, P, and P Curwen, 1990. *Principles of Economics* (Unwin Hyman, Boston).
- Handa, P, and R A Schwartz, 1996, How best to supply liquidity to a securities market, *Journal of Portfolio Management* 22, 44-52.
- Handa, P, R A Schwartz, and A Tiwari, 1998, The ecology of an order-driven market, *Journal of Portfolio Management* 24, 9.
- Harris, L, 2002. *Trading and Exchanges: Market Microstructure for Practitioners* (Oxford University Press, New York).
- Ho, T, R A Schwartz, and D K Whitcomb, 1985, The trading decision and market clearing under transaction price uncertainty, *Journal of Finance* 30, 21-43.
- Jeyasreedharan, N, 2004, The Extremal Expectations Hypothesis, Unpublished PhD Thesis (Curtin University of Technology, Perth).
- Keynes, J M, 1936. *The General Theory of Employment, Interest and Money* (Macmillan, London).
- Quiggin, J, 1982, A theory of anticipated utility, *Journal of Economic Behaviour and Organisation* 3, 232-43.
- Quiggin, J, 1985, Anticipated utility, subjectively weighted utility and Allias paradox, *Organisational Behaviour and Human Performance* 35, 94-101.
- Quiggin, J, 1993. *Generalised Expected Utility Theory: The Rank-Dependent Model* (Kluwer Academic Publishers, Boston).

Schwartz, R A, 1988. *Equity Markets: Structure, Trading and Performance* (Harper and Row Publishers, New York).

Schwartz, R A, 2000, Building a better stock market: New solutions to old problems, The AEI-Brooking Joint Center Conference: ECNs and the Future of the Securities Markets (Washington).

Schwartz, R A, and D K Whitcomb, 1988, Transaction Costs and Institutional Trading Strategies, Monograph Series in Finance and Economics 1988-2/3 (Salomon Brothers Center for the Study of Financial Institutions and Leonard N Stern School of Business, New York University, New York).

Treynor, J L, 1987, The economics of the dealer function, *Financial Analysts Journal* 43, 27-35.