Agent-based Modelling and Market Microstructure

Jing Yang

A Thesis
in
The Department
of
Economics

Presented in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy at
Concordia University
Montreal, Quebec, Canada

October 2002

0-612-78632-3

Abstract

Agent-based Modelling and Market Microstructure

This thesis advances the literature by applying agent-based simulation to market microstructure issues. Although most of the agent-based literature focuses only on generating market dynamics that best resemble those in the real world, a trading structure is presumed to be a Walrasian auction in most of the published literature. This thesis argues that the institutional details of trading are relevant, and need to be reflected when modeling agent behavior in trading models.

Chapter 2 compares the convergence property in a double-auction market versus a Walrasian market. This chapter constructs an artificial equity market in which agents trade a risky asset that pays a stochastic dividend each period. In some of the experiments, market dynamics under double-auction converge to the Rational Expectation Equilibrium. However, this convergence is sensitive to deviations from rationality. In experiments where we introduce noise trading, convergence becomes unattainable. Minimal rationality is not sufficient to generate convergence in a double-auction market when the market price is endogenous.

Chapter 3 constructs a three-stage model of a dealership market. After trading with the customer, the dealer chooses between trading in a quote-driven inter-dealer market or in an order-driven brokered inter-dealer market. We show that the dealers choice of inter-dealer markets depends on the number of dealers available to make the market. This condition determines which inter-dealer market will prevail. We also demonstrate the conditions for the existence of the equilibrium number of dealers that enter the market making industry. At equilibrium, an increase in risk-aversion, the volatility of the customer order flow

and asset shifts inter-dealer trading from the direct market to the broker. Moreover, customers are better off when dealers have a choice between two inter-dealer trading venues. Building on this theoretical framework, an agent-based simulation is employed to investigate the welfare effects of heterogeneous dealers under different transparency regimes.

Chapter 4 examines the factors that affect liquidity-motivated investors preference between a dealership market and a limit-order book market. This study builds on the theoretical market microstructure literature and uses an agent-based computational approach that allows us to examine the equilibrium properties of models that are richer than the extant literature in terms of the heterogeneity of trading and customer attributes. We find that an increase in the thickness of the market, correlated customer order flow and a decrease in the number of dealers cause a rise in the probability of an order book system prevailing. We also find that large-size customer orders and an increase in the degree of dealer heterogeneity are more likely to shift trades to a dealership system.

Acknowledgements

I wish to express my deepest thanks to my thesis supervisors, Prof. Bryan Campbell and Prof. Lawrence Kryzanowski, for their contribution and support. Prof. Campbell provided much-needed guidance during my graduate education in Concordia University. I especially appreciate every effort that he made to come to Ottawa and discuss my thesis since I started working at the Bank of Canada. Prof. Kryzanowski introduced me to market microstructure research which has interested me ever since and eventually became the topic of my thesis. I am grateful to his generous financial support which allowed me to finish my job market paper. Overall, their contribution greatly improved the final version of this thesis. The efficiency they demonstrated at the later stage of my thesis was truly exceptional.

I would also like to thank my other thesis committee members, Gordon Fisher and Anastasios Anastasopoulos, for their valuable advice during my graduate study.

I benefit a lot from the discussions with my colleagues at the Bank of Canada, especially, Toni Gravelle and Chris D'Souza. A special note of gratitude is extended to Nicolas Audet, whose provocative comments and creative ideas have largely influenced this thesis.

The Doctoral Fellowship, provided by Concordia University for my first three years of study, is gratefully acknowledged.

Last but not least, I would like to thank my family in China, especially my parents who tried all their best throughout their lives to provide me good education. This thesis is as much theirs as it is mine.

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 4

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1

A Survey of Agent-based Financial Markets: Theory, Technology and Applications

1.1 Introduction

The newly developing field of agent-based computational economics concerns the computational study of economies modelled as evolving systems of interacting agents. The agent-based approach is widely implemented in fields such as economics, sociology, electrical engineering, political science and law. In this paper, the scope is limited to the application in financial markets. In an artificial financial market, agent-based models stress the interactions and dynamics among traders learning about the relations between prices and market information. A principle concern of agent-based computational finance research is to understand the apparently spontaneous formation of prices in economic processes, such as the unplanned coordination of trading activities in varied market economies that economists usually associate with the fictional Walrasian auctioneer. The challenge is to explain how these market dynamics arise from the bottom up through the repeated interactions of autonomous agents channeled through trading institutions, rather than from the top-down imposition of coordination mechanisms such as market clearing constraints or the assumption of a representative agent.

Financial markets offer some distinguishing features which make them good candidates for agent-based modeling. First, the issues of price and information dissemination and aggregation tend to be sharper in financial settings where agent's objectives tend to be clear. The complex interaction between prices and information is probably more dramatic in financial markets than in any other economic situation. Second, the complexity of financial markets is represented by many different trading institutions; for example, dealership markets and limit-order book markets, which directly affect price formation. These effects cannot be explained via the notion of Walrasian equilibrium as in a traditional rational expectations model. Finally, financial markets experience frequent regulatory and structural change. To understand the potential impact of a new regulation on the behaviors of agents and, in turn, on market efficiency and market liquidity, agent-based modeling provides a natural experimental environment. LeBaron (2001) provides a good guide for the research on agent-based financial markets.

This thesis advances the literature by applying agent-based simulation on market microstructure issues. As surveyed in this chapter, although most of the agent-based literature focuses only on generating market dynamics which most resemble the real world ones, a trading structure is presumed to be a Walrasian auction in most of the papers. This thesis argues that the institutional framework matters. Chapter 2 compares the convergence property in a double-auction market versus a Walrasian market. This chapter constructs an artificial equity market in which agents trade a risky asset that pays a stochastic dividend each period. We find that in some experiments, market dynamics under double-auction converge to Rational Expectation Equilibrium in a close neighborhood. But this convergence is sensitive to deviations from rationality. In the experiments where we introduce noise trading, the convergence become unattainable. Minimal rationality is not sufficient to generate convergence in a double-auction market when the market price is endogenous.

Chapter 3 constructs a three-stage model of a dealership market. The dealer, after trading with the customer, chooses between trading in a quote-driven inter-dealer market or in an order-driven brokered inter-dealer market. We show that the dealers choice of inter-dealer markets depends on the number of dealers available to make a market. The condition is derived to determine which inter-dealer market will prevail. We also demonstrate the conditions for the existence of the equilibrium number of dealers to enter the market making industry. At equilibrium, an increase in risk-aversion, the volatility of the customer order flow and asset shifts inter-dealer trading from the direct market to the broker. Moreover, customers are better off when dealers have a choice between two inter-dealer trading venues. Building on this theoretical framework, an agent-based simulation is employed to investigate the welfare effects of heterogeneous dealers under different transparency regimes.

Chapter 4 examines the factors that affect a liquidity-motivated investors preference of market structure. We suppose that customers can choose between a dealership market structure and a limit-order book structure. This study builds on the theoretical market microstructure literature and uses an agent-based computational approach that allows us to examine the equilibrium properties of models that are richer than the extant literature in terms of the heterogeneity of trading and customer attributes. The main predictions of our model can be summarized as follows. First, an increase in the thickness of the market, measured by the number of customers submitting orders at any point in time, causes a rise in the probability of an order book system prevailing. Second, customers who submit large-size order flow, such as institutional investors, are more likely to trade on dealership trading systems; customers whose order flows tend to be correlated, are more likely to prefer dealership trading systems as well. Third, a decrease in the number of dealers increases the likelihood that the order-book market will prevail as the preferred trading venue by risk-averse customers. Fourth, an increase in the degree of risk-averse dealer heterogeneity or lower dealer risk-aversion levels (relative to customer risk aversion) increases the likelihood that customers will choose to trade on the dealership system.

This first chapter provides a road map for agent-based research and draws the link between agent-based modeling and market microstructure research. We first outline the basic framework of agent-based modeling including the structure of the economy, agents strategies and the way they interact. Then we survey the recent literature on agent-based research focusing on implications for financial markets. Despite the many insights generated from these studies, most of this literature fails to address the importance of trading mechanisms in a simulated financial market. Chapter 2 advances this area by introducing a continuous double-auction trading mechanism in a simulated market. It studies the impact of market mechanism on the learning behaviors of agents and, in turn, market dynamics compared with a benchmark case of Walrasian auction.

This chapter also surveys the market microstructure literature on comparing auction markets with dealership markets. Most research in market microstructure presumes the existence of a single market mechanism. A few researchers have studied issues of competition between market structures. The fact, however, that dealers are observed to switch between market mechanisms and the considerations which determine such a switch have not been investigated fully. Chapter 3 focuses on the determinants of dealer choice of market mechanism in the context of inter-dealer trading. This chapter is motivated by the on-going debate on introducing an order-book alternative trading venue for the Canadian government bond market. Chapter 4 studies the customers choice of an order-book market versus a dealership market. Using agent-based simulation, this chapter presents a novel method to search for a Nash equilibrium in a numerical way.

The rest of this introductory chapter is organized as follows. Section 2 considers the rational expectations model to clarify how that model's shortcomings fostered the development of the bounded rational expectations theory. Section 3 addresses three important aspects of agent-based modeling: the basic framework of the simulation, agents learning behavior and the trading venue that agents use to interact with each other. Among these three, the last one (the trading institution context) is least frequently tackled in the agent-based literature, but is rather a central focus of market microstructure study. This thesis advances the interaction between the two fields of agent-based modeling and microstructure finance. Section 4 reviews the research on agent-based simulations in financial market applications. Even though these studies address many very interesting questions, none of them investigates the link between trading protocols and market dynamics. This is a core of microstructure studies and the focus of this thesis. This section ends up with a brief summary of Chapter 2 of this thesis. Section 5 reviews the literature on market design and transparency. Much of the literature in this section is motivated by the on-going debates on the relative merits of the dealership market versus auction market and on more trading disclosure versus less. At the end of this section, we draw links between Chapter 3 and Chapter 4 of this thesis and the current literature.

1.2 From Rational Expectations to Bounded Rationality

Typically, dynamic problems under uncertainty involve the formation of expectations at the level of the single agent. According to asset pricing models, Cochrane (2001), for example, the demand for a specific asset depends on the expectations of future payoffs of the asset. In this context there are two important issues: (1) the formation of expectations on the part of each single agent; and (2) the relation between these expectations and the true dynamic relation taking place at the level of the economic system.

One example from the theory of asset pricing may be useful in understanding the challenge posed by these problems and the solution proposed by rational expectations. Suppose there are only two assets and one agent in the economy. One asset pays a fixed and known interest rate r over the period, while another will pay dividend and will have a market price at the end of period t. The dividend and the market price at the end of the period are stochastic and can be described by two probability distributions. The question posed at the beginning of period t, concerns the maximum price that the agent is willing to pay for the risky asset. Under Gaussian distributions for dividends and constant absolute risk averse (CARA) utility functions, it is well know that homogeneous rational expectation equilibrium (HREE) is represented by the following market equilibrium price, see Grossman and Stiglitz (1980) and Bray (1982),

$$P_t = E_t(P_{t+1} + D_{t+1}) - \lambda\sigma^2 S/N, \tag{1}$$

where λ is the constant coefficient of risk aversion and N is the total number of traders. Time is indexed by t and the horizon is indefinite. σ^2 is the variance of the risky asset. The risky asset is issued in units, and pays a stochastic dividend, D_t, which follows an exogenous stochastic process.

In order to calculate this price, the agent needs to form expectations of future dividends and prices. The initial proposal for these expectations problems was to assume the existence of reasonable rules of thumb for the formation of expectations. The theory of rational expectations assumes that the agent forms expectations by using the true distribution of future dividends and prices. Note that in order to do this, the agent must know not only the distribution of the dividends but also the equilibrium model that gives rise to the price as a discounted value of dividends. Rational expectations pricing, therefore, requires *ex ante* knowledge both of the statistical distribution of the exogenous variables and of the economic model that aggregates at the level of the actions of all the agents.

When evaluated on the grounds of realism, the hypothesis of rational expectations seems very strong. It certainly requires an incredible computational ability on the part of the agents who are assumed to know the structure of the economy. The second unrealistic aspect of the rational expectations model, which underlies most of traditional economic modeling, concerns the trading institution through which trades are made. In a rational expectations model, the Walrasian auctioneer serves to aggregate every agent's demand schedule and to clear the market at the equilibrium price before trades occurs. A more realistic view argues that the various frictions in market microstructure and departures from symmetric information must have some real impact on the process of convergence to Walrasian equilibrium prices. This concern implies that the assumptions in the rational expectations model need to be relaxed. These key assumptions are that (1) payoff-related information is common knowledge; (2) agents are homogeneous; and (3) the trading mechanism is inconsequential.

The Bounded Rationality literature relaxes the first two assumptions (Bray(1982) and Sargent (1993)). Agents do not have full knowledge of payoff-related information. Instead, they learn to approximate their true but unknown conditional expectation function and hence form boundedly-rational expectations. The view of learning we put forward in our paper is one in which individuals explicitly construct approximate models of their environment which are updated as their information improves through either active or passive learning. The decision rules they adopt reflect the fact that they know they hold misspecified models of reality. Learning, then, is seen to apply both to the structure and to any relevant parameterization that the approximation entails. In principle, the whole branch of non-parametric inference can be brought to bear on this problem, although we consider only one approach based on artificial neural networks in the following chapters.

Artificial neural networks (in what follows ANN) provide a conceptually simple procedural model of how agents might learn to approximate the unknown conditional expectation function. The minimal structural information required, eventually a knowledge of the input and output variables, does not appear to hinder seriously the performance of the method either in theory or in practice. Moreover, in principle, a neural network approximation can evolve in complexity and hence accuracy as knowledge of the environment increases, reflecting the behavioral aspects of learning suggested above. Such non-parametric learning schemes that require no ex ante knowledge of the structure may serve as good rules of thumb for agent's behavior.

Economics has relied on principles of equilibrium that assume that individual agents are able to solve well-defined optimization problems that in turn can be replaced by global optimization problems in some specific cases. The analysis of individual behavior is far too costly and poses many difficult questions. Accordingly, many simple economic frameworks such as supply and demand curves and representative agents have become the backbone of traditional theorizing. By contrast, the agent-based approach or a bottom-up approach suggests another route. Agent-based modelling is a computerized representation of the way real-life complex systems behave. A model consists of individual agents whose behavior is programmed to follow a set of rules, plus a set of conditions representing the state of an economy. Both agents behavior and the exogenous conditions can be modified to represent a situation of interest; for example, the NASDAQ stock market or a foreign exchange market. The agent-based approach relies heavily on computational tools to push beyond the restrictions of analytic methods. As we discussed above, ANN, among many other non-parametric learning techniques, requires no ex ante knowledge of the structure and may serve as a natural tool in an agent-based model.

The agent-based approach is beginning to show promise as a research methodology and hopes to shed light on questions about heterogeneous trading behavior, market efficiency and institutional design. The next section briefly discusses three important elements of artificial market simulation which comprises an economic environment, expectations formation and a trading institution.

1.3 The Core Elements of Agent-based Market Simulation

There are three building blocks in an agent-based artificial market. First, we need to lay out an economic environment where agents interact with each other. Second, we need to describe how agents form their expectations; in other words, what learning algorithm agents use. Third, we must describe a trading mechanism which defines how the trades are processed. In this section we start with a review of the current literature along these three dimensions. Table 1 summarizes the recent developments in this literature.

1.3.1 Economic Environment

The first thing that needs to be described in building an agent-based artificial model is the economic set up. Most of the existing literature uses the rational expectations-based asset pricing models to describe agents economic environments (see Table 1). The market, considered in what follows, consists of a risk-free and a risky asset, and traders decide on their asset allocation in each of the assets. The stock pays a dividend, , which has both predictable and unpredictable components. In the Santa Fe stock market as described below, the dividend for the risky asset is assumed to follow a linear autoregressive process of order one (AR(1)) as follows,

$$D_{t+1} = \overline{d} + \rho(D_t - \overline{d}) + \xi_t, \tag{2}$$

where ξ_t is Gaussian, independent, and identically distributed.

The agents have constant absolute risk aversion (CARA) utility functions and decide on their demand for shares. It is well known that under CARA and Gaussian distribution of dividends, agents demand is a linear function in price; i.e.:

$$Q_t^j = \frac{\widehat{E}_t^j(P_{t+1} + D_{t+1}) - (1+r)P_t}{\lambda \widehat{\sigma}^2_{P_{t+1}+D_{t+1}}}, \tag{3}$$

where is the price of the risky asset at time t, denotes the expectation for trader j at time t, and is the conditional variance of price and dividend. Equation (10) is a well-known linear demand function under CARA and it is valid if the shocks around the above expectations are Gaussian. The agents demand for the shares is increasing in the forecast of future prices and dividends, and decreasing in the estimate of the variance of the stock. Many of the agent-design issues arise when building these forecasts from past time series data. Forecasting is at the heart of this system, which we discuss in the following section.

1.3.2 ANN Learning Mechanism

Agent-based financial modeling tries to characterize boundedly-rational agents who can map a time series of past prices into a forecast of future prices. In much of the agent-based literature, various learning methods are employed as the agents forecasting tool such as the genetic algorithm (Arifovic 1996), classifier system (Arthur *et al.* 1997 and LeBaron *et al.* 1999), neural network (Beltratti *et al.* 1996) and Bayesian learning (Chan, LeBaron, Lo and Poggio 2001). But at this moment, no learning mechanism appears to be more useful than the others.

In our paper, we use artificial neural networks (ANN), a technique that essentially provides non-parametric estimates of unknown parameters or conditional expectation functions representing the boundedly rational expectations. No assumption about the state of economy is required to implement this technique; only information about the set of input variables and the output for ANN to learn to generate expectations.

Since the ANN algorithm constructs a time-varying nonlinear approximation to the unknown conditional expectations function, it is quite different from the least squares or Bayesian algorithms that essentially assume there is a fixed constant structure characterized by the value of the parameters that must be learned. There are also a number of potential theoretical advantages arising from the use of neural networks. In the first place, they provide the ability to combine information from separate models and process this information in a parallel fashion. In addition, given their non-parametric nature, they can be robust to poor information. The main reason why they are of interest in the present context is the theoretical result that demonstrates their ability to approximate any measurable function to any required accuracy (see White 1989, 1992). A detailed neural network structure is presented in Appendix 1.

1.3.3 Trading Institutions

In the classical economics literature, there is no specific role for market institutions. Price formation is performed by a Walrasian auction, which is really a one-shot clearinghouse. Traders submit their demand and supply schedules to the auctioneer and the final allocation arises from the maximal net trade at a single market clearing price. This process is often called institution free. However, while this framework does provide a simple and elegant way to envision price setting, it does not capture the process by which equilibrium prices arise. If trading involves more than simply matching of supply and demand in equilibrium, then the trading mechanism and the behavior of traders will influence the resulting prices.

A market institution is a set of rules specifying which sorts of bids and other messages are legitimate, and how and when specific traders transact. A market institution defines an exchange process by specifying the set of admissible messages (i.e. order price/or quantity), and by specifying a final and an initial allocation. In reality, it is very rare to see a pure Walrasian auction institution. An overview on major types of real-world market institutions will help guide the following discussion.

We begin with a taxonomy of market structures in the real world. Trading systems exhibit considerable heterogeneity in these dimensions, as shown in Table 2. Across developed economies, one observes two predominant types of equity markets: order-driven, auction-agency markets and dealership markets. Order-driven markets or order-book markets, such as the Toronto Stock Exchange (TSE) and the Paris Bourse, are structured as two-sided auctions in which there is no intermediary. Incoming orders submitted to the market are either matched up with offsetting standing orders previously submitted to the market or are placed in an order book themselves. Automated limit order book systems of the type used by the Toronto Stock Exchange and Paris Bourse offer continuous trading with high degrees of transparency, such as public display of current and away limit orders without reliance on dealers.

A dealership market can be classified as a single dealership or a multiple dealership market. NYSE is a specialist (single dealer) market, while NASDAQ and London Stock Exchange (LSE), foreign exchange (FX) and government bond markets are competitive multiple-dealer markets. In terms of market transparency, FX and government bond markets offer very little transparency, while other dealer markets such as NASDAQ and LSE offer moderate degrees of transparency. In fact, NASDAQ provides information on the full order book.

Unfortunately, very few papers in the existing agent-based literature study the impact of trading mechanism on market liquidity and price efficiency. Table 1 summarizes the current agent-based literature in terms of their basic model, learning and trading institutions. Most of the existing agent-based simulation is set up in a Walrasian auction framework. LeBaron (2000, 2001) surveys many of these other computational markets. These more complicated simulated markets are interested in long-run market behavior and tend to ignore the trading mechanisms which are at the center of our research. This research opens up a totally new area for the application of the agent-based approach to the issues related to market microstructure.

The following two sections serve as road maps for research in agent-based simulation and market institutional design respectively. Section 4 surveys the most recent developments in agent-based research and the related technology and basic findings. Section 5 selectively discusses market microstructure research, which is closely related to the issues raised in Chapter 3 of the thesis.

1.4 Literature Review I: Agent-based Financial Markets

There is a growing literature that attempts to model financial interactions starting from the agent's perspective. Computational agent-based models stress interactions and learning dynamics in groups of traders learning about the relations between prices and market information. Building heterogeneous agent rational expectations models are certainly not new in finance. The current literature of computational frameworks tackles the problem of very complex heterogeneity and in so doing leaves the territory of what can be handled analytically.

The Santa Fe stock market, described by Arthur *et al.* (1997) and LeBaron *et al.* (1999), is one of the benchmark artificial market studies. The model tries to combine a well-defined rational expectations environment with inductive learning using a classifier system. The model is set up in a similar framework to that of Bray (1982) and Grossman and Stiglitz (1980). In this one-period framework, myopic and constant absolute risk averse (CARA) agents try to decide on their optimal asset allocation between a risk-free bond and a risky stock which pays a stochastic dividend. As indicated above, under Gaussian distributions for dividends and prices and CARA utility, it is well know that the demand for the risky asset is a linear function of the market price. In equilibrium, where aggregated market demand equals total supply of the shares, there is a well defined linear homogeneous rational expectations equilibrium. Up to this point, this is a simple and traditional economic framework. Where this approach breaks from the traditional framework is in the formation of expectations. The paper introduces a classifier system for each individual agent to forecast the next state of the economy. Agents use the classifier to code current market information into bit strings based on the symbols 1, 0, #. The symbol 1 here represents true, 0 false and # is a wild card which matches anything. Therefore, all of the learning rules can be represented by these strings, and each rule maps into a real vector of forecast parameters which the agents use to form their expectations.

In such a world, agents update their forecasting rules every K periods. For example, in Arthur *et al.* (1997), new rules are generated by using a genetic algorithm (GA) with uniform crossover and mutation. One important parameter in this study is the learning rate. This paper compares the slow learning case with a fast learning one, and concludes that in a slow learning case, the economy appears very close to the rational expectations equilibrium, while in a fast learning case the simulated market can reproduce some features which have been observed in real financial markets, such as volatility persistence and high trading volume.

Even though this study is one of the most adventurous and complex artificial markets, the fact that the market clears as a Walrasian auction makes it less representative in term of mimicking actual financial markets. In the second chapter of this thesis, we try to go beyond this study by introducing a double-auction trading mechanism which is widely used in many real-world trading venues.

In another line of research, Arifovic (1996) makes the first attempt to simulate foreign exchange (FX) markets in the spirit of the two-country overlapping generation model. Agents use the genetic algorithm (GA) to update their savings and portfolio decisions regarding what fraction of savings to place in each currency. Arifovic (1996) conducts some laboratory experiments with human subjects on this foreign exchange model. Two important features emerge from her work. One is that the exchange rate fluctuates without a tendency to settle to a constant value; and the other is that the first-period consumption is close to the stationary equilibrium value. Arifovic then constructs a genetic algorithm learning environment to test learning in this FX market. In this setup, the portfolio decisions of agents endogenously determine the price levels in the market. The population of agents evolves according to their fitness value and a new population is created with crossover and mutation operators in a standard GA fashion. The election operator introduces one new operator for the dynamics of the population. The simulations of GA capture the two important features of the human-subject experiment and result in the persistent fluctuations of the exchange rate. The paper suggests that the stationary rational expectations equilibria are unstable.

Arifovic's paper points to one important issue for artificial markets in the general equilibrium setting with endogenous price formation. The limitation of the paper is that this artificial FX market totally ignores all the features of real FX market structures. It is rather a simulation version of an overlapping generation model at the macroeconomic level. In Chapter 3, we try to simulate an artificial FX market from a microeconomic perspective by considering the fact that the FX market is a hybrid of dealership and limit-order markets.

Although most of the artificial market studies are formulated in a Walrasian auction spirit, there are a few studies that consider different market structures. Beltratti and Margarita and Terna (1996) simulate markets where trade occurs in a decentralized fashion. Traders are randomly paired with each other, and traders trade according to the differences in their expectations. In this case, trading may occur at different prices at the same time.

The main focus of this work is the evolution of the combination of the populations of different types of traders. They analyze the stock dynamics and the changing fraction of different agent types. They find that in the early stage of trading which is very volatile, it is worthwhile to pay to be smart. After those volatile periods, the value added of marginal forecasting power is reduced below its costs and the smart agents disappear.

Another feature of this work is the introduction of neural networks for modeling the formation of agent's expectations. The smart traders are endowed with a neural network which takes last period's prices and dividends as inputs and estimates next periods price as an output. Other types of traders, naive trader and noise trader, can pay certain amounts to buy these ANN; they can pay to be smart. Then the authors investigate the evolution of traders types. In most of their studies, the neural network structure is kept simple.

The second chapter adopts a similar learning mechanism for agents, yet differs from the Beltratti *et al.* (1996) study with regard to the trading mechanism. Beltratti *et al.* (1996) uses a decentralized trading mechanism, while we introduce a double-auction mechanism and investigate the impact of the trading mechanisms on the market dynamics under Walrasian auction and double-auction.

Another study which incorporates trading mechanism and asymmetric information is undertaken by Chan, LeBaron, Lo and Poggio (2001). They use a one-period stock, which pays state-contingent dividends at the end of the period. The distribution of the state of the economy is common knowledge. Risk-neutral agents maximize the end-of-period value of their portfolios by forecasting the dividend using Bayesian methods. Three types of traders, empirical Bayesian traders, momentum traders and noise traders, populate the market and trade in a double-auction environment. The risk neutrality assumption implies that the rational expectations equilibrium price is simply the value of the dividend at a given state. The study focuses on information efficiency by studying how close and how fast prices converge to the REE price. They consider the price deviation of the last 20 transactions from REE price as a measure of closeness to the REE. Using simple learning agents, the paper shows that the simulations reproduce some features of human-subject experiments. The limitation of the paper is its very simple information structure. The underlining distribution of the states is discrete and uniform and it is common knowledge among agents. More precisely, the simulation is set in an economy with three states; for instance, the stock pays a dividend of {\$0, \$1, \$2}. The assumption of risk neutrality and the simple information structure limit the richness of the results.

The models discussed to this point represent a wide range of possible agent constructions. But many questions still remain unanswered. Which type of trading mechanism is relatively more efficient? Which type of trading mechanism provides higher liquidity? Which type of trading mechanism should be adopted for certain types of markets? These policy-related decisions may have major consequences. The need for a centralized market is also an interesting policy-related issue. Among the papers surveyed in this section, only the paper by Beltratti *et al.* (1996) uses a decentralized trading mechanism. In short, it is important to understand the impact on the market quality of the trading mechanism to be used. Among all the issues studied in agent-based models, these may have the most direct relevance for policy.

This thesis attempts to add to the current agent-based-simulation literature by considering the impact of trading mechanisms on market liquidity and price efficiency and comparing different market institutional settings. In Chapter 2 we introduce the double-auction mechanism to mimic a limit-order book stock market such as the TSE and Paris Bourse. Endogenous payoffs, heterogeneous agents and double-auction are the three features of the artificial stock market setting to be used in the analysis. Simulations in Chapter 2 compare the market dynamics resulting from a double-auction setting and a Walrasian auction setting. We find that the market price under double-auction does not always converge to the rational expectations equilibrium as in the Walrasian auction when the heterogeneity of the market participants increases. Welfare effects across different types of agents are analyzed as well.

The trading mechanism is a well-studied issue in the market microstructure literature. In the next section we survey some of the most recent developments in this area as a backdrop to the simulations that we conduct in Chapter 3.

1.5 Literature Review II: Market Institutional Design and Market Transparency

Much research exists on the comparison of order book market and dealership markets, on how different market structures can affect market quality, and on the impact of trade disclosure regulations on both types of markets. Most of these studies are motivated by the ongoing debate on market consolidation, fragmentation and increase of market transparency. Since Chapter 3 addresses these market design questions in an inter-dealer setting, we review the existing literature on market design with an emphasis on the dealership market.

In order to classify the existing research, it is useful to note that the existing dealership markets are organized along two parallel and, in effect, separate markets. These are a public trading environment where customers first trade with market makers, and an interdealer trading environment, where dealers then trade among themselves (called the inter-dealer trade herein). For the inter-dealer trade, dealers have the choice of either trading bilaterally with each other as in a quote-driven market or trading indirectly and anonymously with each other via an interdealer broker (IDB) known as the limit-order book market.

Inter-dealer trading between dealers who act as market makers is a common feature of most major financial markets and accounts for a significant proportion of the total trading volume. In equity markets such as the London Stock Exchange, inter-dealer trading constitutes some 40% of the total volume, and occurs mainly via an anonymous limit-order book; see Hansch, Haik and Viswanathan (1998). In the market for U.S. Treasuries, trading between the market makers accounts for about 40% of total turnover (Viswanathan and Wang 2000). Two-thirds of these transactions are handled by inter-dealer brokerage firms such as Garban and Cantor-Fitzgerald, and the remaining one third is done via direct dealings between the primary dealers. In the FX market, inter-dealer trading accounts for about 85% of the trades (Lyons 1995). Hence, in most financial markets, inter-dealer trading is an integral part of market design, particularly for institutional markets that deal with large orders. At the same time, the structure of inter-dealer trading is undergoing significant change.

One aspect of microstructure characteristics of dealer markets that has been largely ignored in the theoretical literature, at least until recently, is the role played by the existence of interdealer markets that operate in parallel to the customer or public market. In particular, little research examines the microstructure of these interdealer markets in which dealers have a choice of trading bilaterally with other market makers or trading in an interdealer broker market which resembles an order-driven market.

Why do these markets rely on the intermediation of dealers? This issue is referred to as the Dealers Puzzle, which really consists of two parts. First, what are the functions of market makers that make their presence valuable? Second, why do public auction markets not provide the same functions? Another issue relates to the role of market transparency. Is higher transparency always preferred in order to improve market quality? This can be viewed as the transparency puzzle. We review the literature on market design in the following section. Much of this literature is heavily influenced by on-going debates about auction versus dealership markets, and greater trading disclosure versus less transparency.[1]

1.5.1 Theoretical Literature

An excellent comparative review of trading systems can be found in Madhavan (1992). He compares an auction market with a dealership market in the context of asymmetric information. He shows that a market organized as a dealer system gives liquidity traders a better price than a continuous auction system. The reason is that price competition forces the expected profits of market makers on each trade to zero. The price is equal to the

[1] The market transparency issues that we review in this section do not cover the literature on circuit breakers and trading halts. For a review of this literature, and the effectiveness of such trading stoppages to disclosure information, see Kryzanowski and Nemiroff (2001).

asset's expected value and contingent on order size, where in an auction system the transaction price is not known at the time of order submission. By behaving strategically, i.e. conditioning the demand on the market clearing price, each player has some influence on the market clearing price in an auction market. This strategic behavior results in higher price variability and less efficient prices. However, this finding is directly related to the fact that dealership markets are modeled in this paper as a price competition and auction markets are modeled as a quantity competition.

Taking an approach different from Madhavan (1992), Vogler (1997) abstracts from information asymmetries and develops a two-period model to compare a dealership market, which is then followed by an inter-dealer trading round, in a pure auction market. He models both first period public trading and the dealership market as a Bertrand competition. By contrast, the auction market is modeled as a quantity competition game. He shows that, if trading is motivated by non-informational motives, a dealership market also gives a better price as long as the probability that the market maker cannot execute his order in the inter-dealer market is low. He also finds that, when the number of market makers increases, the inter-dealer market shows higher risk-allocation efficiency, as measured by the closeness of market makers asset allocation with the ones in the perfect competition case.

The findings of Naik, Neuberger and Viswanathan (1999) are of particular interest. A two-stage theoretical model is developed that most closely resembles dealership markets. Naik *et al.*. compare the welfare effects under two trading disclosure regimes. The key and realistic feature of their model is the fact that the price the dealer offers to the public investor (in the first stage of trading) depends on his ability and the costs to manage his inventory position. They find that it is essential to model the second stage in order to understand how prices are set in the first stage. An additional innovation over both the Madhavan (1992) and Vogler (1997) studies, where public investors are assumed to be risk neutral, is that Naik *et al.* also model both the public investor and the market maker as being risk averse and, accordingly, sensitive to price risks. Naik *et al.*. show that in some circumstances requiring full and prompt disclosure of public-trade in a dealership market may reduce the welfare of public risk averse investors. The authors also show that an increase in the level of post-trade transparency works against the execution of large trades, but nevertheless tends to benefit small traders.

Madhavan (1995) develops a model to compare upstairs markets and downstairs markets. The upstairs market operates as a search-brokerage mechanism where prices are determined through negotiation. By contrast, downstairs markets are characterized by their ability to provide immediate execution at quoted prices. Upstairs trading captures the willingness of traders to seek execution outside the primary market, and hence is of interest in debates regarding market consolidation and fragmentation. Madhavan argues that large traders are afraid of having their strategies leaked and prefer to use upstairs markets to accomplish large-block trades in one single step. By contrast, small, uninformed traders (uninformed in terms of the stocks fundamental value) are better off using more transparent downstairs markets.

Lyons (1996) also attempts to model the risk sharing aspects of dealership markets in the presence of asymmetric information. He models the realistic feature of two-stage trading in multi-dealer markets. Lyons presents a hot potato model of the foreign exchange market that allows dealers to off-load their order flow to other dealers. While Naik, Neuberger and Viswanathan (1999) focus on the issue of the transparency of public trades, Lyons considers the optimal level of transparency in interdealer trading that dealers would endogenously choose. He does not examine the optimal level of transparency for the market as a whole. This question is of interest since the FX market is one that is unregulated and yet there is a degree of transparency that has arisen without regulatory prodding. For example, in the FX market, dealers can observe the order flow that goes through the IDBs. Lyons shows that dealers, if they could choose the transparency regime ex ante, prefer incomplete transparency over both a completely opaque setting and a completely transparent setting. The key insight that explains this finding is that incomplete transparency allows for the optimal level of risk sharing to occur in the market. If there is no transparency, customers do not wish to trade and as such dealers cannot share the price risk borne by themselves in the first round of trading. Under complete transparency, the dealers bear all of the price risk (in the first round) and the possibility of sharing this risk by trading with customers in the second round may be limited. If post-trade transparency is complete, prices adjust (nearly) instantaneously, and dealers absorb all of the price risk instantaneously. Accordingly, dealers will prefer some middle ground of transparency.

1.5.2 Empirical Literature

Gemmill (1996) examines how changes to the transparency regime on the London Stock Exchange (LSE) that occurred in 1989 (at that time structured as a dealership market) affected market efficiency and liquidity. In reviewing the data that covered 1987 to 1992, Gemmill found no evidence that prices adjusted less quickly (in reaction to news or trading) after a sharp reduction in transparency. Overall, these results do not support the hypothesis that an increase in post-trade transparency will improve market efficiency. This study also found that, on average, the relative spreads (measured as the actual price paid compared to the going quote used to account for negotiated price improvements) for large trades versus small trades were narrower under the low transparency regime. These studies did not reveal any statistically significant price effects of increased post-trade transparency.

Studies of London trading in French and German listed stocks by DeLong *et al.* (1995) find that spreads are lower on continental auction markets than on the London dealer market. Therefore, it appears that spreads in dealership markets (London and NASDAQ) are higher than in auction markets (continental exchange and NYSE). But one should interpret their results with caution. Zhang and Kryzanowski (2002 forthcoming) show that it is not correct to compare spreads across markets unless their mid-spreads are identical.

Porter and Weaver (1998b) investigate the effects of greater pre-trade transparency using data from the TSE. A regime change in pre-trade transparency occurred on April 12, 1990 when the TSE started providing real-time public dissemination of the firm quotations for up to four levels away from the inside market (in both directions). Porter and Weaver find that there is a decrease in liquidity, measured by the realized spread, with the increased transparency regime.

Generally, there is little empirical evidence supporting the theoretical hypothesis that public investors are better off in a dealership market than in an auction market, as well as the hypothesis that market efficiency and the welfare of public investors is improved under more transparent regimes.

However, tests of theories concerning market structure face a serious problem: the absence of high quality data that allows researchers to pose what if questions. Traders adjust their strategies in response to market protocols and information. This feature makes it difficult to assess the impact of market protocols. Further, these studies are event studies, and most empirical studies are limited in that there are not large samples of events to study. Moreover, changes directly resulting from the changes in transparency regime are often accompanied by design alterations in other dimensions as well. It is difficult to distinguish from other unrelated changes that may have occurred during the sample period in question. As a result, it is important to find a cleaner environment in which to test the effects of transparency changes. It is clear that an analysis based on experimental markets in which markets are simulated in a controlled (laboratory) environment could shed considerable light on these issues. But this type of research exposes the researcher to other challenges which we discuss in detail in the next section.

1.5.3 Experimental Literature

Perhaps the research of most direct relevance to the issue of the appropriate level of transparency in a market is that based on experimental economics where hypotheses may be tested in a controlled environment. In a laboratory study, human subjects trade in artificial markets. The questions which researchers study are often the effects of various changes in protocols; for example, the effects of changes in trading disclosure on market quality. Some experimental studies also study quality variables that might not otherwise be observed. These include trader type (e.g., informed trader or uninformed trader), the accuracy of the beliefs of market participants over time, and the rate of convergence of prices to full information values.

Bloomfield and O'Hara (1999) examine the speed at which prices converge to full-information values and the possible impact of changes in disclosure rules in a multiple dealer market. Price bubbles are generated easily, even if the distribution of fundamental values are common knowledge across traders. Moreover, prices in auction markets need not always converge to full information values. Instead, prices may settle at the wrong value, since agents may learn incorrectly. Bloomfield and O'Hara (1999) find that price efficiency is not significantly affected when going from opaque to semi-opaque markets, but find that prices move rapidly to their new fundamental level (when new information about the fundamentals is released) in the fully transparent setting. Price efficiency in this paper is defined as the speed with which the price converges to the full-information value of the securities. Alternatively, price efficiency is the speed with which information is embodied into the price. They also find that changes in transparency cause significant changes in the traders' profits. Informed traders are made worse off by going to the most transparent setting. However, liquidity traders (who need immediacy) are made significantly worse off in the transparent setting. This is surprising since it contradicts the findings of Pagano and Roell (1996) who predict that an increase in transparency shifts the distribution of returns from the informed traders towards the uninformed, liquidity traders. They also find dealers compete less for order flow and spreads are wider in the transparent setting.

Another interesting experimental study that sheds some light on regulatory arbitrage is conducted by Bloomfield and O'Hara (2000) in which dealers can either be those who disclose their trades (thus being post-trade transparent) and those who do not. They find that non-transparent dealers are more aggressive at setting quotes allowing them to capture more order flow. This greater share of order flow in turn provides them with an informational advantage (over the disclosing dealers) and allows them to quote narrower spreads (which, although not directly examined in the study, benefit the non-dealer traders in principal). The low-transparency dealers profits are significantly higher than those for transparent dealers. Bloomfield and O'Hara (2000) suggest that markets will naturally gravitate towards less transparent regulatory settings.

Flood *et al.* (1999) conduct an experimental economic study modelled as a multiple dealership market. Their work differs in that traders face a time pressure which places a greater emphasis on the traders learning ability. Also, they only examine changes in pre-trade transparency, leaving the post-trade environment opaque. They find that, as the level of pre-trade transparency increases, trading activity increases, although the returns of the traders are not significantly affected in a statistical way. They find that greater pre-trade transparency slows down this price discovery process (i.e., price efficiency is significantly decreased). This surprising result is in marked contrast to the Bloomfield and O'Hara (1999) study and to much of the theoretical literature on transparency. They hypothesize that this result is due to the fact that dealers, given that they no longer need to seek out pre-trade information by calling for and making quotations, become less aggressive price setters in the more transparent setting. In turn, the price converges more slowly to the one predicated by the fundamentals. In essence, the greater transparency reduces the incentives of dealers to acquire new information.

This new literature uses more realistic trading regimes and incorporates institutional features found in actual markets. The advantage of this approach is that predictions of theoretical models can be examined in a controlled environment. The disadvantage of this approach is that the results are often sensitive to subtle differences in the experiments or the instructions given to participants and the idiosyncrasies of participants.

This has led to a growing new field of agent-based modelling. Using computer software to replace human subjects, agent-based modelling allows better control on individual features of the participants and the modeling of their learning processes. Even though agent-based modelling is a natural candidate for studying market architecture hypotheses, very little work has been done in this area. This thesis is a contribution to this evolving literature.

Chapter 3 in this thesis attempts to explore the applications of agent-based models on financial markets with special focus on the market institutional design. We try to mimic the structure of a typical dealership market such as the foreign exchange (FX) market and fixed income market. In this chapter, two types of inter-dealer trading are investigated. They are direct inter-dealer trading organized as a bilateral search, quote-driven market, and a brokered inter-dealer trading which is run as an electronic limit-order book market. In reality, dealers are observed to conduct their inter-dealer trading through two different inter-dealer markets. The considerations, however, which determine switching between markets have not been investigated. Chapter 3 uses the agent-based approach to provide some new insights on the factors that determine the decisions of dealers, and the impact on the welfare of both the public and dealers.

1.A Appendix: Artificial Neural Networks

Halbert White and various co-workers have shown that feedforward neural networks are best regarded as approximates of nonlinear function g mapping vectors $x \in X$ into vector $y \in Y$.

A feedforward neural network with one hidden layer is described by the two equations

$$Y = \alpha_0 + \sum_{k=1}^{K} \alpha_k a_k, \tag{4}$$

and

$$a_k = f(\sum_{k=1}^{K} \beta_k^{'} X + \gamma_k) \tag{5}$$

The second equation describes the output of hidden unit k. The first equation generates the output Y of the network by taking an affine function of the vector of outputs of the hidden units. The parameters of the network are the weights W, which include both α_k and β_k. An example of such a network is depicted in Figures 1 and 2. Figure 1 pictures computational elements or node with N inputs and one output (weighted sum of inputs). Three representative examples of nonlinearities are shown in the lower panel of the graph.

The literature has addressed two issues about models of this class: the issue of approximation or representation, and the issue of estimation. The approximation literature describes the class of functions $g(\cdot) : X \rightarrow Y$ that can be arbitrarily well approximated by the above model. Hornik, Stinchcombe, and White (1989) have shown that if enough hidden units are included, then any Borel measurable function from one finite dimensional space to another can be approximated by this model. The parameter that controls the quality of approximation is K, the number of hidden units. For a given K, the best mean-square approximator to a function $g(x)$ is determined by the value of (α, β) that minimizes the squared norm

$$\left\| Y - \sum_{k=1}^{K} \alpha_k f(\beta_k' X + \gamma_k) + \alpha_0 \right\|^2 \tag{6}$$

For a given K, the best approximator can be found by variations of the Newton method. The approximation literature comforts us by assuring us that, if we select K large enough, we can find a pair of (α, β)that makes this mean squared error arbitrarily small.

The estimation problem occurs when we are given a sample $\{Y_t, X_t\}_{t=1}^{T}$ and a particular model of form $f(\cdot)$, with a fixed K, and want to estimate the parameters. This is a version of the nonlinear regression problem. Estimation can proceed by using versions of the stochastic gradient algorithm or stochastic Newton algorithm. The training term discussed in the neural network literature all use versions of those algorithms.

Figure 1. A Node with N Inputs and One Output

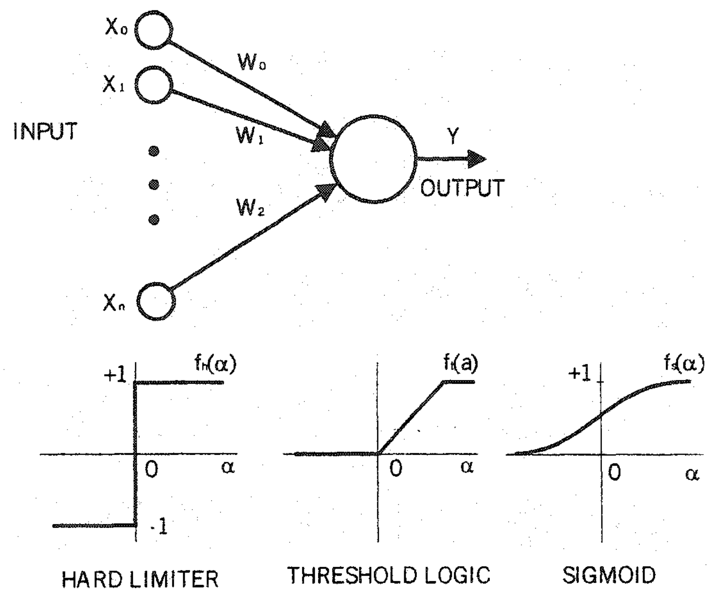

Figure 2. A Three Layered ANN Architecture

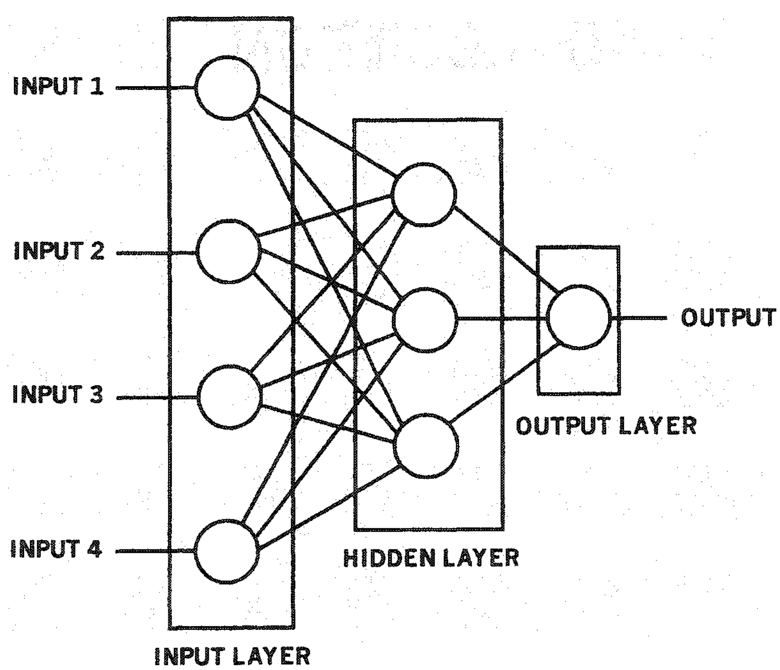

Table 1. Literature on Agent-based Artificial Financial Markets

Authors/Paper	Economic Model	Learning	Trading Mechanism
Arthur et al. (1997) & LeBaron et al. (1999)	R.E.	Classifier System	Walrasian Auction
Beltratti, Margarita (1996)	R.E.	ANN	Bilateral Trade
Arifovic (1996)	Cobweb	GA	Walrasian Auction
Chen, LeBaron, Lo & Poggio (2000)	No Model (Risk Neutral Agents)	Bayesian/nearest neighbour	Double Auction
Chen & Yeh (2002)	R.E.	GA/GP	Walrasian Auction

Note: R.E. is rational expectations model. ANN is artificial neural networks;

GA is genetic algorithm, GP is genetic programming.

Table 2. Trading Systems in the Real World

	Equity Markets					Bonds / FX markets		
	NYSE open	NYSE intra-day	Nasdaq	Paris Bourse	LSE	C→D	Direct ID	Broker ID
Order Book				X				X
Dealership	X	X	X		X	X	X	
Double Auction				X				X
Continuous		X	X	X	X	X	X	X
Open auction	X							
Centralized	X	X	X	X	X			X
Decentralized						X	X	
Pre-trade quotes		X	X	X	X			X
Post-trade quotes		X	X	X	X			

Note: C→D denotes customer-dealer trading sphere. Direct ID and Broker ID denote direct inter-dealer and brokered inter-dealer systems respectively.

Chapter 2

The Efficiency of Double Auction Markets with Neural Network Learning Agents

2.1 Introduction

In this chapter[2] we construct an artificial equity market with a double-auction trading mechanism and study its convergence properties.

Under open-outcry floor trading, the continuous double-auction has been the principle trading system in equity markets for more than 140 years. Technological advances have now introduced a new form of market trading system: the computerized double-auction mechanism (widely used in electronic trading). What are the features of this market? How do these features affect the price formation process? How efficient is this type of market system compared to those with alternative trading systems?

Experimental studies answer these questions by examining two different types of artificial markets. One uses a laboratory market where human subjects trade with each other and generate market dynamics. The other analyzes a computer simulated market where artificial traders replace human traders. The main weakness of the laboratory market is that some parameters are unobservable or uncontrollable, like a trader's risk preference, gen-

[2] A version of this chapter has been presented in CEF (1999) at Boston, the workshop at the University of Chicago on agent based simulation (1999). It is published later in *Evolutionary Computation in Economics and Finance*, Chen S. H. (ed.), Springer-Verlag (2002), Page 85-106.

eral learning behavior and trading strategies. In this sense, a computer simulated artificial market improves upon the laboratory market, since those parameters as well as the trading system can be carefully controlled and modified to test many different hypotheses. On the other hand, this computational tool involves new untested parameters.

Many laboratory market studies have been conducted under continuous double-auction. Early work such as Plott and Sunder (1982) and Forsythe, Palfrey and Plott (1982) introduced laboratory asset markets featuring markets with insider traders and uninformed traders and conclude that the equilibrium prices do reveal insider information. The lesson from the double-auction asset market experiments of Smith, Suchanek and Williams (1988) seem to be rather different. They report frequent large bubble episodes where transaction prices rise well above the Rational Expectations Equilibrium (REE) values for an extended time period, which usually ends in a sudden price crash to one that is close or below the fundamental value. Camerer and Weigelt (1992) studied a market where an asset pays a stochastic dividend each period. This is a more complicated market where token valuations are not specified exogenously by the experimenter but are determined endogenously and must be inferred by traders over the course of the trading process. They concluded that " the double-auction does not fail completely at generating convergence to competitive equilibrium prices for this stochastic asset, but the auction mechanism performs much more slowly and erratically than in simpler settings."

Computer simulated markets extend the experimental study by permitting manipulation of the agents' trading strategies and specifying a trading mechanism. After conducting many real trading experiments in laboratories, Gode and Sunder (1993) were interested in just how much "intelligence" was necessary to generate the results they were seeing. They ran a computer experiment with "Zero Intelligence" (ZI) agents who are almost completely random in their behavior. They used a double-auction market similar to those used in many laboratory experiments. Their results showed that the budget constraint is critical to achieving market efficiency. Further examples of a simple double-auction can be found in Rust, Miller and Palmer (1992). They focus on the performance of different trading strategies.

Experiments involving simulated markets frequently appear in the literature. Chan, LeBaron, Lo and Poggio (2001) construct a double-auction asset market experiment to investigate information dissemination and aggregation. In tests of information dissemination, they find prices that converge to fundamental value in almost all cases. But, in many cases, convergence is difficult to achieve in tests of information aggregation. To test the robustness of the results, technical traders and noise traders are introduced. The Santa Fe stock market (Arthur *et al.* 1997 and LeBaron *et al.* 1999) investigates market efficiency and price convergence with a rational expectations asset-pricing model, genetic algorithm learning, and Walrasian tatonnement as the market clearing mechanism. LeBaron (2002 forthcoming) provides a very useful survey of some portions of this literature.

In this chapter we construct an artificial equity market where agents trade a risky asset that pays a stochastic dividend each period. Artificial Neural Networks (ANN) assume the role of traders, who form their expectations about the future return and place orders based on these expectations. Market prices are set endogenously by trading among agents in a *double-auction* market.

This chapter addresses the market efficiency question by measuring the price convergence to a rational expectations equilibrium price. Market efficiency is defined such that the prices generated from a double auction market converge to rational expectations equilibrium prices. In particular, we consider the price deviation from rational expectations prices and trading volumes. In a rational expectations environment, the price deviation should decrease and trading volumes will tend to zero when the information is fully reflected in the prices. We conduct this study with three purposes in mind: (1) to simulate a computerized double-auction platform in order to help us better understand the price formation process in this trading structure and to test other hypotheses on market design and trading behavior; (2) to replace the human trader with the artificial neural network traders to observe whether they are able to learn to form rational expectations, and how they adjust their expectations within each period; (3) to compare this double-auction market outcome with theoretical benchmarks and with results from laboratory markets. It is worth mentioning that, in most of the laboratory market studies where double auctions are used, the "redemption value" (return) of the risky asset is exogenously set by the designers and the underlying distribution of this value is discrete and uniform . By contrast, in our simulated market, the market price, a major part of the return, is endogenously determined by the trading process. Dividends, the other part of the return, follows an exogenous and stochastic process with known structure and unknown parameters.

We find that market dynamics under double-auction converge to the REE in some experiments. This convergence, however, is sensitive to the deviation from rationality among the agents. In the experiments where we introduce noise trading, convergence becomes unattainable. Minimal rationality within the market is not sufficient to generate convergence in a double auction market when the market price is endogenous.

The remainder of this chapter is organized as follows. Section 2 presents the market structure. In this section, our basic asset pricing model, the agents' expectation formation, their trading strategies and a double-auction trading mechanism are described. Section 3 provides the design of all the experiments. The results of these computational experiments are presented in Section 4 where market simulation results are evaluated, tested and compared with the findings in the literature. Section 5 concludes and suggests future extensions to the work presented in this chapter.

2.2 Market Structure

2.2.1 Basic Framework

We set up a model of an asset market along the lines of Grossman and Stiglitz (1980) and Hussman (1992) that was used subsequently by Arthur *et al.* (1997). Consider a market with N traders, indexed by j. Traders are assumed to have constant absolute risk aversion (CARA) utility functions with a constant coefficient of risk aversion λ. Time is indexed by t and the horizon is indefinite. Traders decide on their desired asset composition between a risky stock and a risk-free bond. The risk free bond is in infinite supply and pays a constant interest rate r. Stock is issued in total units of $\overline{Q}$, and pays a stochastic dividend, D_t, which follows an AR(1) stochastic process as follow

$$D_t = \bar{d} + \rho D_{t-1} + \xi_t, \tag{7}$$

where $\xi_t \sim N(0, \sigma_\xi^2)$. The parameters of the dividend process are unknown.

Each trader j attempts to maximize next period's wealth, by optimizing the allocation between the risky asset and the risk-free asset as follows:

$$Max \;\; \hat{E}_t^j[-exp(-\lambda W_{t+1}^j)], \tag{8}$$

$$s.t \;\; W_{t+1}^j = (1+r)(W_t^j - Q_t^j P_t) + Q_t^j [P_{t+1} + D_{t+1}]; \tag{9}$$

where P_t is the stock price and $\hat{E}_t^j$denotes the expectation of trader j conditional on the set of information Ω_t available to j at time t, including historical information on market prices and dividend process. The parameters of the stochastic dividends must be learned and used to determine the expected utility.

Agents have a wealth constraint, borrowing cash and short sales are not allowed. Each agent has given initial holdings in stock and the sum of the total supply of the market is equal to $\overline{Q}$.

It is well known, as in Grossman and Stiglitz (1980) and Hussman (1992), that under CARA utility and Gaussian distributions for predictions, the agent's desired demand, Q_t^{j*}, for holding the risky asset is linear in the expected excess return:

$$Q_t^{j*} = \frac{\hat{E}_t^j[P_{t+1} + D_{t+1} - (1+r)P_t]}{\lambda \hat{\sigma}^2_{j, P_{t+1}+D_{t+1}}} \tag{10}$$

Risk aversion prevents traders from taking an infinite position based on expected return differentials. $\hat{\sigma}^2_{j, P_{t+1}+D_{t+1}}$ is the variance of the expected returns.

The market clearing price can be found by balancing aggregate demand with aggregate supply, so that:

$$\sum_{j=1}^{N} Q_t^{j*}(P_t) = \sum_{j=1}^{N} \frac{\hat{E}_t^j[P_{t+1} + D_{t+1} - (1+r)P_t]}{\lambda \hat{\sigma}^2_{j, P_{t+1}+D_{t+1}}} = \bar{Q} \tag{11}$$

Under *full information* and homogeneous expectations, i.e. where every trader observes each of the components in the dividend process and holds identical expectations, the index j can be omitted from the expectations operator.

Using equations (7) and (11), it is not difficult to solve for a homogeneous rational expectations equilibrium (REE) price under full information as

$$P_t = \frac{\rho}{1+r-\rho} D_t + \frac{1}{r}(1+f)[\bar{d} - \lambda(1+f)\sigma_\xi^2(\frac{\bar{Q}}{N})]. \tag{12}$$

where $f = \frac{\rho}{1+r-\rho}$. This is an important mapping which provides the theoretical values for REE prices. See Appendix B for details.

Finally, as a benchmark for evaluating the different equilibria in subsequent experimental setups, it is useful to find the analytical expression for the expected return in equilibrium. Given equation (11) and the direct relationship between prices and dividends in equation (12), the optimal forecasts in the full information rational expectations equilibrium is straightforward to find. One form of these forecasts is

$$E_t(P_{t+1} + D_{t+1}) = \rho(1+f)D_t + [(1+\frac{1}{r})(1+f)\bar{d} - \lambda\frac{1}{r}(1+f)\sigma_\xi^2(\frac{\bar{Q}}{N})] \tag{13}$$

This is the homogeneous rational expectations equilibrium (HREE) forecast we seek. This forecasting equation is one of the crucial items that traders need to estimate. Moreover, it is used as a benchmark result for our agents to learn. When the underlying components of the dividend process are not directly observable, the key question is how the next period's price and dividend expectations might be formed by each trader. We will consider this issue in the next section.

2.2.2 Heterogeneous Traders and Their Trading Strategies

We assume this market is populated by two types of traders: value traders (Artificial Neural Network traders) and momentum traders. Among them, the value traders follow the rational expectations model and try to learn the state of nature and the rational expectations forecasting given by equation (13) simultaneously. Their existence is expected to keep market prices close to the rational expectations equilibrium. Momentum traders are chartists and do not follow the rational expectations model. They act according to some technical indicators. The presence of this type of traders is to test the robustness of the results obtained when all the traders are value traders.

ANN traders with their trading strategies

For value traders, the parameters in equation (13) are all that they need to learn to compute their expectations. We assume value traders employ artificial neural networks (ANN)[3] to learn the parameters of the rational expectations equilibrium in equation (13).

Artificial neural networks provide a conceptually simple procedural model of how agents might learn to approximate their true but unknown conditional expectations function and hence form "boundedly rational" expectations. This approach will offer more desirable properties than traditional learning methods (say, least squares). It avoids any presumption regarding beliefs and requires no *ex ante* knowledge of the modeling structure. Moreover, it will better mirror actual cognitive processes, in which different individuals might well

[3] Smith (1993) is a good introductory book that concentrates on one type of neural network; namely, feed-forward neural networks, one of the most commonly used ANNs. Hertz, Krogh and Palmer (1991) provide a clear and concise description of the theoretical foundations of neural networks using a statistical mechanical framework.

"recognize" different patterns and arrive at different forecasts from the same market information.

Feedforward neural networks are employed by value traders with one input, one hidden layer (with three neurons) and one output (for more details, see Appendix 1 of this chapter). There are two reasons for choosing this relatively simple network design. First, the single hidden layer feed-forward network possesses the universal approximation property that it can approximate any nonlinear function to an arbitrary degree of accuracy with a suitable number of hidden units (White 1992). Second, since the rational expectations mapping in equation (13) is a simple linear one, a more complex architecture is unnecessary. The motivation for keeping one hidden layer is to provide value traders with the possible learning structure for out-of-equilibrium behavior. The homogeneous rational expectations equilibrium (HREE), represented by equation (13), relies on some strong assumptions such as homogeneous beliefs among agents, common *ex ante* knowledge of the distribution of the state variable, and a Walrasian tatonnement. In our market simulation, with all of those assumptions relaxed, value traders can expect to encounter some out-of-equilibrium behavior, especially when momentum traders participate in the market.

Each value trader possesses an ANN model which shares a similar structure. But there are differences in the initial values of the parameters. Traders train their model according to newly-revealed market information on returns at the end of each period and then use the updated model to forecast the next period's return.

We now consider the question of when a trader will choose to trade via a market order or limit order, or not to trade at all. The problem is structured as follows. Each trading period contains multiple trading rounds. Let the trader determine whether to rebalance his portfolio at the beginning of every trading round. To simplify the analysis, we assume that all orders are for a fixed number of shares($\triangle Q$), and that when any market or limit order is executed, it is satisfied fully at the stated price. We assume that an unfulfilled limit order is canceled prior to the next decision point.

At any of the time t decision points, trader j is faced with five possible options. Those options are indexed by i:

1. Post a limit buy-order at price $P^b_{j,t}$;
2. Post a limit sell-order at price $P^a_{j,t}$;
3. Post a market order to buy shares at the current best ask price, a;
4. Post a market order to sell shares at the current best bid price, b;
5. Do nothing.

The formal way for a trader to find the optimal order placement strategies to maximize his expected utility of wealth, $iMax\hat{E}^j_t(U_{j,i})$, for $i = \{1, 2, .., 5\}$,where $\hat{E}^j_t(U_{j,i})$ is the expected utility derived from option i. In what follows, we use a simpler way to approximate traders' optimal trading strategies.

In order to choose his action, an ANN trader will find his reservation price first. Since at this price his optimal holding is equal to his current holdings, he has no incentive to trade.

A risk-averse trader finds his optimal demand by maximizing his utility in terms of next period's wealth. His reservation price can be solved from the optimal demand function given by equation (10). Let $Q_t^{j*} = Q_t^j$, where Q_t^j is the current holding and Q_t^{j*} is the desired holding in the current period. Then the reservation price for trader j is

$$P^{R,j} = \frac{\hat{E}_t^j(P_{t+1} + D_{t+1}) - \lambda\hat{\sigma}^2_{j,P_{t+1}+D_{t+1}} Q_t^j}{1+r} \tag{14}$$

This reservation price is negatively related to his current holding level, and positively related to the expectation value of the next period's return.

The trader will never buy higher or sell lower than this reservation price. Let a denote the best ask, and b the best bid ($a > b$) in the current trading round. The value trader will

1. Post a market buy order, if $a < P^R$;
2. Post a market sell order, if $b > P^R$;
3. Do nothing, when a or b is equal to P^R;
4. Otherwise, post a limit order to buy or sell at the price

$$P_{j,t} = P^{R,j} - S^j B \tag{15}$$

Equation (15) is a mechanical way to decide the quoting price which should be done ideally via optimization. B is an indicator variable where $B = +1$ for an ask order and $B = -1$ for a bid order. S^j is a spread between the reservation price and the price quoted. It is a linear function[4] of order size $\triangle Q$. The order size is fixed to one unit in this study.

[4] It is important to keep in mind that the spread should also be a function of time. The closer to the end of a trading day, the higher the bid price should be to reflect the probability of the order execution. This aspect of the market will be studied in the future.

Since there is multiple trading rounds in each trading period, this fixed order size will not be a constraint for a trader who wants to trade more than one unit. Meanwhile, S^j is different across traders since they have different forecasting variances,

$$S^j = (\frac{\lambda \hat{\sigma}^2_{j,P_{t+1}+D_{t+1}}}{1+r})\triangle Q \tag{16}$$

The reservation price here serves as a mean-reverting device for the market. When traders hold very similar expectations, the differences between their reservation prices come from the difference between their current holding in the risky asset. The low-reservation-price trader who has a large position posts a low bid and ask price; while the high-reservation-price trader who has a small position posts a high bid and ask price. As a result, the low-reservation-price trader could only become a seller, since his ask order will have a high probability of being executed. Therefore, he can reduce his position and keep selling until it reaches his optimal holding.

Momentum traders and their trading strategies

Momentum traders are chartists who believe that future price movements can be determined by examining patterns in past price movements represented by various moving averages (MAs). Brown *et al.* (1998) support the notion that price history can be used to predict market movements. The moving average trading rule states that when the short-term (usually 1–to–5 day) moving average price is greater than the long-term moving average (usually more than 50 days), a rising market is indicated. Thus, this trading rule would generate a buy signal. Based on such market trends, the momentum trader decides to enter or exit the market.

In this chapter, momentum traders are divided into two groups according to their choice of trading rules. The first group of momentum traders compares the current market price P_t with the average of the past 5 days denoted as MA(5) and responds as follows,

- If P_t >MA(5), they buy $\triangle Q$ shares.
- If P_t =MA(5), they hold current position.
- If P_t <MA(5), they sell $\triangle Q$ shares.

The second group of momentum traders identifies a trading opportunity by comparing MA(5) with the average of past 10 days denoted as MA(10). Specifically:

- If MA(5)>MA(10), they buy $\triangle Q$ shares.
- If MA(5)=MA(10), they hold their current position.
- If MA(5)<MA(10), they sell $\triangle Q$ shares.

2.2.3 The Trading Institution: Double-auction

We chose the double-auction mechanism for two reasons. First, major stock, commodity and many other markets are organized as double auctions. A classic example of a double-auction market is the NYSE. Second, laboratory double auctions with human traders are known to yield data that approximate competitive equilibrium in a variety of environments[5].

---

[5] It is worth mentioning that most of the laboratory double-auction markets set the redemption values of the risky asset exogenously. When these values become endogenous, as Friedman and Rust (1992) point out: 'we

Any trading mechanism can be viewed as a type of trading game in which the players meet at some venue and act according to some rules. These rules dictate who can trade, when and how orders can be submitted, who may see or handle the orders, how orders are processed, and how prices are eventually set. It has been recognized that these rules are very important because they can affect bidding incentives and, therefore, the terms and the efficiency of an exchange.

Rational expectations equilibrium (REE) rely on a Walrasian tatonnement auction which has long served the needs of general equilibrium price theory. In this type of auction, the time at which the orders arrive is not relevant. It is a path-independent process. In this auction, the Walrasian auctioneer, who can observe all the traders' demand functions, adjusts prices in various markets to equilibrate aggregate supply and aggregate demand, which reflects a *competitive equilibrium (CE)*. The key feature of these processes is that agents are not permitted to trade "out of equilibrium" until the adjustment process has determined a CE price vector.

By contrast, in a double auction (known by experimentalists as the *continuous double-auction* and by practitioners as an *open-outcry market*) used in this paper, the price is path-dependent. The time at which orders arrive is relevant. In a double-auction (DA), transaction prices arise from a two-sided auction where buyers improve (raise) bids and sellers improve (lower) asks until one of the buyers and one of the sellers reach agreement.

---

start to see how the nice properties of DA markets can start to break down in more complicated enviroments where token valuations are not specified exogenously by the experimenter but are determined endogenously and must be inferred by traders over the course of the trading process.'

In our simplified double-auction market, which is similar to the one in Chan, LeBaron, Lo and Poggio (2001), traders can either submit a bid (ask), or accept an existing bid (ask). If a bid (ask) exists for the stock, any subsequent bid (ask) could be higher or lower than the current one, but only the best bid (ask) price can be observed by traders. There is no role for the market maker in this setup. A transaction occurs when an existing bid (ask) is accepted. Given the absence of opportunity to borrow or sell short, traders must trade subject to their budget constraints. A trader leaves the market when either the market price equals his reservation price or he runs out of money.

Each trading period contains 40 trading rounds. At the beginning of each trading round, we assume a random permutation of the traders, which determines the order in which traders enter the market. Initially, the value traders come to the market with their own reservation prices and attempt to post or accept a bid (ask) order by comparing their reservation prices with the existing best ask (best bid). They can only observe the best bid and the best ask price. Momentum traders go to the market with their "buy", "sell" or "hold" signals and submit market orders according to these signals. At the end of each trading period, the last transaction price, or *closing price*, is recorded as the market price of this period. Agents trade with each other according to the following scenarios:

- Scenario 1. If the best bid, b, and the best ask, a ($a > b$), exist on the market
  - If $P^{R,j} > a$, he will post a market order, buy at this ask price;
  - If $P^{R,j} < b$, he will post a market order, sell at this bid price;

- If $b < P^{R,j} < a$ and $P^{R,j} < (a+b)/2$, he will post a sell order at a price of $(P^{R,j} + S^j)$;
- If $b < P^{R,j} < a$ and $P^{R,j} > (a+b)/2$, he will post a buy order at a price of $(P^{R,j} - S^j)$.

- Scenario 2. If only the best ask, a, exists
  - If $P^{R,j} > a$, he will post a market order, and buy at this ask price;
  - If $P^{R,j} < a$, he will post a buy order at a price of $(P^{R,j} - S^j)$.
- Scenario 3. If only the best bid, b, exists
  - If $P^{R,j} < b$, he will post a market order, sell at this bid price;
  - If $P^{R,j} > a$, he will post a sell order at a price of $(P^{R,j} + S^j)$.
- Scenario 4. If no bid or ask exist,
  - with equal probability he will post a buy or a sell order at price of $(P^{R,j} - S^j)$ or $(P^{R,j} + S^j)$ respectively.

2.3 Experimental Design

The rational expectations equilibrium price, given by equation (12), will be used as our benchmark for the four experiments. The traders are assumed to work in the following environment and the values of parameters follow those used in the Santa Fe artificial stock market in Arthur *et al.* (1997).

1. The risk-free interest rate, r, is fixed to be 0.1. The risk aversion parameter, λ, is set to be 0.5.

2. Throughout the experiments, we specify the dividend process as equation (7) where $\bar{d} = 0.5$, $\rho = 0.7$, and $\sigma_\xi^2 = 0.25$. Notice that based on these parameters we can calculate the theoretical REE prices in equation (12).

3. Traders are assumed to hold an endowment of 2 share and 2000 dollars in cash.

4. Value traders are assumed to be equipped with an ANN learning mechanism. The neural network is programmed to learn the value of the parameters in the rational expectations equation. The initial starting values of weights for all the hidden nodes are randomly drawn from a uniform distribution between [-1 1].

5. Momentum traders are initially endowed with "hold" signals until they get enough information to form their trading rules. Moving averages of the stock prices are calculated with equal-weighted parameters. Momentum traders using MA(5) or MA(10) pricing rules wait for 5 or 10 periods respectively to enter the market. They enter the market with their "buy" or "sell" signals generated from their trading rules.

6. The minimum unit of time corresponds to a trading round. Each trading period contains 40 trading rounds. Each trading round begins with a permutation of the

traders to determine the order of the participation in the market. The trading round is completed when each trader is considered in turn.

7. The market is run for 1000 trading periods.

8. Each of the four experiments to be described below is repeated 25 times.

- *Experiment 0* 10 value traders in a Walrasian auction

  The purpose of this experiment is to check whether the value traders, using their ANN models, are able to learn the rational expectations equilibrium. The rational expectations equilibrium is a mapping between next period's return and the lagged dividend which is given by equation (13).

- *Experiment 1* 10 experienced value traders in a double-auction market

  The purpose of this experiment is to test the convergence of prices to the REE with a double-auction trading institution. Value traders in this experiment are trained in a market set up for 200 periods before they enter the market.

- *Experiment 2* 10 inexperienced value traders in a double-auction market setup.

  The value traders in this experiment are not trained by the presample data before they enter the market. This experiment is designed to check the effect of traders' experience on the convergence property.

- *Experiment 3* 10 value traders, 10 momentum traders, among them, 5 traders use MA(5) and 5 traders use MA(10).

The purpose of this experiment is to check the sensitivity of the convergence on this double-auction market to deviations from rationality. The presence of momentum traders should add some noise to the market. The rationale for this experiment is to determine whether the value traders can maintain rational expectations equilibrium given the existence of some erroneous signals caused by the presence of "irrational" momentum traders. We call the prices generated from these artificial markets as bounded rational expectations (BRE) prices. Next section reports the time series properties of BRE prices compared with REE prices.

2.4 Computational Results

In this section, we compare the time series properties of simulated data, generated from four bounded rational expectations environments (BRE), with the rational expectations equilibrium prices. rational expectations prices are simulated using equation (12). Prices in the bounded rational environment of the model are generated from the artificial markets with both the Walrasian and double-auction trading systems. The efficiency of a market in this paper is measured by how close and how fast the prices converge to REE. The variables we consider are price deviations and trading volume. Diminishing trading volume and price deviation suggest that the market is approaching its equilibrium.

In general, ANN traders are able to learn rational expectations parameters collectively. Convergence to the REE occurs rapidly in Experiment 1, the case of experienced traders and identical trading strategies. Such convergence occurs more slowly in experiment 2, where value agents are not trained before they enter the market. In experiment 3, with the presence of momentum traders, convergence is unattainable. The reason is that momentum traders only obey their own trading rules (e.g. MA(5) or MA(10)), which are not consistent with rational expectations. They disregard the views of value traders, and value traders are not able to learn those technical trading rules either. The value traders may not even be aware of the presence of momentum traders. In other words, the fundamental value of our stock is not common knowledge.

2.4.1 General features of each experiment

Before we describe the results of various tests conducted on the simulated data with respect to the time series properties, we give some flavor of each experiment.

In order to make valid comparisons, the same draws of the random errors, $\{\xi_t\}$in the dividend process equation (7), are used to generate REE prices and BRE prices. Given the draws of the random errors and the values of exogenous parameters, rational expectations prices are simulated using equation (12). The rational expectations prices have a mean of 16.61 over 25 repetitions of the random draws.

Experiment 0: In order to isolate the effect of the trading system and the learning behavior, we construct experiment 0 where 10 ANN agents submit their bids simultaneously to a Walrasian auctioneer. The auctioneer calculates the market price where aggregated demand equals supply. As is seen from Fig.1, the price series appears to be nearly identical to the price in the rational expectations regime. Figure 2 plots ANN expectations and rational expectations. The difference between these two are fairly small. In a Walrasian environment, value traders can recover the REE very well.

The first two panels of Table 3 report the means, standard deviations of the market prices and trading volumes in this experiment during sample periods of 1-250, 251-500, 501-750 and 750-1000. Agents on this market learn the rational expectations function in equation (13) very quickly. We observe the reservation price of each agent converges to the rational expectations price and the bid-ask spreads get narrower due to diminishing uncertainty. Equation (16) indicates that the size of the spread is a function of the uncertainty embedded in the agents' expectations. When the variance of their expectations decreases, the spread gets smaller. As a result, the means of the market prices, 16.59, 16.58, 16.52 and 16.64 for each sample period, respectively, are close to the mean of the REE prices. Meanwhile, since market price reflects all the information and agents hold the same expectations, the trading volume diminishes rapidly, and after certain trading periods, the trading volume tends to zero.

Experiment 1: We introduce a more realistic double-auction to replace the Walrasian auctioneer, while maintaining the homogeneity of the agents. The market participants are purely value traders. One noticeable feature of this experiment is that we give agents time to train themselves before we start to record the data. We call those agents experienced traders. They are identical in terms of their information set, information processing ability, but heterogeneous in terms of their network parameters. In this experiment, after a certain period of learning, the homogeneous rational expectations are reached. As shown in Figure 3, the deviations of market prices from REE are bounded and close to zero. Figure 4 shows the realized BRE prices and trading volumes. In the lower part of the graph, we see that the average trading volume approaches zero after roughly 100 periods. Market prices reflect all the pay-off related information, and trading ceases. This result confirms Tirole's famous "No-trade Theorem" in Tirole (1982).

The first panel of Table 4 shows the mean of the market prices and volume in this market. After certain periods, the mean of the market price tends to the one of REE prices and trading volume diminishes. The second panel of Table 4 reports the standard deviation of BRE prices from RRE prices. The standard deviations in general are larger than the ones in experiment 0, but they decrease quickly. It suggests that even though the double-auction trading system causes some friction, value traders are still able to learn the REE prices after acquiring experience.

Experiment 2: This experiment has a set-up identical to that of experiment 1 except that the value traders in this experiment are not given time to train their neural network model before they enter the market. In the early period of this experiment, the market shows relatively high price deviations. Compared with the first experiment, convergence is slower and the price deviation is bounded in a wider range. Figure 5 plots the price deviations from REE prices for both experiment 1 and experiment 2. It is clear that the deviation of the market prices in experiment 2 is more volatile than the ones in the experiment 1.

The first two panels of Table 5 show that market prices approach the REE price slowly and the variance of the market prices decreases in the later sample period. Overall the value traders do learn the REE prices but the learning takes a longer time. The standard deviations in the second panel stay larger than the ones in Table 4 in experiment 1. The reason for this could be that the ANNs in this market are not pre-trained by any market data. The difference between experiment 1 and 2 can be interpreted as the difference between experienced traders and inexperienced traders. In the inexperienced traders case, the convergence to REE is slow. This result is consistent with Camerer and Weighelt (1992). This result can be related to the difference in price behavior for new public issues of stocks and long outstanding stocks. For new public issues of stocks, there is no historical data available to train traders. In this case, we observe high volatility in market prices compared with experiment 1.

Experiment 3: This experiment adds another layer of complexity by allowing heterogeneity among market participants. In this experiment, we introduce momentum traders who obey only technical trading rules and do not use pay-off relevant information about dividends. In experiment 3, with the presence of irrational momentum traders, convergence appears unattainable. Figure 6 shows the market prices, and the price deviations from REE prices. There is no tendency for the price deviations to diminish with time. The figure shows that the value traders have difficulty learning the REE prices in the presence of the noise traders.

The first two panels of Table 6 present the mean and the standard deviation of market prices and trading volume in experiment 3. In this experiment, the variance in the market price hardly decreases over the periods and the trading volume persists through all the sample periods. These results suggest that the value traders have difficulty learning in the presence of another type of trader. This result is in contrast to the findings in Gode and Sunder (1993) "zero intelligence" (ZI) traders case, but it is consistent with the results in

Camerer and Weighelt (1992)'s laboratory market. The difference in the results is explained by the market setting. In the ZI case, the return of the risky asset ("redemption value") is exogenous and set by the designer. In our experiments, this end-of-period return is determined *endogenously* and must be inferred by traders over the course of the trading process. In this case, convergence is more difficult to achieve than in the ZI case.

Correlations: Now we take a closer look at each experiment through the correlations between BRE prices, and REE price as well as the correlations between price changes and trading volumes.

In non-parametric learning, the correlation between REE and BRE prices should increase over time. Correlation coefficients are presented in Tables 3-6 for the four sample periods:1-250, 251-500, 501-750 and 751-1000. Average correlations over the 1000 periods between REE and BRE prices should increase over time as the learning algorithm of the bounded rational environment converges to the REE. This result may not hold in an irrational environment. Correlation coefficients are presented in Tables 3-6 in the third panel. The REE prices sequence is computed using equation (12). BRE prices are endogenously determined by the conditions of the experiment. For each simulation, the correlation between the two price sequences is computed for each period. The tables report the average correlation over the 25 repetitions.

For the first and the second experiments, we find similar evidence that REE and BRE prices have a higher correlation coefficient in the later periods of the simulation. The same pattern fails to hold for experiment 3. For experiment 3, the average correlations are 0.1565, 0.1760, 0.1762 and 0.1912 (Table 6). This experiment suggests that the value traders have difficulty learning the REE prices. As a result, the BRE prices show small correlations with the REE prices and the correlations hardly change in the later sample periods.

Andersen (1996) among other studies find evidence of a positive relationship between the volatility of price changes and trading volumes in stock markets. The correlation between the size of price changes and trading volumes in the boundedly rational environment will indicate if environments conform to the empirical evidence.

Our results reported in Tables 3-6 in the third panel, second column, are qualitatively similar across all experimental settings in the ANN learning environment. First, correlations between absolute changes in BRE prices and trading volumes are all positive. Second, correlation coefficients decrease over the sample. For example, in the experiment with the presence of homogenous experienced value traders, the correlation coefficients are: 0.3411, 0.2732, 0.1742 and 0.083 for the four sample periods respectively (Table 4). In the last experiment, with the presence of noise traders, correlations are nearly constant throughout the entire sample (Table 6): 0.5424, 0.5044, 0.5321 and 0.4867. Agents are not able to fully extract information from endogenous prices, and actively trade on the basis of expectations differences.

2.4.2 Time series features of BRE prices

In the previous section we discussed the general features of each experiment. In this section we present some statistical tests on the price series of the markets. Under the null hypothesis of a linear homogeneous REE, price and dividend should be a linear function of a lagged dividend. This is clear from the dividend process in equation (7) and the dividend to price mapping (12). To check this, the market price and dividend is regressed on a lag of the dividend and a constant term as indicated in equation (13):

$$P_{t+1} + D_{t+1} = c_0 + c_1 D_t + \varepsilon_t. \tag{17}$$

In the homogeneous REE this estimated residual series ε_t should be independent and identically distributed, N(0, 1.89). (See Appendix 2). This regression is estimated for experiments 1, 2 and 3. Results are given in Table 7.

The first column of Table 7 shows the variance of the residuals from all three experiments. All three cases show a higher variability than 1.89 but to different degrees. The first two experiments are very close. The next column gives the excess kurtosis, which should be zero under normal distribution. In the three experiments, the third one shows a significant amount of excess kurtosis. The third column presents the autocorrelation in the residuals. In the first two experiments, the autocorrelation are small and are close to zero. The residuals appear to be correlated in the third experiment.

In general, the case of heterogeneous preferences and trading strategies generates a market that exhibits richer "psychology". It is consistent with the fast learning case in the Santa Fe artificial stock market of Arthur *et al.* (1997). It is not consistent, however, with the attainment of a REE.

The results in experiment 3 appear to be different from those in experiment 1. In the third experiment, a large proportion of the variance in the prices and dividends remains to be explained.

What are other potential explanatory variables? To answer this question, two other technical indicators used by momentum traders are added to the simple linear regression of price and dividend on the lagged dividend,

$$P_{t+1} + D_{t+1} = c_0 + c_1 D_t + c_2 I_{t,MA(5)} + c_3 I_{t,MA(10)} + \xi_{t+1}. \tag{18}$$

Here the first indicator variable is a 5-period moving average indicator, which is equals to one when price is above the 5-period moving average and zero when price is below that level. The second indicator variable represents whether the 5-period moving average price is above the 10-period moving average. Table 8 shows the average of regression results for equations (17) and (18) for only experiment 3. The numbers presented in the Table 8 are the average of the estimated parameters and average R^2 values. The standard deviation over the 25 repetitions are given in parenthesis. The results show that both technical indicators add significant predictability. The parameters are small but statistically significant. The R^2 values in the last column confirms the fact that adding technical trading indicators explains a higher proportion of the variance in price and dividend.

It is obvious that in the presence of momentum traders, ANN traders cannot drive the market price to the rational expectations equilibrium. They cannot pick up the part of the variation which can only be explained by the technical indicators. To understand this question more clearly, we need to look at the market microstructure aspect in terms of order flow.

A concrete example will help. In the experiment with only value traders, the individual reservation price serve as a mean-reverting device to drive the market price to equilibrium by adjusting their positions. Consider a two-trader case. Suppose for certain periods, Trader 1 keeps bidding high and buying stocks from Trader 2 and this buying pressure drives up the market price. As a result, Trader 1 has a large stock holding and hence a relatively low reservation price. Trader 2, however, has a higher reservation price. Since agents will not buy above or sell below these prices, it is impossible for Trader 1 to buy shares from Trader 2, since the limit buy order that Trader 1 posted is always below Trader 2's reservation price. Therefore, the low-reservation-price trader could only possibly be a seller, by which he could reduce his holding level until it reaches the optimal holding. This selling pressure will drive the market price to its equilibrium. With the presence of the momentum traders, however, the link between the reservation price and the market price is broken. Since even in the case where Trader1 posts a low bid quote, it is still very possible for a momentum trader to "sell at the market", since they are not rational traders. As a result, a high-position trader still keeps increasing his holding level, and this will drive the market price a long way away from its equilibrium level. Convergence to the REE fails. In this case, minimal rationality seems insufficient to produce convergence to REE prices.

To check whether the large divergence in experiment 3 originates in the trading institution, it would be interesting to replace the double-auction in this experiment with a Walrasian auction, and then compare the market dynamics from these two different trading institutions. Unfortunately, we cannot perform this test since the 10 momentum traders have no demand functions to submit to the Walrasian auctioneer. In this case, it is impossible for the Walrasian auctioneer to find a market clearing price since the link between the individual demand and aggregate demand is broken.

2.5 Conclusions and Future Research

Our artificial double-auction market is capable of generating behavior close to REE under certain circumstances. The speed of convergence varies across each trader's experience. When irrational traders emerge, the market is driven by more noise factors other than just fundamental trading. The convergence is unattainable in this case. The value trader cannot learn the chartists' strategies: they may not even realize the existence of the chartists. But it is still not clear whether the observed market inefficiencies should be attributed to the learning effects, or to other differences in trading institution design.

When the intrinsic value of the risky asset becomes endogenous, the convergence is sensitive to deviations from rationality and minimum rationality. As we saw in experiment 3, it is not sufficient to converge to REE.

Some interesting hypotheses that can be tested on this market include informational efficiency, the role of market transparency and market microstructure. Another interesting extension to this initial work is related to the choice of market mechanism. In this chapter, we have used a double-auction framework. It would be appealing to simulate different types of auctions or to endow some participants with more market power to observe whether and how they could corner a market.

2.A Appendix 1

We assume each value trader possess a ANN(1-3-1) model. They have similar structures, but there are differences in the initial values of the parameters.

Five steps are taken to form the prediction of next period's return and risk.

1. Initialize a network with K hidden nodes ($K = 3$), J inputs ($J = 1$) and one output. The parameters start with initial values drawn

from a uniform distribution whose range is set at $[-1, 1]$.

2. Forecast with the current values of the parameters. An input at time t is presented to the network and forwarded through the network as follows:

$$\widehat{Y}_t = h(\sum_{k=1}^{K} \alpha_k f(\beta_k' X + \gamma_k) + \alpha_0), \tag{19}$$

where $\beta_k = [\beta_{k1}, \beta_{k2}, ..., \beta_{kJ}]'$, $X = [X_1, X_2, ..., X_J]'$ and $\gamma = [\gamma_1, \gamma_2, ..., \gamma_K]'$; $f(\cdot)$ is the logistic function, $f(u) = [1 + \exp(u)]^{-1}$, and h is a linear function, $h(s) = s$. The signals, X, here containing only one element, the dividend, are transmitted to the hidden neurons by a logistic function. The outputs of the hidden units is then weighted again, and are then summed and transformed by $h(\cdot)$. At the beginning of each period, we include the most recent dividend in the input layer. Each trader invokes his own network to forecast the return in the next period.

3. Compute the forecasting error. At the end of each trading period, when the market price and dividend are revealed, the sum of squared deviations between the output of the network and the target are computed as:

$$ERR_t = \frac{1}{2}\sum_{t=1}^{t}(Y_t - (\widehat{Y_t}))^2 \tag{20}$$

4. Estimate the parameters. Given an input and output pair X_t, Y_t, the estimation of all the parameters in this network is done by minimizing the sum of the squared errors above. The optimization is done by a conjugate gradient algorithm.

5. Update the forecasting errors. At the end of each trading period, agents update their estimated conditional variances according to an exponentially-weighted average of squared forecasting errors,

$$\sigma^2_{j,P_{t+1}+D_{t+1}} = \theta(P_t + D_t - \widehat{E}_{t-1}(P_t + D_t))^2 + (1-\theta)\sigma^2_{j,P_t+D_t} \tag{21}$$

This variance will be used in traders' demand function (10), the reservation price function (14) and the spread function (16). We keep these values fixed inside each trading period, updating them according to equation (21) at each invocation of ANN. In our experiment, θ is fixed to be 0.013. (This is the value choose by the SFI artificial stock market.)

2.B Appendix 2

We calculate the homogeneous REE for the case where we assume market price is a linear function of the dividend

$$P_t = fD_t + g, \tag{22}$$

We then calculate f and g from the market conditions at equilibrium.

Under *full information* and homogeneous expectations, Eq. (11) can be rewritten as

$$\frac{\hat{E}_t[P_{t+1} + D_{t+1} - (1+r)P_t]}{\lambda\hat{\sigma}^2_{P_{t+1}+D_{t+1}}} = \frac{\bar{Q}}{N} \tag{23}$$

by omitting the index j from the expectation operator.

From the dividend process in Eq. (7) and the linear assumption in Eq. (22), we can calculate the variance of this forecast as

$$\hat{\sigma}^2_{P+D} = (1+f)^2\sigma^2_\xi, \tag{24}$$

where σ^2_ξ is the variance of the error term in dividend process.

Substituting the dividend process in Eq. (7) and the linear assumption in Eq. (22) into Eq. (23), we can calculate the homogeneous REE forecast as

$$\hat{E}_t[P_{t+1} + D_{t+1}] = (1+f)[(1-\rho)\bar{d} + \rho d_t] + g \tag{25}$$

where

$$\begin{aligned} f &= \frac{\rho}{1+r-\rho} \\ g &= \frac{1}{r}(1+f)[\bar{d} - \lambda(1+f)\sigma^2_\xi(\frac{\bar{Q}}{N})]. \end{aligned}$$

Rearranging this equation we have the expression in Eq. (13), which is the homogeneous REE forecast that we seek.

Substituting the expression g into the price function in Eq. (22), we have the REE price shown in Eq. (12).

Table 3. Summary Statistics for Experiment 0

	Market Price	Trading Volume
Mean:		
Periods 1-250	16.59	1.2680
Periods 251-500	16.58	0.0000
Periods 501-750	16.52	0.0000
Periods 751-1000	16.61	0.0000
Standard deviation:		
Periods 1-250	0.8680	0.2680
Periods 251-500	0.8437	0.0000
Periods 501-750	0.8221	0.0000
Periods 751-1000	0.8259	0.0000
Correlation:	REE and BRE prices	Prices changes and volume
Periods 1-250	0.5223	0.40
Periods 251-500	0.9211	NA
Periods 501-750	0.9321	NA
Periods 751-1000	0.9389	NA

Note: The numbers reported in this table are the average over the 25 repetitions of the same experiment.

Table 4. Summary Statistics for Experiment 1

	Market Price	Trading Volume
Mean:		
Periods 1-250	17.96	3.7813
Periods 251-500	16.87	0.0000
Periods 501-750	16.84	0.0000
Periods 751-1000	16.26	0.0000
Standard deviation:		
Periods 1-250	1.6880	0.6280
Periods 251-500	1.4387	0.0000
Periods 501-750	1.1421	0.0000
Periods 751-1000	0.8925	0.0000
Correlation:	REE and BRE	Prices changes and volume
Periods 1-250	0.4565	0.3411
Periods 251-500	0.6756	NA
Periods 501-750	0.8762	NA
Periods 751-1000	0.9211	NA

Note: The numbers reported in this table are the average over the 25 repetitions of the same experiment.

Table 5. Summary Statistics for Experiment 2

	Market Price	Trading Volume
Mean:		
Periods 1-250	17.46	4.7318
Periods 251-500	17.11	3.1387
Periods 501-750	16.87	2.1055
Periods 751-1000	16.82	2.1631
Standard deviation:		
Periods 1-250	1.8086	0.7801
Periods 251-500	1.8342	0.6752
Periods 501-750	1.5210	0.5091
Periods 751-1000	1.2109	0.4210
Correlation:	REE and BRE	Prices changes and volume
Periods 1-250	0.3125	0.4412
Periods 251-500	0.4761	0.3109
Periods 501-750	0.6760	0.2770
Periods 751-1000	0.8211	0.1838

Note: The numbers reported in this table are the average over the 25 repetitions of the same experiment.

Table 6. Summary Statistics for Experiment 3

	Market Price	Trading Volume
Mean:		
Periods 1-250	20.12	3.8913
Periods 251-500	22.61	2.8907
Periods 501-750	26.21	2.8052
Periods 751-1000	28.43	2.7011
Standard deviation:		
Periods 1-250	4.6829	0.7208
Periods 251-500	3.5607	0.6741
Periods 501-750	4.4211	0.4221
Periods 751-1000	4.1225	0.4211
Correlation:	REE and BRE	Prices changes and volume
Periods 1-250	0.1565	0.5424
Periods 251-500	0.1760	0.5044
Periods 501-750	0.1762	0.5321
Periods 751-1000	0.1912	0.4867

Note: The numbers reported in this table are the average over the 25 repetitions of the same experiment.

Table 7. Basic Summary Statistics for the Residuals and Trading Volumes

Experiment	Variance	Kurtosis	ρ	Trading Volume
Experiment 1	1.91	0.33	0.036	0.25
	(0.12)	(0.31)	(0.007)	(0.03)
Experiment 2	2.93	0.71	0.078	0.96
	(0.65)	(0.56)	(0.012)	(0.05)
Experiment 3	4.23	4.76	0.351	2.42
	(0.97)	(1.01)	(0.098)	(0.11)

Note: The experiments are run for 1000 periods and 25 times for each. The statistics shown in this table are the average over 25 runs. Numbers in parenthesis are standard errors estimated using the 25 runs. The kurtosis reported is excess kurtosis. The trading volume is the average value for each period over all trading periods.

Table 8. Regressions on Fundamental and Technical Indicators

Regression equation	constant	D	$I_{MA(5)}$	$I_{MA(10)}$	R^2
Eq.(11)	4.65	1.85	NA	NA	0.43
	(0.08)	(0.54)			(0.05)
Eq.(12)	4.21	2.02	0.054	0.086	0.57
	(0.13)	(0.87)	(0.023)	(0.034)	(0.08)

Note: The numbers in the parenthesis are the standard errors estimated in the 25 runs.

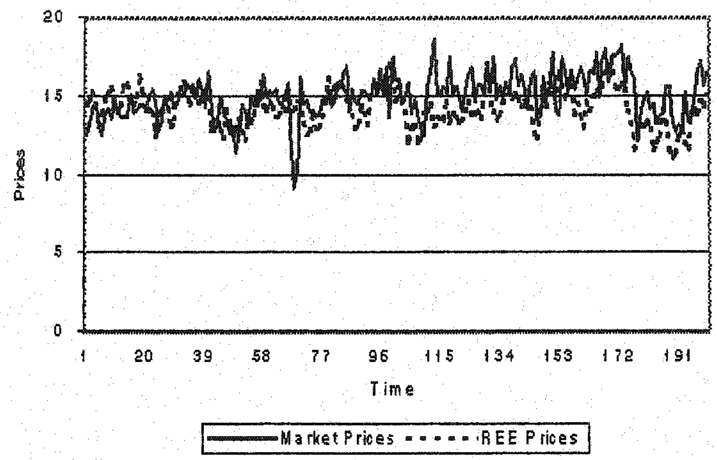

Figure 1. Experiment 0: Market Prices vs. REE Prices

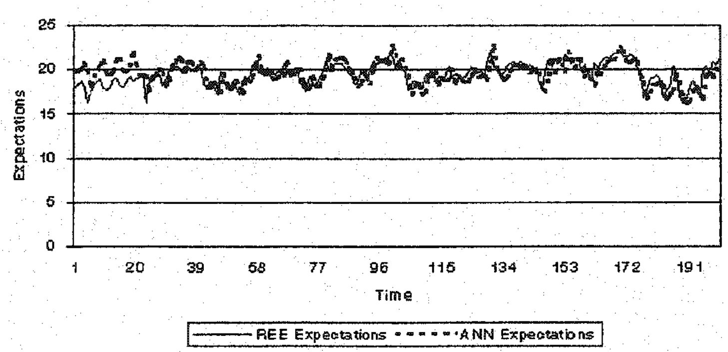

Figure 2. Experiment 0: Rational vs. ANN Expectations

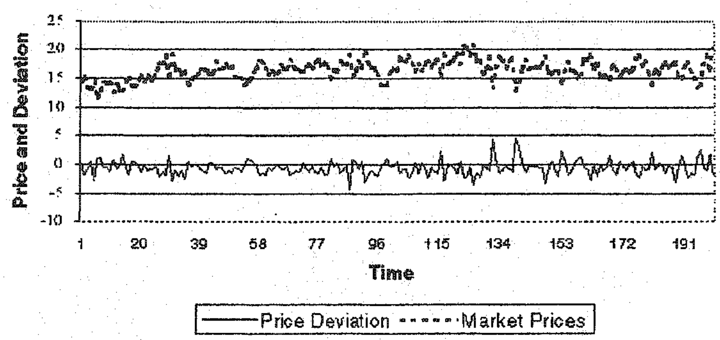

Figure 3. Experiment1: Market Prices vs. Price Deviations

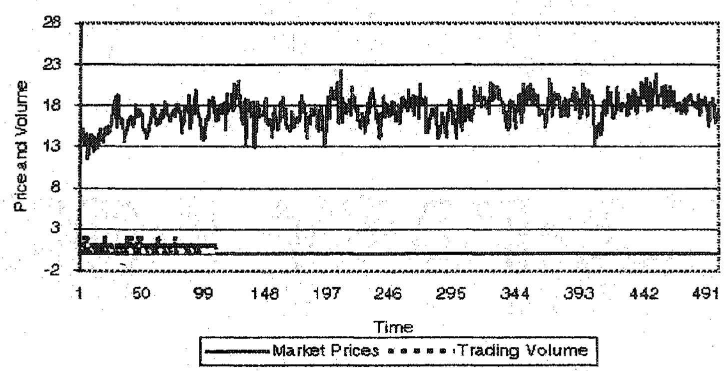

Figure 4. Experiment 1: Market Prices vs. Trading Volume

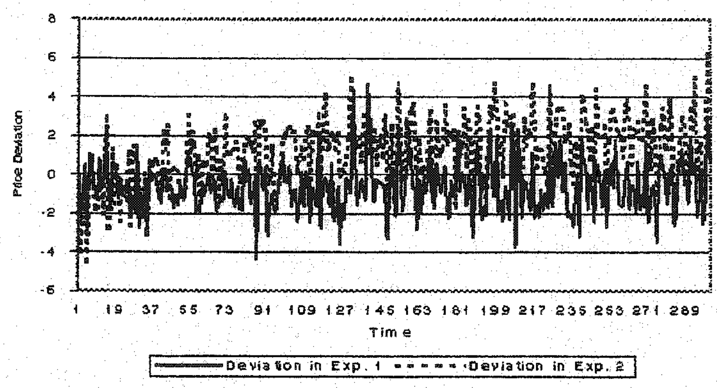

Figure 5. Price Deviations in Experiment 1 and Experiment 2

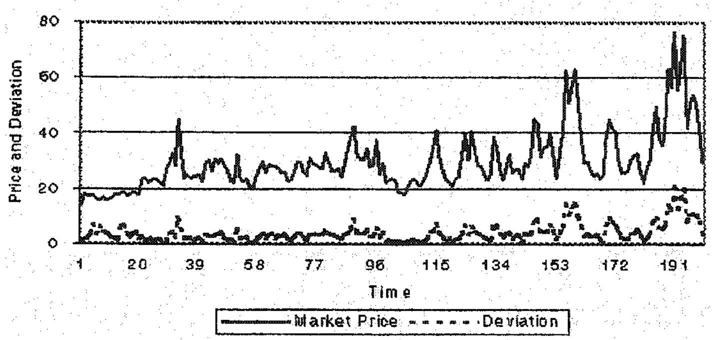

Figure 6. Experiment 3: Market Prices vs Price Deviations

Chapter 3

Endogenous Entry and Inter-dealer Markets

3.1 Introduction

Inter-dealer markets have received relatively little attention in the literature. Until recently, little research has addressed the microstructure of inter-dealer markets (IDM) in which dealers have a choice of trading directly with each other or trading in an inter-dealer broker market. While a direct inter-dealer market is quote-driven, a brokered inter-dealer market (IDB) resembles a screen-based order-driven market.

Some variation in the usage of IDBs exists across different dealership markets. For example, over 90% of interdealer trading is by IDB in the US Treasury market (Fleming and Remolona 1999). For the Canada government bond market, IDB trading accounts for 85% of total trading volume in 1998 (Gravelle 2002). The most recent BIS survey indicates that 98% of inter-dealer trading is conducted through IDB for UK government bonds. In contrast, Evans and Lyons (1999) shows that for FX, 50% of inter-dealer trading is on the IDB market, and that IDB accounts for 40-60% of total inter-dealer trading on the London Stock Exchange. This leads to the following question: why do dealers in markets like those for government securities tend to rely almost exclusively on interdealer brokers when conducting trading amongst themselves, while dealers in FX or the LSE trade bilaterally to a greater extent?

Lyons (1996) suggests that IDBs play a role in the information extraction process for FX dealers. Lyons posits that dealers endeavour to ascertain market-wide order flow information by observing the order flow passing through all IDB systems. In a decentralized market, this information is vital since it signals information that market makers have gathered from public orders or from other private sources. Recent work by Viswanathan and Wang (2000) shows that dealership markets with interdealer trading are preferred by customers under certain market structures. Viswanathan and Wang (2002) explain that customers prefer to trade in hybrid markets where small orders are routed to an order-driven market and large orders to a dealership (or upstairs) market.

However, these studies are unable to explain fully why dealers choose to use IDBs to execute interdealer trades instead of bilateral interdealer transactions. As well, the issues of why we observe proportionally more IDB trading in government security (GS) markets versus equity dealership markets remain unanswered. Does the scale of inter-dealing trading or IDB trading have an effect on the liquidity offered by dealers to customers in terms of bid-ask spreads, depths, or volumes? How does a decentralized multi-dealer market arrive at a particular level of IDB usage? Why do direct inter-dealer markets and brokered inter-dealer markets coexist simultaneously?

One possible explanation for the greater use of interdealer brokers in GS markets relative to other markets is that GS market makers receive orders that, relative to their desired (risk-adjusted) inventory levels, are on average larger as well as less frequent than their counterparts in dealership equity markets. Both Proudman (1995) and Vitale (1998) note that the U.K. Gilt market is characterized by large and infrequent order flow relative to that observed in the U.K. equity market. Since institutional investors (such as pension funds, mutual funds and insurance companies) are the dominant GS customers, the order flow observed by GS market makers is relatively large.

Thus, the fundamental questions addressed in this chapter are the following: Under what circumstances would a dealer prefer trading in a direct inter-dealer market to a brokered inter-dealer market? Under what circumstances would a dealer enter into market making? Would a hybrid inter-dealer market that allows a dealer to have a choice between direct inter-dealer markets and a brokered inter-dealer market improve a customer's welfare relative to an exclusive dealership structure?

In this chapter, the focus is on the determinants of a dealer's choice of inter-dealer markets as well as on the dealer's decision on entering into market making for a security. To this end, the paper compares customer welfare across three market structures; namely, direct inter-dealer, brokered inter-dealer and hybrid inter-dealer markets.

We construct a three-stage model to study dealers' choice of interdealer trading venue. At the first stage of the model, dealers make their decision on whether or not to enter market-making for the security. At the second stage of the model, each dealer makes a quote to a customer and the customer chooses to trade with the dealer who provides the best quote. At the third stage, the dealer decides whether he wants to unwind his inventory obtained by trading with a customer by trading directly with another dealer, or whether to submit a sell order to the inter-dealer broker.

We model the direct inter-dealer market as a price competition game. We assume all the bargaining power is on the customer side and that the dealer competes for customer orders. This is not a strong assumption. Empirical studies (Lee 1993, Ready 1999 and Stoll 2001) provide evidence on price improvement across different markets (i.e., effective spreads are shown to be smaller than the quoted spreads). Dealers have an incentive to attract incoming customer orders by allowing customers to trade within the quoted spreads. Consequently, this price competition among dealers has the result that no dealer expects to gain positive surplus in equilibrium. In contrast, the inter-dealer broker system is modeled as an order-driven system where dealers pay fixed brokerage fees in order to submit their orders to the broker. The broker clears the market and determines a price. All the dealers behave strategically by considering the impact that their orders have on price. In equilibrium, dealers expect to achieve a positive expected utility.

Trading motivation in the inter-dealer market results from the dealer-customer trade in the second stage. Although we assume that the customer does not possess any private information about the value of the asset, the customer's order itself still carries non-payoff related information such as the partial information on market demand and supply, which the trading dealer can further explore on the inter-dealer market.

We demonstrate that the dealer's third stage choice of inter-dealer trading venue depends on the number of dealers who provide market making due to the built-in institutional features of the two inter-dealer venues. In the direct inter-dealer market dealers compete on price. The risk-sharing opportunities offered by this market is independent of the equilibrium number of dealers as long as two or more dealers compete for the incoming inter-dealer order flow. In contrast, the risk-sharing opportunities increase with the number of dealers in a brokered market. We show that a critical number of dealers exists at any point, and that the risk-sharing benefit offered by the brokered market outweighs the benefit of competitive price in direct inter-dealer markets when the equilibrium number of dealers exceeds the critical number of dealers.

Recently a growing number of papers examine market competition in an inter-dealer trade setting. Volger (1997) develops a model that compares a dealership market to an auction market. He models the dealership market as a two-stage game where dealers first trade with customers and then trade among themselves in an inter-dealer market. Volger shows that the price offered by the dealer to the liquidity customer depends on the risk-sharing opportunities in the inter-dealer market. For a liquidity trader, his model predicts that the dealership market provides a better price. Naik, Neuberger and Viswanathan (1999) develop a two-stage theoretical model with inter-dealer trading to examine how customer welfare is affected under two trade-disclosure regimes. They show that dealership markets with a full disclosure requirement may reduce the welfare of risk-averse customers. The authors also show that an increase in post-trade transparency hinders the execution of large trades and tends to benefit small traders. Viswanathan and Wang (2002) examine the welfare of customers achieved across three different market structures: limit order-book market, dealership market and a hybrid market which is a combination of both an order book and a dealership market. To study these market structures, the authors use a discriminatory price auction, a uniform price auction and a hybrid of these two. They find that a hybrid market can dominate the dealership market and the limit-order book market. In their paper, there is no inter-dealer trading. Viswanathan and Wang (2000) investigate the importance of inter-dealer trading with a two-stage model. Comparing an order book market with a dealership market followed by a second round of inter-dealer trading, they show that the dealership market with inter-dealer trading improves customer welfare.

Our study differs from these papers in several important ways. First, the existing literature on inter-dealer trading only considers one type of inter-dealer trading, namely, the order-driven, inter-dealer broker (IDB) market. None of the papers model quote-driven, direct inter-dealer trading. We explicitly model the two types of interdealer markets and the competition between these two markets. Second, unlike the literature that focuses on customer choice among different market structures, we focus on the dealer choice of trading venues, as well as the impact of those choices on the welfare of customers. Third, unlike the existing literature, we study the endogenous entry of market-making activity.

The rest of this chapter is organized as follows. In Section 2, we describe the model in general terms. Section 3 characterizes the equilibrium in the inter-dealer trading round, and dealer choice on inter-dealer markets. We recursively solve for equilibrium in a customer-dealer trading round in Section 4. Section 5 analyzes the choice of endogenous entry and characterizes the necessary and sufficient conditions for the existence of the equilibrium number of dealers becoming active market makers, as well as their impact on customer welfare. Section 6 concludes the chapter.

3.2 Overview of the Model

The model is constructed as a three-stage game to study the dealer's choice of inter-dealer market, and the impact of this choice on the welfare of a public investor. At the first stage, dealers decide whether or not to enter the market based on expected excess returns.

The dealer decides to enter only if the expected return from market-making is no less than the cost of being a market maker. At the second stage of the model, an external customer ask for quotes from dealers, and only trades with the one who offers the best quote. At the third stage of the model, the dealer who has taken the customer order, decides whether he wants to rebalance his portfolio bilaterally with another dealer or go to the inter-dealer market to do so.

The motivation for inter-dealer trading in the third stage results from the second stage customer-dealer trade. We argue that even when customer order flow does not carry any pay-off relevant information, the dealer can still speculate on it since it conveys partial information on the demand condition of the market. At this stage, the trade between the customer and the trading dealer is not disclosed until the end of the inter-dealer trading round. This assumption partially resembles many real-world markets, such as those for government bonds and FX, which are fully opaque with respect to customer-dealer trades. The recent transparency literature argues that dealers offer a better quote for customers under an opaque regime, since it gives a dealer an incentive to attract customer order flow, and to recoup the incentive in a later trading round (see Lyons 1997 and Madhavan, 1995). We adopt their convention that dealers compete for customer orders.

In the third stage, the dealers' choice on inter-dealer markets depends on the equilibrium number of dealers who enter market-making in the first stage. Since the direct inter-dealer market is modeled as a price-competition game, the equilibrium price and quantity become independent from the number of dealers available on the market as along as more than two dealers compete for the incoming flow. Since the trade is bilateral, the risk sharing opportunity offered by the direct inter-dealer market is independent of the number of dealers available on the market. In contrast, the equilibrium price and the opportunity for risk sharing are always a function of the number of dealers in a brokered inter-dealer market. The greater the number of dealers, the closer the results are to those in perfect competition. The inventory-rebalance cost and the risk-sharing opportunity is a function of the number of dealers who engage in market-making. Thus, at any point in time, a critical number of dealer exists such that the risk sharing benefit of the brokered market outweighs the price competitive benefit of the direct market if the equilibrium number of dealers exceeds this critical number.

Since trading through the brokered inter-dealer market generates positive expected returns and requires a fixed cost for the brokerage fee, we derive the equilibrium conditions, using comparative statistics, for the equilibrium number of dealers who enter market making when inter-dealer trading is conducted through the broker. We obtain a simple characterization of the equilibrium number of dealers. We show that a sufficient increase in asset volatility, volatility of customer order flow and in the risk aversion of dealers can cause a shift of order flow from the direct to a brokered inter-dealer market.

The welfare of customers depends on the price that the customer is able to obtain from the dealer at the second stage, and, in turn, this depends upon which inter-dealer market prevails. We show that a large institutional trader is not worse-off when the dealer is given a choice between the two types of inter-dealer markets.

We now present the model more formally, and introduce some notation which is used later in this chapter. The model is a three-stage game and the game is solved by *backward induction*. The sequence of the events is as follows:

Stage 1: N out of an existing pool of dealers $\mathcal{N} = \{1, ...N\}$ decide to enter the market-making for a risky security. A dealer makes the decision to engage in market making activity by comparing the *expected utility* of being a market maker for a risky asset with that of not entering the market. The risky asset has a liquidation value of F that follows a normal distribution with $E(F)$ and variance of $var(F)$, and the opportunity cost of money is normalized to zero. We assume that no market participant has private information about the future liquidation value F. Each dealer has a negative exponential utility function with risk aversion parameter λ. We define the expected utility of not trading as the *reservation utility*, U^R. The reservation utility is equal to -1 when wealth equals zero. Dealers choose to enter market-making when the expected utility is greater than or equal to this reservation utility.

Stage 2: An outsider customer, who wishes to trade C units of the risky asset (exogenously set) for liquidity purposes, asks the dealers for quotes. Then he trades with the dealer who offers the best quote. In the case of several best quotes, the customer randomly chooses one dealer I to trade with without splitting the order. If C is positive, the customer pays a price of P_2 per unit for purchase of C units from the dealer. If C is negative, the customer receives a price of P_2 per unit from the sale of C units to the dealer. $\mathcal{I}$ is a set of indexes of dealers who provide the best quotes, and is given by:

$$\mathcal{I} = \left\{ i : P_{2,i} = \left\{ \begin{array}{l} \min(P_{2,j})_{j=1}^{N} \text{ if } C > 0 \\ \max(P_{2,j})_{j=1}^{N} \text{ if } C < 0 \end{array} \right\} \right. \tag{26}$$

Therefore, I is a random variable such that

$$\Pr(I = i) = \left\{ \begin{array}{l} \frac{1}{\#\mathcal{I}} \text{ for } i \in \mathcal{I} \\ 0, \text{otherwise.} \end{array} \right. \tag{27}$$

$\#\mathcal{I}$ is the cardinality of $\mathcal{I}$.

The price dealers quote to the customer depends on their conjectures about the trading opportunities in an inter-dealer market at stage 3 and on the bargaining power of the two counterparties in stage 3 of inter-dealer market trading.

Stage 3: The N-1 dealers who did not trade with the customer during stage 2 set quotes at which they are willing to trade with the dealer I. The incentive for $N - 1$ dealers sharing the inventory risk with dealer I is given by a risk premium that they charge on the dealer I. Dealer I decides whether he wants to adjust his position by trading with one of the N-1 dealers directly or posting an order with a broker at a fixed positive cost of K. We assume dealer I cannot split his order between dealers or between two inter-dealer markets, as in Naik, Neuberger and Viswanathan (1999) and Vogler (1997).

3.3 Equilibrium of Inter-dealer Trading in Stage 3

The game among the dealers is solved by backward induction, i.e., we solve first for an equilibrium in the brokered inter-dealer market P_3^b , Q_3^b and the one in the direct inter-dealer market P_3^d , Q_3^d. Given these equilibrium prices and quantity, we solve for the equilibrium price and quantity in the public market. Throughout our discussion, we assume that dealer I *sells* C units of the risky asset in stage 2 to a customer and *buys* Q_3^d units at stage 3. The analysis is analogous when the signs of the transaction are changed.

We use the following assumptions in the remainder of this chapter.

Assumption 1: Utility function of all dealers is a negative exponential function in wealth with constant absolute risk aversion (CARA).

Assumption 2: The fundamental value F and, in turn, the wealth are normally distributed.

Assumption 3: Dealers behave symmetrically to positive or negative customer order flow. Therefore we solve the equilibrium assuming that dealer 1 sells to the customer at Stage 2 and buys from other dealers at Stage 3. Thus, the proofs are analogous when the signs of the transactions are changed.

Assumption 4: The customer is a liquidity trader (i.e., the strategic behavior of customers are not modeled in this chapter).

3.3.1 Direct inter-dealer market

Suppose that dealer I chooses to trade in the direct inter-dealer market to *buy* Q_3^d units of the risky asset. Each dealer j among the rest of the $N-1$ dealers quotes a *supply schedule* $P_{3,j}(Q_3^d)$, and the one who quotes the lowest price sells Q_3^d units to dealer I. If more than one dealer offers the minimum ask price, dealer I chooses to trade with one of them randomly with equal probability. We denote the set of $N-1$ dealers as $\mathcal{N}^{-I}$; that is, the set of dealers excluding dealer I. The transaction price at stage three P_3^d is defined as the minimum price among all the ask prices:

$$P_3^d = \min\{P_{3,j} : j \in \mathcal{N}^{-I}\}. \tag{28}$$

Formally, we define $\mathcal{J}$ as a set of indexes of dealers who provide the best quotes, or:

$$\mathcal{J} = \{j : P_{3,j} = \min\{P_{3,j} : j \in \mathcal{N}^{-I}\}\} \tag{29}$$

and J as a random variable such that

$$\Pr(J = j) = \begin{cases} \frac{1}{\#\mathcal{J}} \text{ for } j \in \mathcal{J} \\ 0, \text{otherwise.} \end{cases}, \tag{30}$$

where $\#\mathcal{J}$ is the cardinality of $\mathcal{J}$.

Given that the $N-1$ dealers are homogeneous, we now focus on the decision problem of a representative dealer, dealer j. Dealer j decides on $P_{3,j}$ such that,

$$MaxE(U_j) = E[-\exp(-\lambda W_j)] \tag{31}$$

where the change in wealth after the trade is given by

$$W_j = \begin{cases} 0 \quad if \quad P_{3,j} = P_3^d \text{ and } J \neq j \text{ , with probability } \frac{N-2}{N-1} \\ Q_3^d(P_{3,j} - F) \quad if \ \ P_{3,j} = P_3^d \text{ and } J = j, \text{with probability } \frac{1}{N-1} \end{cases} \tag{32}$$

Dealer I, who traded with the customer during stage 2, decides on the quantity Q_3^d to maximize his expected utility which is given by

$$MaxE(U_I) = E[-\exp(-\lambda W_I)], \tag{33}$$

with wealth as defined by

$$W_I = (Q_3^d - C)F + CP_2 - Q_3^d P_3^d \tag{34}$$

Proposition 1 Given assumptions1-4, the direct inter-dealer market equilibrium quantity Q_3^d and price P_3^d are given by

$$Q_3^d = \frac{1}{2}C \tag{35}$$
$$P_3^d = E(F) + \frac{1}{2}\lambda var(F)Q_3^d.$$

The proof for this proposition and all following propositions are given in the Appendix.

At equilibrium, dealer I posts an order which is equal to half of his inventory, C, at a price which is above the expected value of the asset. Since P_3^d is an ask price, $\frac{1}{2}\lambda var(F)Q_3^d$ is the half bid-ask spread which is the risk premium that dealer j charges for providing risk sharing to dealer I. The spread is an increasing function of the size of dealer $I's$ order, dealer $j's$ risk aversion and the uncertainty of the liquidation value F. The bigger the order size, the higher the price dealer I has to pay to the rest of the dealers to absorb the excess quantity. Substituting the equilibrium price and quantity into the expected utility function (31), we obtain the reservation utility. This equilibrium confirms that dealers set a price schedule so that they are indifferent between absorbing the quantity at a higher price and not undertaking on any position in the risky asset.

3.3.2 Brokered inter-dealer market

Now suppose dealer I chooses instead to pay a fixed brokerage fee of K to place an order through the inter-dealer broker. In this inter-dealer broker market, each market participant places his order simultaneously. As a result, dealer I maximizes his utility over his demand and the rest of the $N-1$ dealers maximize their utility against their supply curves.

We assume that dealer I conjectures that the other $N-1$ dealers submit orders which are a linear function in prices and we show that the conjecture is true at equilibrium. Dealer I maximizes his utility against his order size Q_3^b given as

$$MaxE\{-\exp[-\lambda((Q_3^b - C)F - Q_3^b P_3^b + CP_2^b - K)]\}, \tag{36}$$

where P_3^b is the market clearing price at stage 3, and P_2^b is the price dealer I trades at with the customer at stage 2.

Proposition 2 Given assumptions1-4, the linear rational expectations equilibrium in a brokered inter-dealer market is given by

$$\begin{cases} Q_3^b = \frac{(N-1)(N-3)}{(N-1)^2-2}C \\ P_3^b = E(F) + \frac{N-2}{(N-1)(N-3)}\lambda var(F) Q_3^b \end{cases} \tag{37}$$

where Q_3^b is the quantity dealer I trades with the rest of the dealers, P_3^b is the market clearing price, and $D_{j|j\in\Delta_j}$ is the demand of each of the other $N-1$ dealers. Since Q_3^b must be positive, $N \geq 3$ is the market opening condition for the brokered inter-dealer market.

In the brokered equilibrium, the orders of dealers and the clearing price are all linear in dealer I's inventory $-C$. After selling C units of the risky asset to the customer in stage 2, dealer I buys back a fraction $\frac{(N-1)(N-3)}{(N-1)^2-2}$ of C at a price which is above the expected value of the asset at stage 3.

When N is finite, this equilibrium characterizes an imperfect risk sharing. It is easy to show that the position of dealer I in the risky asset after inter-dealer trading is $-\frac{2(N-2)}{(N-1)^2-2}C$. Each of the rest of the $N-1$ dealers net position in the risky asset is $D^b_{i,3} = -\frac{(N-3)}{(N-1)^2-2}C$. It is easy to check that the asset holdings of dealer I is larger than each of the rest of $N-1$ dealers since $|-\frac{2(N-2)}{(N-1)^2-2}C| > |-\frac{(N-3)}{(N-1)^2-2}C|$ for $N \geq 3$. More precisely, the smaller the number of competing market makers, the higher dealer I's portion of the customer's order. Dealer I bears the price risk by holding an unhedged position. Later we show that dealer I charges the customer a risk premium to bear inventory risk for the fraction of the risky asset that he does not buy back.

As $N \to \infty$, the quantity that dealer I buys back, Q^b_3, tends to C and the clearing price is equal to the expected value of the asset. That is, as long as N is finite, dealer I restricts his order monotonically and retains a fraction of his Stage 2 inventory. However, as $N \to \infty$, the brokered market equilibrium price, P^b_3, tends to the unconditional mean of the liquidation value, $E(F)$, and the amount , Q^b_3, that he can buy tends to C, which characterizes perfect risk-sharing.

3.3.3 Dealer's choice of inter-dealer market

To determine dealer $I's$ choice of inter-dealer market we need to compare the expected utility he receives by trading in direct inter-dealer markets with that obtained by trading through a broker.

Regardless of which inter-dealer market is chosen, dealer I's expected utility is given by:

$$E(U) = -\exp[-\lambda((Q-C)E(F) - QP(Q) + CP_2 - \tilde{K}) + (1/2)\lambda^2 var(F)(Q-C)^2, \tag{38}$$

where $P(Q)$ denotes the price that dealer I has to pay to obtain his optimal order Q at stage 3. Define the brokerage fee $\tilde{K}$ as a random variable. $\tilde{K}$ is equal to zero for trading in the direct inter-dealer market, and to K for trades in the brokered inter-dealer market.

The certainty equivalence of expected utility, as in equation (38), is given by

$$CE = (Q-C)E(F) - QP(Q) + CP_2 - \tilde{K} - (1/2)\lambda var(F)(Q-C)^2. \tag{39}$$

The quantity and the price in equation (39) depend on dealer I's choice of the inter-dealer market, such that

$$Q = \begin{cases} Q_3^b = \frac{(N-1)(N-3)}{(N-1)^2-2}C & \text{if dealer } I \text{ trades in a brokered IDM} \\ Q_3^d = \frac{1}{2}C & \text{if dealer } I \text{ trades in a direct IDM} \end{cases}, \tag{40}$$

$$P = \begin{cases} P_3^b = E(F) + \frac{(N-2)}{(N-1)(N-3)}\lambda var(F)Q_3^b & \text{if dealer } I \text{ trades in a brokered IDM} \\ P_3^d = E(F) + \frac{1}{2}\lambda var(F)Q_3^b & \text{if dealer } I \text{ trades in a direct IDM} \end{cases}. \tag{41}$$

We now can determine dealer I's choice of inter-dealer market given a fixed number of dealers. Let $CE^b(Q_3^b)$ denote the certainty equivalence that dealer I receives if he chooses to trade through the broker, and $CE^d(Q_3^d)$ denote the certainty equivalence that dealer I attains if he chooses to trade in the direct market. Substituting Q_3^b and P_3^b of (37) and Q_3^d and P_3^d of (35) into equation (39) gives the certainty equivalence of a brokered inter-dealer market CE^b and direct inter-dealer market CE^d, respectively. Dealer I chooses to trade in an inter-dealer market which allows the dealer to achieve a higher expected utility. For example, dealer I chooses to enter the brokered inter-dealer market if his certainty equivalence in this market is greater than the one expected in a direct inter-dealer market.

This implies that:

$$CE^b - CE^d \tag{42}$$
$$= -K + \frac{1}{4}\lambda var(F)C^2 - CE(F) - \frac{(N-2)[(N-1)(N-3)+2(N-2)]}{[(N-1)^2-2]^2}\lambda var(F)C^2 > 0.$$

Proposition 3 *Under assumptions1-4, we define the integer N^* as the minimum number of dealers necessary for an IDB market to dominate the direct inter-dealer market. N^* is given by:*

$$N^*(K, \lambda, var(F), C) = 1 + int\left[\frac{2-\sqrt{2(A^2+1)}}{A+1}\right],$$
$$where\ A = -\frac{4CE(F)+4K}{\lambda var(F)C^2},$$

and $A < -1$ given $N^* \geq 3$.

From proposition 3, and since the minimum number of dealers, N^*, decreases in A, we observe that a larger brokerage fee results in a greater N^* brokered interdealer market to prevail. In contrast, an increase in $var(F)$, λ or size of C shifts inter-dealer trading to an IDB.

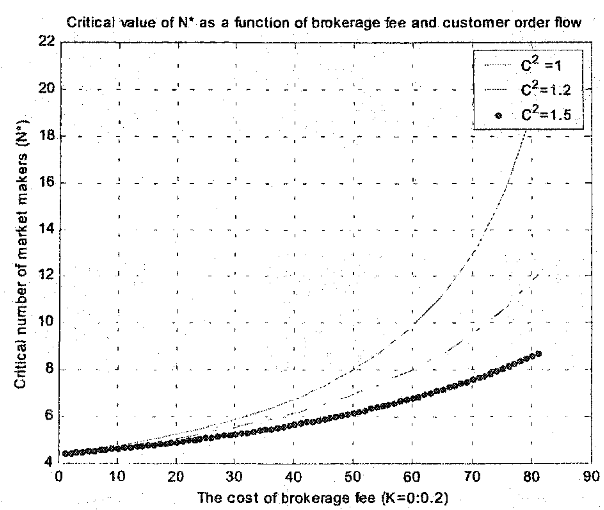

Figure 1: The Critical Value of N*

For every configuration of the initial parameters, a critical number of market makers N^* exists such that inter-dealer trading is affected by the broker when the number of dealers is equal to or greater than N^*, and occurs directly through bilateral trading otherwise.

Figure 1 plots the relationship between the critical number of dealers versus the other parameters. The X-axis is the value of the brokerage fee, K. The Y-axis is the critical number of N^*. We set both the risk aversion and the variance of the fundamental value F equal to 1, and vary the cost of trading and the size of the customer order flow. We observe that the value of N^* increases monotonically with the cost of trading when $1 > A \geq 0$. The intuition is that the dealer trades in a brokered inter-dealer market only if the return from trading is able to offset the cost of participation since trading in a brokered inter-dealer market is costly. Thus, the higher the cost, the less competitive this market is compared to the direct inter-dealer market. We also observe that the critical value of N^*decreases in the average size of customer order flow given a fixed trading cost. For example, for the same value of K=0.2 (the last observation on each curve), the critical value of N^* decreases from 20 to 9 when the square of the customer order size C^2 increases from 1 to 1.5. This implies that the market maker tends to rely on inter-dealer trading to share the inventory risk when a market is featured by large customer order flows and the market maker, in turn, is hit frequently by large unwanted inventory. Therefore, the market maker chooses to trade in the inter-dealer market to lay-off more of the order from his original customer. From equations (35) and (37), it is easy to see that dealer I can trade a larger proportion of his inventory in a brokered inter-dealer market when $N > 4$, all else held equal.[6] Therefore,

[6] This is partially driven by the assumption that dealers cannot split their orders in a brokered inter-dealer market. In order to check this effect, a different model is needed for the IBD market instead of a model of

when the dealer is faced with *large* customer order flow, he tends to choose to trade in the brokered inter-dealer market to rebalance his inventory. The greater the need for risk sharing, the more attractive are trades in the brokered inter-dealer market than in the direct inter-dealer market. In other words, when the customer order flow is large, a small number of dealers are required for a brokered market to dominate a direct market.

3.4 Equilibrium of Stage 2: Public Trading

Given equilibrium in the third stage, we now consider the price setting of dealers in the second stage. In stage 2, each dealer sets an ask price at which he is willing to sell C units of the risky asset to a customer; recalling that $C > 0$ implies that the customer wants to buy. The customer trades with the dealer who quotes the best price, denoted as dealer I.

We focus our discussion on the decision problem of the representative dealer, denoted as dealer i. We define P_2 as the price quoted by the rest of dealers except dealer i :

$$(P_{2,j})_{j \in \mathcal{N}^{-i}} = P_2, \tag{43}$$

where the set $\mathcal{N}^{-i} = \{1, ...i-1, i+1...N\}$. We show at equilibrium, that dealer i is better off if he does not deviate from the rest of the dealer's quotes.

If dealer i quotes an ask price, $P_{2,j}$, which is lower than the quote from the rest of the dealers, P_2, his quote is accepted and dealer i trades with the customer. If dealer i quotes an ask price which is higher than the rest of the dealers, his quote is not accepted. If $P_{2,i} < P_2$, then $i = I$; if $P_{2,i} > P_2$, then $i \neq I$; if $P_{2,i} = P_2$, then $i = I$ with probability $\frac{1}{N}$, where I

quantity competition.

is the index of the dealer who trades with the customer and $\mathcal{N}^{-I}$ (as defined in stage 3) is a set of the rest of the $N-1$ dealers who did not trade with the customer at stage 2.

From our above discussion of inter-dealer trading, we know that the dealer's payoff depends on the number of dealers who enter market making. Proposition 1 implies that dealer i sets price such that

$$E(U_i) = \begin{cases} E(U_i^d) & \text{if } N < N^*, \text{for a direct IDM,} \\ E(U_i^b) & \text{if } N \geq N^*, \text{ for a brokered IDM,} \end{cases} \tag{44}$$

where $E(U_i^d)$ and $E(U_i^b)$ denote the expected utility of dealer i when the third stage trading occurs in the direct inter-dealer market or the IDB market, respectively. Now we first consider the expected utility of dealer i when $N < N^*$, or the case where inter-dealer trading is conducted directly.

The expected utility of dealer i depends on what type of dealer he is. In turn, this depends upon the results of the game. For example, if dealer i wins the auction and trades with the customer, he becomes the dealer I at stage 3 who decides to buy the quantity Q_3^d to rebalance his portfolio. Therefore, the expected utility of dealer i at stage 2 when he quotes $P_{2,i} < P_2$ is equal to the expected utility of dealer I at stage 3 in the direct inter-dealer market, which is

$$E(U_i^d)|_{P_{2,i}<P_2} = E[-\exp(-\lambda(Q_3^d - C)F + CP_{2,i} - Q_3^d P_3^d)]. \tag{45}$$

If dealer i loses the auction by quoting a price higher than P_2 at the current stage, he becomes one of the $N-1$ dealers at stage 3 who needs to provide another quote. Given the uniform distribution of the winning chance, he could be the one who is randomly chosen to trade with dealer I with probability $\frac{1}{N-1}$, or not to be chosen with probability of $\frac{N-2}{N-1}$.

Therefore, the expected utility of dealer i when $P_{2,i} > P_2$ is given by

$$E(U_i^d)|_{P_{2,i}>P_2} = \frac{N-2}{N-1}(-1) + \frac{1}{N-1}E[-\exp(-\lambda Q_3^d(P_3^d - F))]. \tag{46}$$

If dealer i quotes a price which is equal to P_2, then he can win the auction with probability of $\frac{1}{N}$ to be the dealer who trades with the customer or lose the auction to become one of the $N-1$ dealers with probability of $\frac{N-1}{N}$ in stage 3. It follows that

$$E(U_i^d)|_{P_{2,i}=P_2^d} = \begin{cases} E[-\exp(-\lambda(Q_3^d - C)F + CP_{2,i} - Q_3^d P_3^d)] & \text{with } \Pr = \frac{1}{N} \\ \frac{N-2}{N-1}(-1) + \frac{1}{N-1}E[-\exp(-\lambda Q_3^d(P_3^d - F))] & \text{with } \Pr = \frac{N-1}{N} \end{cases}. \tag{47}$$

In the Appendix, we solve for the equilibrium price P_2 and show that dealer i is better off by quoting the same price as the rest of the dealers than by deviating from the equilibrium price. We also show that dealers receive their reservation utility at the equilibrium price.

We now consider the expected utility of dealer i when $N > N^*$, or inter-dealer trading in stage 3 is conducted through a broker. If dealer i quotes $P_{2,i} > P_2$, his expected utility is

$$\begin{aligned} E(U_i^b)|_{P_{2,i}>P_2} &= E(U_i^b)|_{i\neq I} \\ &= E[-\exp(-\lambda(D_{i,3}^b F - D_{i,3}^b P_3^b - K))]. \end{aligned} \tag{48}$$

If dealer i quotes $P_{2,i} < P_2$, he wins the auction to trade with the customer in stage 2 and becomes dealer I in stage 3. His expected utility over the game is

$$\begin{aligned} E(U_i^b)|_{P_{2,i}<P_2} &= E(U_i^b)|_{i=I} \\ &= E[-\exp(-\lambda((-C + Q_3^b)F + CP_{2,i} - Q_3^b P_3^b - K)]. \end{aligned} \tag{49}$$

If dealer i quotes the same price as the rest of the dealers, $P_{2,i} = P_2$, he will win the auction with probability $\frac{1}{N}$ and lose the auction with probability $\frac{N-1}{N}$ given a uniform distribution of wining the auction by chance.

$$
\begin{aligned}
E(U_i^b)|_{P_{2,i}=P_2} &= \frac{1}{N}E(U_i^b)|_{i=I} + \frac{N-1}{N}E(U_i^b)|_{i\neq I} \\
&= \frac{1}{N}E[-\exp(-\lambda((-C+Q_3^b)F + CP_{2,i} - Q_3^b P_3^b - K)] \\
&+ \frac{N-1}{N}E[-\exp(-\lambda(D_{i,3}^b F - D_{i,3}^b P_3^b - K))].
\end{aligned} \tag{50}
$$

Proposition 4 Under assumptions 1-4, and given that the market opening condition is satisfied, the customer receives a price P_2 at stage 2, that is given by

$$
P_2 = \begin{cases} E(F) + \frac{1}{4}\lambda var(F)C & \text{if } N < N^* \\ E(F) + \frac{2(N^2-N-4)(N-2)-(N-3)^2}{2[(N-1)^2-2]^2}\lambda var(F)C & \text{if } N \geq N^*. \end{cases} \tag{51}
$$

Proposition 4 describes the link between the choice of inter-dealer market and the prices obtained by the customer in stage 2. Dealer I only buys back a fraction of his trade in stage 3 inter-dealer trading. At the end of stage 3, dealer I is still short of the asset by a fraction of $\frac{1}{2}$ when $N < N^*$ and $\frac{2(N-2)}{(N-1)^2-2}$ when $N \geq N^*$. The price P_2 in equation (51) is the compensation dealer I requires to take on the inventory risk for the fraction of the asset he does not buy back.

Naik *et al.* (1999) obtain a pricing rule similar to (51) in their two-period model of an opaque dealership market. In their model, inter-dealer trading is conducted multilaterally, costlessly and competitively, in the sense that dealers receive their reservation utility. Unlike this paper, the focus of their analysis is to compare the prices obtained by an investor in a standard auction market and a dealership market with or without disclosure.

Substituting the equilibrium price (51) back into the expected utility when $N < N^*$ in equation (47), we can easily show that the expected utility at equilibrium is equal to:

$$
E(U_i^d)|_{P_{2,i}=P_2} = -1. \tag{52}
$$

Similarly substituting the equilibrium price (51) back into the expected utility in equation (50), the expected utility at equilibrium when $N > N^*$ is equal to:

$$E(U_i^b)|_{P_{2,i}=P_2} = -\exp(-\lambda E(F)C - \frac{(N-1)(N-3)}{2[(N-1)^2-2]^2}\lambda^2 var(F)C^2 + \lambda K). \qquad (53)$$

For greater details, see the Appendix.

3.5 Equilibrium of Stage 1: Endogeneous Entry for Market Making

In the analysis of Stage 3, we showed that we can predict when inter-dealer trading will occur as long as we know the number of dealers for a given set of initial parameters. Furthermore, in our analysis of stage 2, we showed that the price received by the customer depends on the dealer's choice of inter-dealer trading venue. The customer's welfare depends upon whether the dealer has a choice between inter-dealer trading venues.

In this section, we first discuss the factors that determine the number of dealers who enter into active market making. Then we address the impact on a public investor's welfare of dealers having a choice between various types of inter-dealer market.

3.5.1 The equilibrium number of dealers

In the analysis of stage 2, we considered that, if the number of dealers in the market exceeds or is equal to N^*, then each dealer attains an expected utility equal to $E(U_i^b)|_{P_{2,i}=P_2} = -\exp(-\lambda E(F)C - \frac{(N-1)(N-3)}{2[(N-1)^2-2]^2}\lambda^2 var(F)C^2 + \lambda K)$ in (53). If the number of dealers is between 3 and N^*, each dealer attains her reservation utility -1 in equation (52).

If an equilibrium number of active dealers $\widetilde{N}$ exists, then the expected utility of those agents who become dealers is at least as large as the expected utility of those agents who do not enter the industry; that is:

$$E\{-\exp(-\lambda E(F)C - \frac{(N-1)(N-3)}{2[(N-1)^2-2]^2}\lambda^2 var(F)C^2 + \lambda K)\} \geq -1. \tag{54}$$

Rearranging we obtain:

$$\exp(-\lambda K) \geq E(\exp(-\lambda E(F)C - \lambda AC^2)), \tag{55}$$

where

$$A = \frac{(N-1)(N-3)}{2[(N-1)^2-2]^2}\lambda^2 var(F).$$

We know that the customer order flow is normally distributed as $C\tilde{}N(0,\sigma_C^2)$, $E(C)=0$, therefore equation (55) can be simplified to

$$\exp(-\lambda K) \geq E(\exp(-\lambda AC^2)).$$

Now we need to find $E(\exp(-\lambda AC^2)) = (1-2A\sigma_C^2)^{-\frac{1}{2}}$as follows:

$$\begin{aligned}
&E(\exp(-\lambda AC^2)) \\
&= \int_{-\infty}^{\infty} \exp(-\lambda AC^2)\frac{1}{\sigma_C\sqrt{2\pi}}\exp(-\frac{C^2}{2\sigma_C^2})dC \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma_C\sqrt{2\pi}}\exp(-\frac{C^2}{2\sigma_C^2} - \lambda AC^2)dC \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma_C\sqrt{2\pi}}\exp[-\frac{C^2}{2(\frac{\sigma_C^2}{1+2\lambda\sigma_C^2 A})}]dC \\
&= \frac{1}{\sqrt{1+2\lambda\sigma_C^2 A}}\int_{-\infty}^{\infty} \frac{1}{\sigma_C\sqrt{\frac{2\pi}{1+2\lambda\sigma_C^2 A}}}\exp[-\frac{C^2}{2(\frac{\sigma_C^2}{1+2\lambda\sigma_C^2 A})}]dC \\
&= (1+2\lambda\sigma_C^2 A)^{-\frac{1}{2}}
\end{aligned} \tag{56}$$

Therefore, equation (55) yields

$$\exp(-\lambda K) \geq (1 + 2\lambda A\sigma_C^2)^{-\frac{1}{2}}. \tag{57}$$

We rearrange the expression as

$$(1 + 2\lambda A\sigma_C^2) \geq \exp(2\lambda K). \tag{58}$$

Using a Taylor's expansion in equation (57) on the right hand side of (58) and ignoring terms of order $O((\lambda K)^2)$, the equilibrium condition can be approximated by

$$A\sigma_C^2 \geq K$$

or

$$\frac{(N-1)(N-3)}{2[(N-1)^2-2]^2}\lambda var(F)\sigma_C^2 \geq K, \tag{59}$$

or

$$H(N, \lambda, var(F), \sigma_C^2) \geq K.$$

where $H(N, \lambda, var(F), \sigma_C^2) = \frac{(N-1)(N-3)}{2[(N-1)^2-2]^2}\lambda var(F)\sigma_C^2$ is the *expected return* of entering active market-making.

This approximation is the equilibrium condition for a dealer entering market-making.[7] The left hand side of the equation represents the expected return that a dealer receives by entering market making, and the right hand side is the brokerage cost. This equation states that the dealer decides to enter market-making if the expected return from making a market is greater than or equal to the cost of entering the market.

This determination leads to the necessary and sufficient condition for the equilibrium number of dealer $\widetilde{N}$ to exist:[8]

---

[7] This approximation is not valid for large values of K and large coefficients of risk aversion.

[8] Note that if equation (26) is violated, direct inter-dealer trading dominates brokered inter-dealer trading.

Proposition 5 Under assumptions 1-4, the necessary and sufficient condition for an equilibrium number of dealers $\widetilde{N}$ ($\geq N^*$) to exist is the following: Since the expected return is decreasing in N for $N \geq N^*$ and equal to zero at the limit which are

$$\frac{\partial H(N, \lambda, var(F), \sigma_C^2)}{\partial N}|_{N \geq N^*} \leq 0 \tag{60}$$

and

$$\lim_{N \to \infty} H(N, \lambda, var(F), \sigma_C^2) = 0,$$

and as long as

$$H(N^*, \lambda, var(F), \sigma_C^2) \geq K \tag{61}$$

then there always exists a $\widetilde{N}$ which makes the expected return equal to the brokerage fee,

$$H(\widetilde{N}, \lambda, var(F), \sigma_C^2) = K.$$

This value of $\widetilde{N}$ is the equilibrium number of dealers entering the market making function.

In the Appendix we show that dealers' expected return $H(N, \lambda, var(F), \sigma_C^2)$ decreases in N. This proposition states that $\widetilde{N}$ is the equilibrium number of dealers where the expected return is equal to the cost. The entrance of another dealer into the market decreases the expected return per capita. Adding one more dealer has two effects. First, it lowers the inventory risk since dealer I can rebalance his inventory at a lower cost, since the partial derivative of the equilibrium price (37) with respect to N is negative. Sec-

Our modeling of the direct inter-dealer market does not allow us to determine the exact number of dealers who enter the market-making industry when inter-dealer trading is conducted directly. The latter occurs when equation (26) is violated and equation (27) is satisfied. Therefore, we use the term *active market-making* to indicate that the equilibrium number of dealers derived from this model is conditional on inter-dealer trading via IDB.

ond, it also lowers price volatility. In equation (37), the variance of the price is given by $var(P_3^b) = [\frac{(N-1)(N-2)(N-3)}{(N-1)^2-2}\lambda var(F)]^2\sigma_C^2$. It is easy to show that the partial derivative of the variance with respect to N is negative for $N \geq 5$. When the number of dealers participating in the brokered market is small, the first effect dominates the second one. However, when the number of dealers participating in the brokered market is large, the second effect always dominates.

Corollary 6 The equilibrium number of dealers increases in the variance of customer order flow, risk-aversion of the dealer, the volatility of the asset, and it decreases in the brokerage fee, K.

Corollary 6 indicates the link between the dynamics of customer order flow and the equilibrium number of dealers actively making the market. The more volatile the customer order flow is, the greater number of dealers that will enter active market-making, and, in turn, choose to trade via the IDB market.

This result is particularly interesting in explaining the greater concentration of IDB trading in government security markets. As we have discussed, the government bonds markets are driven by the big institutional investors, and therefore, are characterized by large, infrequent customer order flow. In other words, the customer order flow is more volatile relative to that in an equity market. According to Corollary 6, an increase in the volatility of customer order flow tends to shift trading between dealers to the IDB market. This result supports the observation on government security markets.

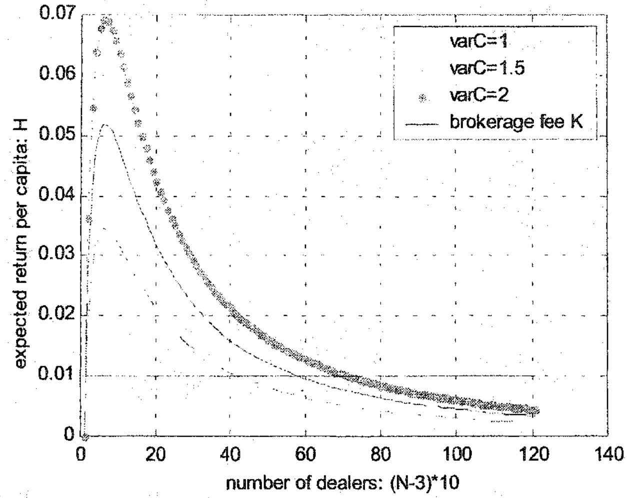

Figure 2: Dealer's Expected Return vs. Number of Dealers

Figure 2 shows the variation of the expected return (per capita) for N dealers, for a given set of exogenous parameters. The X-axis is the number of dealers N, and the Y-axis is the expected return from entering market making, given by $H(N, \lambda, var(F), \sigma_C^2)$. The value for the parameters are 0.01 for fixed brokerage fee K, 1 for the variance of fundamental value F, and 1 for the risk aversion parameter of dealers λ. Given this set of parameters, we plot the expected return as a function of the variance of customer order flow, σ_C^2, which takes the value of $\{1, 1.5, 2\}$. We observe that when $N \geq 5$, the expected return is decreasing in the number of dealers entering the market-making industry.

3.5.2 Impact on Public Welfare

As is discussed in the introduction to this chapter, most dealership markets converged in recent years to a hybrid structure that has a direct inter-dealer market and a brokered inter-dealer market. This gives dealers a choice of which inter-dealer market to trade in. In this section, we examine if a public investor is better-off when dealers can choose between quote-driven direct trading and order-driven brokered trading.

Suppose that the inter-dealer brokered market does not exist, so that all inter-dealer trading is conducted bilaterally. Further assume that $\widetilde{N} \geq N^*$ dealers enter the market making function for the security. From Proposition 4, the customer who wishes to buy C units of the security expects to negotiate the following price:

$$E(P_2) = E(F) + \frac{1}{4}\lambda var(F)C \tag{62}$$

with volatility

$$var(P_2) = \frac{1}{16}\lambda^2 \sigma_F^4 \sigma_C^2. \tag{63}$$

In a hybrid inter-dealer market structure, Proposition 3 predicts that the brokered inter-dealer market will dominate the direct inter-dealer market when $\widetilde{N} \geq N^*$. From Proposition 4 and the equilibrium condition (51), it follows that the customer expects to negotiate a price given by

$$E(P_2) = E(F) + \frac{2(N^2 - N - 4)(N-2) - (N-3)^2}{2[(N-1)^2 - 2]^2}\lambda var(F)C \tag{64}$$

with volatility

$$var(P_2) = [\frac{2(N^2 - N - 4)(N-2) - (N-3)^2}{2[(N-1)^2 - 2]^2}]^2 \lambda^2 \sigma_F^4 \sigma_C^2. \tag{65}$$

Inspection of equation (62) versus (64), and (63) versus (65) reveals that the introduction of the brokered system reduces both the mean and the variance of the price for the customer. In the Appendix, we prove the following proposition on the impact of the dealer choice on customer welfare.

Proposition 7 Under assumptions 1-4, suppose that the customer has a mean-variance utility function with a risk aversion of β given as

$$E(-CP_2) - \beta var(P_2)C^2.$$

Under this utility function, a customer who wants to trade large orders is at least no worse-off when dealers have a choice between inter-dealer trading venues.

Proposition 7 results from the fact that the bargaining power between the dealer and the customer is assumed to rest with the customer. If the customer's order is large enough so that the per unit fixed costs are insignificant, this implies that higher benefits derived by dealers from hybrid inter-dealer markets are passed on to the customer through a lower expected price. Moreover, the customer expects that the hybrid dealership market also will lead to a reduction of price volatility.

3.6 Conclusion

This paper constructs a three-stage model of a dealership market where dealers trade with customers first, and then trade on an inter-dealer market (IDM) to rebalance their portfolios. We suppose that dealers can choose between an order-driven brokered IDM and a quote-driven direct IDM. The choice of two inter-dealer trading venues reflects the existence of IDB markets in a number of dealership markets such as the foreign exchange markets and government bond markets. We show that the brokered IDM dominates the direct IDM for a given set of exogenous parameters when the number of dealers on the market is greater than or equal to N^*. Under this condition, the brokered inter-dealer market provides a better risk-sharing opportunity in the sense that the trading dealer can unwind a greater proportion of his inventory at a lower cost.

We also endogenously derive the equilibrium number of dealers who become active market makers. We demonstrate that, in equilibrium, a dealer quotes the customer a price such that the dealer receives his reservation utility at this price as a result of price competition with other dealers. Therefore, a dealer's final payoff is determined by the choice of IDM.

In the absence of information asymmetry, we show that a sufficient increase in volatility of the asset and the risk aversion of the dealer tends to shifts IDMs from being direct to being brokered. Moreover, our findings suggests a link between the choice of inter-dealer trading venue with the dynamics of customer order flow. We show that an increase in the volatility of customer order flow tends to shift the IDM to be brokered. This finding is particularly interesting for the government bonds markets, where private information is less of an issue. Unfortunately, this empirically testable implication of the model only applies to markets where trading on private information is not significant. For example, Proudman (1995) shows empirically that trading in the UK gilts markets reveals little private information.

We also demonstrate that the dealers' choice of inter-dealer trading venue has an impact on customer welfare. When dealers have a choice between inter-dealer markets, we show that a customer can negotiate a better or at least not-worse price than when they do not. In turn, this implies that the customer welfare improves or at least is not worse-off when dealers have a choice between inter-dealer trading venues. This result may explain the wide-spread use of the hybrid of two inter-dealer trading mechanisms in a variety of dealership markets which are populated by institutional investors.

However, there are a few limitations in this study. One is the absence of asymmetric information. These results can only apply to a market where the presence of private information is less of a concern. They may not be generalizable when private information is an issue. In the future study, considering the presence of asymmetric information, one should be able to answer the questions dealing with the extent that inter-dealer trades are driven by signal extracting versus risk sharing, or the contribution of asymmetric information to the bid ask spread of a dealer's quote and dealer's choice of inter-dealer market. Another limitation of this study is that dealers have limited means for risk sharing. We assume that dealers immediately offset the inventory impact of the customer's order in the inter-dealer market after trade in the public trading stage. In reality, dealers may have other choices for risk sharing. The dealer may find it more advantageous to spread the trade over time with other dealers, to attract customer order flow from the opposite side of the market, or hedge his spot market position on future markets. Our analysis does not explore any of these possibilities. The existence of more trading opportunities could lead to different results. When dealers face more choices for risk sharing, hypothetically the dealer's optimal strategy is determined by the trade-off between the benefits from postponing the trades to the next stage to wait for a better price and the risk of holding larger unhedged positions as well as price uncertainty in the next stage. We leave this for future work.

3.7 Appendix

In all the proofs that follow, we assume that Dealer I sells a quantity of C units of the risky asset to the customer at Stage 2 and buys a quantity of Q_3^d (or Q_3^b) in an inter-dealer market at Stage 3. The proofs are analogous when the signs of the transaction are changed.

Proof of Proposition 1:Direct inter-dealer market equilibrium

Each of the N -1 dealers who did not trade with the public at Stage 2 set a price schedule at which his expected utility of selling Q_3^d to Dealer I is at least equivalent to his reservation utility. Due to the homogeneity of the N -1 dealers, the rest of the $N-1$ dealers quote the same price schedule at the equilibrium. This symmetric equilibrium implies that the expected utility of the representative dealer j by quoting $P_{3,j} = P_3^d$ is greater than or equal to the one he receives by deviating from P_3^d. Formally, the two characteristics of the equilibrium can be written as

$$E(U_j) \mid_{P_{3,j}=P_3^d} \geq E(U_j) \mid_{P_{3,j}>P_3^d} \tag{66}$$

and

$$E(U_j) \mid_{P_{3,j}=P_3^d} \geq E(U_j) \mid_{P_{3,j}<P_3^d}. \tag{67}$$

It follows that:

$$E(U_j) \mid_{P_{3,j}>P_3^d} = -1.$$

$$E(U_j) \mid_{P_{3,j}<P_3^d} = E[-\exp(-\lambda Q_3^d(P_{3,j} - F))] \tag{68}$$

$$E(U_j) \mid_{P_{3,j}=P_3^d} = \frac{N-2}{N-1}(-1) + \frac{1}{N-1}E[-\exp(-\lambda Q_3^d(P_{3,j} - F))] \tag{69}$$

Since $F \sim N(0, \sigma_F^2)$ and $\lambda Q_3^d F \sim N(0, \lambda^2 (Q_3^d)^2 \sigma_F^2)$, equations (68) and (69) can be rewritten as

$$E(U_j) \mid_{P_{3,j} < P_3^d} = -\exp(\lambda Q_3^d E(F) - \lambda Q_3^d P_{3,j} + (1/2)\lambda^2 (Q_3^d)^2 \sigma_F^2)) \tag{70}$$

$$E(U_j) \mid_{P_{3,j} = P_3^d} = \frac{N-2}{N-1}(-1) + \frac{1}{N-1}[-\exp(\lambda Q_3^d E(F) - \lambda Q_3^d P_3^d + (1/2)\lambda^2 (Q_3^d)^2 \sigma_F^2))]. \tag{71}$$

Let $\varepsilon > 0$ and $P_{3,j} = P_3^d + \varepsilon$. The equilibrium condition in equation (66) implies

$$\frac{N-2}{N-1}(-1) + \frac{1}{N-1}[-\exp(\lambda Q_3^d E(F) - \lambda Q_3^d P_{3,j} + (1/2)\lambda^2 (Q_3^d)^2 \sigma_F^2))] \geq -1.$$

It then follows that

$$P_3^d \geq E(F) + \frac{1}{2}\lambda \sigma_F^2 Q_3^d. \tag{72}$$

Let $P_{3,j} = P_3^d - \varepsilon$. The equilibrium condition in equation (67) implies

$$\begin{aligned} &\frac{N-2}{N-1}(-1) + \frac{1}{N-1}[-\exp(\lambda Q_3^d E(F) - \lambda Q_3^d P_3^d + (1/2)\lambda^2 (Q_3^d)^2 \sigma_F^2))] \\ &\geq -\exp(\lambda Q_3^d E(F) - \lambda Q_3^d (P_3^d - \varepsilon) + (1/2)\lambda^2 (Q_3^d)^2 \sigma_F^2)), \end{aligned} \tag{73}$$

which can be simplified to

$$[(N-1)\exp(\lambda Q_3^d \varepsilon) - 1]\exp(\lambda Q_3^d E(F) - \lambda Q_3^d P_3^d + (1/2)\lambda^2 (Q_3^d)^2 \sigma_F^2) \geq (N-2).$$

We first notice that the l.h.s. of the inequality is an increasing function of ε. Therefore, the inequality holds $\forall \varepsilon > 0$ if and only if it holds for $\varepsilon \to 0$. In this case, $\exp(\lambda Q_3^d \varepsilon) \to 1$. It then follows that

$$P_3^d \leq E(F) + \frac{1}{2}\lambda \sigma_F^2 Q_3^d \tag{74}$$

Combining equations (72) and (74), we get the pricing strategy for $N-1$ dealers in stage 3 as

$$P_3^d = E(F) + \frac{1}{2}\lambda\sigma_F^2 Q_3^d$$

In order to obtain the equilibrium order Q_3^d, we derive the first order condition from dealer I's utility and wealth functions, given by equations (33) and (34), as

$$\lambda E(F) - \lambda(P_3^d + Q_3^d(P_3^d)') - \lambda^2 var(F)(Q_3^d - C) = 0$$

Substituting into the equilibrium pricing rule we get

$$Q_3^d = \frac{1}{2}C$$

Proof of Proposition 2: brokered inter-dealer market equilibrium

Dealer I conjectures that each of the other $N-1$ dealer submit linear orders D_3^b. We now show that this conjecture is true at equilibrium. The linear order D_3^b is given by

$$D_{j,3}^b|_{j\in\mathcal{N}-I} = a - bP_3^b \tag{75}$$

Market clearing conditions imply that

$$(N-1)D_3^b + Q_3^b = 0 \tag{76}$$

which is equivalent to

$$P_3^b = \frac{a}{b} + \frac{1}{(N-1)b}Q_3^b \tag{77}$$

Dealer I maximizes his utility given by equation (36) against his order size Q_3^b, which is equivalent to maximizing the quadratic function given by

$$E(F)(Q_3^b - C) - Q_3^b P_3^b + CP_2^b - K - (1/2)\lambda^2 var(F)(Q_3^b - C)^2. \tag{78}$$

The first order condition of equation (78) w.r.t.Q_3^b is

$$E(F) - Q_3^b(P_3^b)' - P_3^b - var(F)(Q_3^b - C) = 0 \tag{79}$$

Substituting equation (79) into the equation above, we obtain

$$E(F) - \frac{2}{(N-1)b}Q_3^b - \frac{a}{b} - \lambda var(F)(Q_3^b - C) = 0$$

which yields

$$Q_3^b = \frac{E(F) + \lambda var(F)C - a/b}{\frac{2}{(N-1)b} + \lambda var(F)} \tag{80}$$

Now we consider the rest of the $N-1$ dealers. Given each of the other $N-1$ dealers is identical, we focus on the decision problem of the representative dealer, denoted by dealer i. We assume that each dealer conjectures that the other $N-2$ dealers submit limit orders. We show that their conjectures are true at equilibrium. Dealer i submits his order taking into account the impact of his order $D_{i,3}^b$ on the equilibrium price. The market clearing condition implies that

$$(N-2)(a - bP_3^b) + D_{i,3}^b + Q_3^b = 0$$

or

$$P_3^b = \frac{a}{b} + \frac{1}{(N-2)b}(Q_3^b + D_{i,3}^b) \tag{81}$$

It follows that dealer i maximizes a quadratic function given by

$$MaxE(U) = -\exp(\lambda D_{i,3}^b E(F) - \lambda P_3^b D_{i,3}^b + (1/2)\lambda^2 var(F)(D_{i,3}^b)^2 - \lambda K)$$

The first order condition w.r.t. $D_{i,3}^b$ is given by

$$-E(F) + P_3^b + D_{i,3}^b (P_3^b)' - \lambda var(F) D_{i,3}^b = 0 \tag{82}$$

Given equation (81), we can rewrite the equation above as

$$\left(\frac{1}{(N-2)b} - \lambda var(F)\right) D_{i,3}^b = E(F) - P_3^b$$

and solve for the parameters a and b as

$$\begin{aligned} a &= \frac{N-3}{(N-2)\lambda var(F)} E(F) \\ b &= \frac{N-3}{(N-2)\lambda var(F)} \end{aligned} \tag{83}$$

Substituting a and b into equation (80) and (77), we obtain the equilibrium price and quantity (37) as follows:

$$\begin{aligned} P_3^b &= E(F) + \frac{(N-2)}{(N-1)(N-3)} \lambda var(F) Q_3^b \\ Q_3^b &= \frac{(N-1)(N-3)}{(N-1)^2 - 2} C \end{aligned}$$

.

Proof of Proposition 3

To prove proposition 3, we needed to find the conditions under which the $CE^b - CE^d \succeq 0$, that is, the condition where the expression in equation (42) is non-negative, or:

$$CE^b - CE^d = -K + \frac{1}{4}\lambda var(F)C^2 - CE(F) - \frac{(N-2)(N^2-2N-1)}{[(N-1)^2-2]^2}\lambda var(F)C^2 > 0.$$

After some rearranging we show that expression (42) is positive if and only if

$$\frac{N-2}{(N-1)^2-2} \leq \frac{1}{4} - \frac{E(F)}{\lambda var(F)C} - \frac{K}{\lambda var(F)C^2}. \tag{84}$$

Let

$$A = -\frac{4CE(F)+4K}{\lambda var(F)C^2}.$$

It follows that the roots of $CE^b - CE^d \succeq 0$ are a roots of

$$A + 1 - \frac{4(N-2)}{(N-1)^2-2} \succeq 0. \tag{85}$$

The roots are

$$N^* = 1 + \frac{2 \pm \sqrt{2(A^2+1)}}{A+1}$$

Since $N \geq 3$ is a necessary condition for a brokered market to open, $A < -1$ or $A \geq 1$. Since $A = -\frac{4CE(F)+4K}{\lambda var(F)C^2} \leq 0$, we get the condition on A that $A < -1$. Therefore, the larger root of N^* is ignored. Given that the necessary condition for a brokered market to open is that $N \geq 3$, we can have $CE^b - CE^d \geq 0$ only if N exceeds the larger root by N^*.

For $N > N^*$, $CE^b - CE^d$ is positive and dealer I chooses to trade through the inter-dealer broker. However, if N is less than N^*, then $CE^b - CE^d$ is negative and dealer I trades on the direct market.

From the equilibrium of the two inter-dealer markets (35) and (37), we observe the following relationship: a larger brokerage fee results in a greater N^* brokered interdealer market to prevail. In contrast, an increase in $var(F)$, λ or size of C shifts inter-dealer trading to an IDB.

Proof of Proposition 4 when $N < N^*$

A symmetric equilibrium requires that the following two conditions are satisfied:

$$E(U_i^d)|_{P_{2,i}=P_2} \geq E(U_i^d)|_{P_{2,i}>P_2} \tag{86}$$

$$E(U_i^d)|_{P_{2,i}=P_2} \geq E(U_i^d)|_{P_{2,i}<P_2}. \tag{87}$$

If $P_{2,i} = P_2 + \varepsilon$, given (46) and (47), condition (86) requires that

$$\begin{aligned}
&\frac{1}{N}E[-\exp(-\lambda(Q_3^d - C)F + CP_{2,i} - Q_3^d P_3^d)] + \frac{N-1}{N}\{\frac{N-2}{N-1}(-1) \\
&+ \frac{1}{N-1}E[-\exp(-\lambda Q_3^d(P_3^d - F))]\} \\
&\geq \frac{N-2}{N-1}(-1) + \frac{1}{N-1}E[-\exp(-\lambda Q_3^d(P_3^d - F))]
\end{aligned} \tag{88}$$

The certainty equivalence of equation (88) is

$$\begin{aligned}
&\frac{1}{N}(CP_{2,i} - Q_3^d P_3^d - (1/2)\lambda(Q_3^d - C)^2 var(F)) + \frac{N-1}{N}[\frac{N-2}{N-1}(-1) \\
&+ \frac{1}{N-1}(Q_3^d P_3^d - (1/2)\lambda(Q_3^d)^2 var(F))] \\
&\geq \frac{N-2}{N-1}(-1) + \frac{1}{N-1}(Q_3^d P_3^d - (1/2)\lambda(Q_3^d)^2 var(F))
\end{aligned} \tag{89}$$

Substituting P_3^d and Q_3^d from equation (35) into (89) and rearranging the terms, we obtain

$$P_2 \geq \frac{1}{4}\lambda var(F)C \tag{90}$$

If $P_{2,i} = P_2 - \varepsilon$, given equation (45) and (47), condition (87) requires that

$$\begin{aligned}
&\frac{1}{N}E[-\exp(-\lambda(Q_3^d - C)F + C(P_2 - \varepsilon) - Q_3^d P_3^d)] + \frac{N-1}{N}\{\frac{N-2}{N-1}(-1) \\
&+ \frac{1}{N-1}E[-\exp(-\lambda Q_3^d(P_3^d - F))]\} \\
&\geq E[-\exp(-\lambda(Q_3^d - C)F + C(P_2 - \varepsilon) - Q_3^d P_3^d)]
\end{aligned} \tag{91}$$

The certainty equivalence of this above expression is:

$$\frac{1}{N}(C(P_2-\varepsilon)-Q_3^dP_3^d-(1/2)\lambda(Q_3^d-C)^2var(F))+\frac{N-1}{N}[\frac{N-2}{N-1}(-1) \tag{92}$$
$$+\frac{1}{N-1}(Q_3^dP_3^d-(1/2)\lambda(Q_3^d)^2var(F))]$$
$$\geq (C(P_2-\varepsilon)-Q_3^dP_3^d-(1/2)\lambda(Q_3^d-C)^2var(F))$$

Substituting P_3^d and Q_3^d from equation (35) into (92) and rearranging the terms, we receive

$$P_2 \leq \frac{1}{4}\lambda var(F)C \tag{93}$$

Combining equations (90) and (93), the equilibrium price at the second stage when $N < N^*$ is

$$P_2 = \frac{1}{4}\lambda var(F)C$$

The expected utility that dealer i receives by quoting $P_{i,2} = P_2$ is greater than or equal to the one he receives by quoting $P_{i,2} > P_2$ or $P_{i,2} < P_2$. Therefore, at equilibrium, dealer i has no incentive to deviate. As a result, we have is symmetric equilibrium. Substituting the equilibrium price into the expected utility $E(U_i^d)|_{P_{2,i}=P_2}$, we can easily show that $E(U_i^d)|_{P_{2,i}=P_2} = -1$ at the equilibrium.

Proof of Proposition 4 when $N > N^*$

Symmetric equilibrium requires that the following two conditions are satisfied:

$$E(U_i^b)|_{P_{2,i}=P_2} \geq E(U_i^b)|_{P_{2,i}>P_2} \tag{94}$$

$$E(U_i^b)|_{P_{2,i}=P_2} \geq E(U_i^b)|_{P_{2,i}<P_2} \tag{95}$$

If $P_{2,i} > P_2$ and $P_{2,i} = P_2 + \varepsilon$ and given the expected utility in equations (48) and (50), condition (94) implies

$$\begin{aligned}
&\frac{1}{N}E[-\exp(-\lambda((-C+Q_3^b)F + CP_{2,i} - Q_3^b P_3^b - K)] \\
&+ \frac{N-1}{N}E[-\exp(-\lambda(D_{i,3}^b F - D_{i,3}^b P_3^b - K))] \\
&\geq E[-\exp(-\lambda(D_{i,3}^b F - D_{i,3}^b P_3^b - K))]
\end{aligned} \tag{96}$$

The certainty equivalence of the above expression is given by

$$\begin{aligned}
&\frac{1}{N}(-K + CP_{2,i} - Q_3^b P_3^b - (1/2)\lambda var(F)(-C+Q_3^b)^2) + \\
&\frac{N-1}{N}(-K - D_{i,3}^b P_3^b - (1/2)\lambda var(F)(D_{i,3}^b)^2) \\
&\geq -K - D_{i,3}^b P_3^b - (1/2)\lambda var(F)(D_{i,3}^b)^2
\end{aligned} \tag{97}$$

Substituting the equilibrium price and quantity at stage 3 in equation (37) into (97) and rearranging, we obtain

$$P_2 \geq \frac{2(N^2 - N - 4)(N-2) - (N-3)^2}{2[(N-1)^2 - 2]^2}\lambda var(F)C \tag{98}$$

If $P_{2,i} < P_2$ and $P_{2,i} = P_2 - \varepsilon$, and given the expected utility in equations (49) and (50), condition (95) implies that

$$\begin{aligned}
&\frac{1}{N}E[-\exp(-\lambda((-C+Q_3^b)F + C(P_2 - \varepsilon) - Q_3^b P_3^b - K)] \\
&+ \frac{N-1}{N}E[-\exp(-\lambda(D_{i,3}^b F - D_{i,3}^b P_3^b - K))] \\
&\geq E[-\exp(-\lambda((-C+Q_3^b)F + C(P_2 - \varepsilon) - Q_3^b P_3^b - K)]
\end{aligned}$$

The certainty equivalence of the above expression is given by

$$\frac{1}{N}(-K+C(P_2-\varepsilon)-Q_3^bP_3^b-(1/2)\lambda var(F)(-C+Q_3^b)^2)$$
$$+\frac{N-1}{N}(-K-D_{i,3}^bP_3^b-(1/2)\lambda var(F)(D_{i,3}^b)^2)$$
$$\geq -K+C(P_2-\varepsilon)-Q_3^bP_3^b-(1/2)\lambda var(F)(-C+Q_3^b)^2$$

Simplifying this inequality yields

$$N[-C(P_2-\varepsilon)+Q_3^bP_3^b+(1/2)\lambda var(F)(Q_3^b-C)^2]$$

$$\geq [-CP_2+Q_3^bP_3^b+(1/2)\lambda var(F)(Q_3^b-C)^2]+(N-1)[D_{i,3}^bP_3^b+(1/2)\lambda var(F)(D_{i,3}^b)^2]$$

Since the l.h.s. of this inequality is an increasing function of ε, this inequality holds if and only if $\varepsilon \to 0$. Setting $\varepsilon = 0$ and substituting P_3^b, Q_3^b and $D_{i,3}^b$ from equation (37) into this inequality, we obtain

$$P_2 \leq \frac{2(N^2-N-4)(N-2)-(N-3)^2}{2[(N-1)^2-2]^2}\lambda var(F)C \tag{99}$$

Combining equations (98) and (99) we obtain the equilibrium price P_2 when $N > N^*$.

Proof of the expected utility at the equilibrium when $N > N^*$

At equilibrium, the expected utility is

$$E(U^b)|_{P_{2,i}=P_2} = \frac{1}{N}E[-\exp(-\lambda((-C+Q_3^b)F+CP_2-Q_3^bP_3^b-K)]$$
$$+\frac{N-1}{N}E[-\exp(-\lambda(D_{i,3}^bF-D_{i,3}^bP_3^b-K))]$$

Substituting Q_3^b, P_3^b, $D_{i,3}^b$ and P_2 into the above equation, we obtain:

$$\frac{1}{N}[-\exp(-\frac{2N-2}{[(N-1)^2-2]}\lambda CE(F)-\frac{(N-1)(N-3)}{2[(N-1)^2-2]^2}\lambda^2 var(F)C^2+\lambda K)]$$
$$+\frac{N-1}{N}[-\exp(-\frac{(N-3)}{[(N-1)^2-2]}\lambda CE(F)-\frac{(N-1)(N-3)}{2[(N-1)^2-2]^2}\lambda^2 var(F)C^2+\lambda K)]$$

which yields

$$E(U_i^b)|_{P_{2,i}=P_2} = -\exp(-\lambda E(F)C - \frac{(N-1)(N-3)}{2[(N-1)^2-2]^2}\lambda^2 var(F)C^2 + \lambda K).$$

as equation (53).

Proof of Proposition 5

- Proof that H is decreasing in N for $N \geq N^*$.

We assume that N varies continuously. Recalling that dealers trade in the inter-dealer market when $N \geq N^* \geq 5$, we only need to consider the variation in H for $N \geq 5$.

Taking the partial derivative of the expected return function $H(N, \lambda, var(F), \sigma_C^2) = \frac{(N-1)(N-3)}{2[(N-1)^2-2]^2}\lambda var(F)\sigma_C^2$ with respect to N we obtain:

$$\begin{aligned}\frac{\partial H}{\partial N}|_{N\geq 5} &= \frac{4[(N-1)^2-2]^2(N-2) - (N-1)(N-3)[8((N-1)^2-2)(N-1)]}{4[(N-1)^2-2]^4} \\ &= \frac{-4[(N-1)^2-2][(N-1)^2(N-4)+2(N-2)]}{4[(N-1)^2-2]^4} = \frac{Num}{Den}\end{aligned}$$

where $Den > 0$, and $Num < 0$ for $N \geq 5$. Therefore, $\frac{\partial H}{\partial N}|_{N\geq 5} < 0$, and equation (60) is true.

Dealers enter the market as long as their expected return from market making is greater than or equal to the cost of brokerage fees. This is the case as long as (59), $H(N, \lambda, var(F), \sigma_C^2)|_{N\geq N^*} \geq K$, holds. First note that $H(N, \lambda, var(F), \sigma_C^2)$ is a continuously decreasing function in N for $N \geq 5$. It follows that as $N \to \infty$, the expected return H goes to zero. Hence, as long as the expected return function $H|_{N=N^*}$ evaluated at $N = N^*$ is greater than K, which is when $H(N, \lambda, var(F), \sigma_C^2)|_{N=N^*} \geq K$, there must exist one positive value $\widetilde{N} \geq N^*$ which satisfies the equality of $H(\widetilde{N}, \lambda, var(F), \sigma_C^2) = K$.

So far we have shown that $H(N, \lambda, var(F), \sigma_C^2)|_{N=N^*} \geq K$ is a sufficient condition for the existence of an equilibrium number of dealers greater than or equal to N^*. To prove the proposition, we also need to show that if (60) is true, then $H(N, \lambda, var(F), \sigma_C^2)|_{N=N^*} \geq K$ is also a necessary condition for the existence of an equilibrium number of dealers greater than or equal to N^*. That is, if $H(N, \lambda, var(F), \sigma_C^2)|_{N=N^*} < K$, then the equilibrium number of dealers greater than or equal to N^* does not exist. Suppose that N^* dealers enter the market. Given that H decreases in N, $\frac{\partial H}{\partial N}|_{N \geq N^*} < 0$,(60), the entry of an extra dealer lowers the expected return H, (that is, $H(N^* + 1) < H(N^*) < K$) Similarly we can show that $H(N^* + n) < H(N^*) < K$ for all n. Therefore, it follows that if (60) is true and $H(N, \lambda, var(F), \sigma_C^2)|_{N=N^*} \geq K$ is violated, an equilibrium number of dealers greater than or equal to N^* does not exist.

Proof of Corollary 6

To prove that the equilibrium number $\widetilde{N}$ is increasing in the variance of customer order flow, we need to show that:

$$\frac{\partial \widetilde{N}}{\partial \sigma_C^2}|_{N \geq 5} = -\frac{\partial H / \partial \sigma_C^2}{\partial H / \partial N}.$$

From Proporsition 5 we know that $\partial H / \partial N < 0$, and $\partial H / \partial \sigma_C^2 = \frac{(N-1)(N-3)}{2[(N-1)^2-2]^2} \lambda var(F) > 0$, and therefore, that $\frac{\partial \widetilde{N}}{\partial \sigma_C^2}|_{N \geq 5} > 0$.

Similarly,

$$\frac{\partial \widetilde{N}}{\partial var(F)}|_{N \geq 5} = -\frac{\partial H / \partial var(F)}{\partial H / \partial N} > 0,$$

$$\frac{\partial \widetilde{N}}{\partial \lambda}|_{N \geq 5} = -\frac{\partial H / \partial \lambda}{\partial H / \partial N} > 0,$$

and

$$\frac{\partial \widetilde{N}}{\partial K}|_{N \geq 5} = -\frac{\partial H / \partial K}{\partial H / \partial N} < 0.$$

Proof of Proposition 7

Suppose that the customer has a mean-variance utility function with a risk aversion of β given as:

$$E(-CP_2) - \beta var(P_2)C^2 \tag{100}$$

As long as the number of dealers who have entered the market is greater than or equal to N^*, the introduction of an inter-dealer broker market reduces the mean and variance of the price since $C > 0$. Hence, (100) increases and the customer is better-off. If less than N^* dealers enter market-making, then the introduction of brokered inter-dealer trading does not have an impact on customer welfare.

Now suppose that $\widetilde{N} < N^*$, and that the only way dealers can trade with each other is through the inter-dealer broker. Then the price expected by the customer is given by (64) and the volatility of the price is given by (65). Suppose now that direct inter-dealer trading is introduced. From Proposition 2, and given that $\widetilde{N} < N^*$, the expected price is given by (62) and the volatility by(63). It is easy to show (62) is lower than (64) and (63) is lower than (65) when $\widetilde{N} < N^*$. Hence, customer welfare improves given a choice between inter-dealer markets. However, the introduction of direct trading does not change customer welfare if $\widetilde{N} > N^*$.

Chapter 4

Alternative Trading Systems: Who Wins and Who Loses?

4.1 Introduction

Advances in information technology have fostered the emergence of new electronic trading systems for various financial assets. In foreign exchange markets, electronic interdealer brokers, such as Reuters dealing 2002-2 and EBS have replaced virtually all traditional voice broker systems as the means to transact brokered interdealer trading. Stock exchanges also face increased competition from various new electronic trading systems or platforms, also known as alternative trading systems (ATSs) or electronic communication networks (ECNs). This is the case particularly for US equities where nearly 40 per cent of dollar trading volume in Nasdaq shares have migrated to ATSs (Securities and Exchange Commission, 2000).

In explaining the competitive success or failure of various trading venues, it is useful to examine how these trading venues differ in terms of their market structure and design. Two equity ATSs, Instinet and Island, account for the largest proportion of trading of Nasdaq shares with 10.9 percent and 14.1 percent of trading volume, respectively, in year 2000. The fact that these systems offer an order-book trading environment, which is a good complement to the existing dealership system of Nasdaq, is an important factor behind their success (Financial Services Authority, 2000, pages 12-13).

However, ATSs that offer an order-book trading structure have failed to become a significant force in the *customer-dealer trading* segment of the foreign exchange (FX) or bond market, where the vast majority of public trading continues to take place in a dealership trading environment. However, in the *inter-dealer trading* segment of FX or bond markets, order-book style ATSs such as Reuters Dealing 2000-2 and EBS in the FX market and EuroMTS in the European bond market, account for a significant proportion of the trading activity in these markets (Bank for International Settlements, 2001).

To determine the success of an ATS or how different trading venues succeed in capturing trading activity in a particular security, one has to understand which attributes of the trading venue are most important for trading activity. Moreover, the ability of different trading mechanisms to offer lower trading costs to an investor depends critically on the trading needs or characteristics of the investors themselves. Given that investors seek best execution of their trades, and that best execution encompasses traded price, market impact, immediacy, timing, anonymity and commissions, then it is not surprising to see investors choose different trading venues based on how well each venue fulfils different combinations of these aspects of execution quality.

Canadian regulators recently considered rules that allow customers in the government securities dealership market to trade among themselves through a centralized electronic system rather than trade through dealers. In other words, they propose to introduce an order-book style ATS for the public trading segment of the fixed income markets. Motivated by these recent regulation developments, this chapter seeks to investigate which market structure a customer would choose *ex ante*. We examine what determines a market participant's choice of trading in an order-book ATS versus trading in a dealership market. Moreover, we seek to analyze how the trading environment and how customer trading characteristics affect their choice of trading mechanism.

Despite the fact that new trading systems have emerged and attracted significant trading activity, little theoretical or empirical work examines the considerations that determine the participant's choice of market mechanism. That investors are observed to send their orders to different trading systems is not the focus of much research. Most of the existing research concentrates on modelling one trading mechanism and often simply examines mechanism efficiency in terms of price discovery. O'Hara (1995) provides a survey of continuous order-book and dealership models. However, given that each surveyed study used different modelling techniques, it is generally difficult to compare their results.

To study the optimal market structure preferred by market participants, this chapter constructs an agent-based computational model of both a dealership and limit-order market.[9] This methodology starts where theoretical market microstructure models leave off, in that it allows researchers to examine questions that are analytically intractable. The agent-based simulation methodology is structurally grounded by an analytical model that guides the behavior of the artificial or simulated traders. A secondary objective of this chapter is to illustrate the applicability of the artificial financial markets approach for the study of market design issues.

---

[9] See Chapter 1 for a review of agent-based modelling.

The benefit of using an agent-based method is that this method allows us to investigate questions related to optimal market structure in a unified analytical framework. Another benefit of the approach is that it allows for greater flexibility relative to analytical models, both in terms of how we specify the trading environment and how we model the characteristics of the market participants themselves. For example, most market microstructure models assume that agents are limited to using linear pricing strategies to in order to derive analytic solutions. The agent-based simulation approach allows us to relax this restriction.

We compare the welfare of rational public investors who trade in a *dealership market* with those who trade in a *limit-order book* style ATS. In a *dealership market*, utility maximizing public investors (customers) submit orders to dealers based on their liquidity shock and the competitiveness of dealers' quotes. Given the opaqueness of this market, customer order flow is private information to dealers, which can be further explored by dealers by trading in an inter-dealer market. In contrast, in a *limit-order market,* customers trade directly among themselves without the intermediation of dealers. Customers submit limit orders to the market based on the liquidity shocks they receive just prior to entering the market. In essence, customer trading is motivated by the desire to share liquidity risk among a greater set of participants. Given that customers are identical except for the realization of their liquidity shocks, the degree of customer heterogeneity is defined over their liquidity shocks. Adjusting the distribution from which the liquidity shocks are drawn allows us to vary customer characteristics related to order size. Customers are also characterised by the correlation of their trading needs. When this correlation is high, customer orders tend to be on one side of the market.

By varying these customer characteristics, we examine a range of market structure issues. Our study investigates how the optimal market structure for customers whose order size is sometimes large and varies considerably may differ from that for customers characterised by relatively small homogenous order flow.

Another concern of this study is to examine customer choice of market architecture under different trading environments. The trading environment is defined over the thickness of the market, the number of market makers, the degree of market maker heterogeneity, and the risk aversion differential between market makers and customers. Different market structures may be better suited to overcome various coordination or trading frictions in financial markets. For example, a feature of fixed income and to a somewhat lesser extent FX markets, which differs markedly from equity markets, is their thickness. The number of buyers and sellers trading in the market at any point in time are much less than in equity markets.[10] As such, the observation that these markets are relatively thin might explain why the public trading segment of fixed income and FX markets are structured as dealership systems.

Our findings suggest that the trading environment has an important impact on the optimal market structure. The public investor's choice to trade in a particular market depends on thickness of the market measured in terms of the number of customers active in

[10] The fixed income market is largely a wholesale market consisting of a relatively small number of large institutional investors. On the other hand, equity markets consist of thousands of traders, a large proportion of which are small retail investors.

the market within a short time span. The dealership market structure is preferred by customers when there are few customers that are potentially available to trade. As the number of potential public investors increases, there comes a point where the number of investors exceeds a critical threshold, so that these investors prefer to trade in the order-book market structure.

We find that as the number of dealers decrease or as the risk appetite of the dealers decreases relative to that of the customers, the risk-bearing capacity of the dealership market decreases. This makes this market structure less attractive to liquidity motivated public traders. As a result, the critical number of customers necessary for the limit-order book market to prevail decreases.

Customers who are subject to larger and more volatile liquidity shocks prefer to trade in a dealership system. Specifically, customers characterised as generators of large-sized order flow, which is a trait of institutional investors, are more likely to submit their orders to dealership systems than is the case for small-sized, homogeneous order suppliers. This is consistent with the observed regularity in capital markets where markets dominated by a small number of large institutional investors tend to be organized as dealership markets.

The chapter is organized as follows. Section 2 gives a literature review. The model specification of the trading environment is presented in section 3. Section 4 reports our computational findings. Section 5 summarizes the key findings and reviews some of the limitations of the current framework. Section 6 provides a technical appendix.

4.2 Literature Review

A small but growing number of papers study issues related to optimal market design by modelling a trading mechanism, and then adjusting this trading mechanism or trading protocol, to create a modified trading structure. Various authors recognize the importance of market design and explain the coexistence of markets with differing levels of transparency (e.g., Lyons 1996, Flood *et al.* 1999, Bloomfield and O'Hara 1999, and Naik *et al.* 1999). In general, these studies show that a trade-off exists between market quality, measured in terms of liquidity and implicit trading costs, and market efficiency, measured in terms of effective price discovery. They show that this trade-off may explain why market mechanisms with differing levels of price transparency coexist.

Vogler (1997) develops a theoretical model that compares dealership markets to auction markets in the absence of asymmetric information. The key and realistic feature of the Vogler model is that the price the dealer offers to the liquidity motivated public investor (in the first stage of trading) depends on the dealer's ability to manage his inventory position in the second stage, where she is assumed to trade via an auction mechanism. Vogler shows that a dealership market provides customers with a better price than an auction market if customer trading is motivated only by liquidity shocks. The result holds as long as the market maker can rebalance his inventory in the inter-dealer market. As noted by Vogler, the Walrasian market-clearing mechanism assumed in both the interdealer trading or second stage of the multi-dealer market and in the auction market is a major limitation of this model. the reason is that this mechanism assumes away the probability with which some orders are not executed in an order-book market. Moreover, in Vogler (1997), as is the case for most of the literature, customers are exogenous liquidity traders. By endowing each customer with a utility function and endogenizing his decision making, we can analyze the determinants of the optimal strategies of customers.

As in Vogler (1997), Naik, Neuberger and Viswanathan (1999) develop a two-stage theoretical model that resembles most multi-dealer markets. They examine how customer welfare is affected under two trade-disclosure regimes, one where post-trade information is made public and one where it is not. An additional innovation over the Vogler (1997) study is that Naik *et al.* (1999) model market makers in addition to investors as being risk averse. In other words, Naik *et al.* endogenize the investor's trading strategy which is conditional on the price schedules of dealers. Naik *et al.* show that dealership markets requiring full and prompt disclosure of investor-to-dealer trade details may reduce the welfare of the risk averse investors in some circumstances. The authors also show that an increase in the level of post-trade transparency works against the execution of large trades but nevertheless tends to benefit small traders.

Recent theoretical work by Viswanathan and Wang (2002) considers the comparison between limit-order books and dealership markets abstracting from asymmetric information. They examine the determinants of customer choice for a trading venue, and how customer welfare change across three market structures: limit-order book, dealership and hybrid. The hybrid market structure is a combination of both the order-book and dealership structures. The authors model these three different structures using three different market-clearing mechanisms: discriminatory price auction, uniform price auction and a hybrid of these two auction mechanisms. They show that risk neutral customers always prefer to participate in a discriminatory price auction, that is, the order-book structure. However, a uniform price auction is preferred by risk averse customers when the number of market makers is sufficiently large. They also find that the hybrid market structure can dominate the uniform and discriminatory price auction for risk averse customers.

Our study differs from Viswanathan and Wang in several important aspects. First, Viswanathan and Wang (2002) use a one-sided auction methodology where all dealers are buyers of the security and all customers sellers (or the symmetric opposite of this). That is, the dealership and order-book market structures are restricted to being one-sided markets, very reminiscent of the primary market for government securities. In the standard theoretical set-up, each dealer posts a bid (offer), while an individual customer simply tenders an amount she seeks to sell (buy). We develop a double auction market-clearing framework, which is much closer to real order-book market structures. Second, the use of a double-sided auction framework implies that customers trade with each other in the order-book market and can be both suppliers and demanders of liquidity. In the Viswanathan and Wang (2002) one-sided auction framework, customers can trade only with dealers in both the dealership and order-book markets. Third, unlike Viswanathan and Wang, in modelling the dealership market, we model inter-dealer trading where dealers trade among themselves after receiving customer orders.

4.3 Model Description

The model assumes the existence of two types of agents, customers and dealers. Both possess a portfolio composed of a risk-free and a risky asset. The liquidation or fundamental value of the risky asset is normally distributed and the distribution is common knowledge across all the participants. We assume there is no opportunity cost for holding the risk-free asset.

Moreover, all agents adjust their portfolios to optimize their utility. Dealers are obliged to make the market and thus provide liquidity for customers. The two agent types differ only in terms of their objective function, the roles that they play on the market and the market mechanisms they can access to rebalance their portfolio. This framework flexibility allows us to use comparative statics to compare different types of market organization with respect to the welfare of the participants and the liquidity of the market.

In the following section, we describe the models for two different markets, namely, a limit-order book market and a dealership market.

4.3.1 Limit-Order Book Market Model

Trading Environment

Assume that M customers (public investors) trade a risky asset in the market. The customers are identical in every way except for the realization of their exogenous liquidity shocks $(X_i)_{i \in \mathcal{M}}$, that are positive or negative. For example, if an agent receives a negative demand shock on his risky asset position, he tends to rebalance his portfolio by posting a buy order for the risky asset. The random liquidation value of the risky asset F is independent and normally distributed, and the distribution is common knowledge to all the participants.

Consider the following sequence of events. First, each customer receives a liquidity shock in units of X_i, which is non-payoff related information. These shocks initially are uncorrelated across customers. Each customer then simultaneously submits a bid and an ask price-quantity pair, (B_i, Q_i^B) and (A_i, Q_i^A) respectively, conditional on his realization of the shock. A bid quote indicates that the customer agrees to buy Q_i^B units of the risky asset at the market price if the market clearing price is below his bid, B_i. The ask quote indicates that the customer agrees to sell at the market price up to Q_i^A units of the security if the market-clearing price is above his ask, A_i. Each customer knows the distribution from which inventory shocks are drawn, but they do not observe the actual shocks of other customers.

The market clearing price at which indicated trades are executed is determined via a double-auction system. Since the market clearing price is determined after the customers have submitted their orders, each customer realizes that their bid-ask orders may influence the clearing price to the extent that they are the marginal buyer or seller in the auction. Finally, the value of the risky asset is revealed to the market participants. They can evaluate their wealth and utility.

The problem faced by each customer is to find a strategy which maximizes his expected utility. A strategy is a function which maps the random shock X_i to a bid and ask price-quantity pair, (B_i, Q_i^B) and (A_i, Q_i^A).[11] Each market participant has exponential utility functions defined over wealth W with risk aversion coefficient λ. A liquidity shock

[11] Customers generally only act on one side of the market or the other. Allowing customers to provide two-side quotes is easier to handle numerically. A customer who needs to buy can always manage to post a competitive bid price and a less competitive ask price to increase/decrease the chance of execution of his buy/sell order.

$(X_i)_{i \in \mathcal{M}}$ on his risky asset position is independent and normally distributed.[12]. The utility function of customer i is given by:[13]

$$U_i = -\exp\left(-\lambda W_i\right), \tag{101}$$

The wealth of customer i can be written as a summation of the dollar value of his risky position and his cash position, so that,

$$W_i = (X_i + Q_i)F - PQ_i, \tag{102}$$

where Q_i is the executed quantity of the risky asset, and P is the market clearing price of the risky asset. In the next section, we discuss how this price is determined.

Market Clearing System: Double Auction

We model the limit-order book market as a double auction (DA). The standard setup of a DA is fairly restrictive since order size is fixed exogenously (Werner 1997, Rustichini, Satterthwaite and Williams 1994). The double auction setup is generalized herein to allow each order size to be endogenously decided by market participants. This market clearing mechanism has some intuitive appeal as the quantities demanded for prices above the clearing price are sequentially matched with quantities supplied at prices below this price, until no orders exist at which the purchase price exceeds the selling price. In other words, orders

[12] The normality assumption on F is necessary for deriving an analytical solution for a customer's trading strategy in a 2-stage dealership market. The normality assumption on the inventory shock, X_i , is not necessary for obtaining similar simulation results. The numerical method that we develop in this chapter is general enough to accomodate different distributions for the inventory shock.

[13] We choose the negative exponential utility function because it is the most widely used utility form in this literature due to its analytical tractability.

Table 9. Order Matching and Trading Outcome

participants	liquidity shock in units	bid order (P, Q)	ask order (P, Q)	quantity executed	wealth at F=1
player 1	-0.5	(\$0.7, 2)	(\$0.8, -1)	2	0.5
player 2	0.6	(\$0.4, 2)	(\$0.5, -3)	-1(partial)	0.1
player 3	0.4	(\$0.1,0.5)	(\$0.2, -1)	-1	-0.1

are matched sequentially until the aggregate demand curve lies below the aggregate supply curve. This auction clearing system implies that all trades clear at the same price.[14]

Table 9 gives an example of a trading outcome for a market with a DA clearing system that is comprised of three participants. A negative shock in quantity indicates a sell order. The distribution of the liquidation value, $F\tilde{}N(0,1)$, is the common knowledge among three players.

Each participant starts with a zero endowment of cash and a zero amount of a risky asset.[15] Conditional on the initial liquidity shocks (in units) for players 1, 2 and 3 given in the second column, each customer chooses an action according to his strategy function. An action consists of a bid price-quantity pair and an ask price-quantity pair. For example, after receiving a negative liquidity shock, participant 1 tends to rebalance his portfolio by posting a greater buy order. Actions taken by other players are in columns 3 and 4. The incentive for players 2 and 3 to place sell orders is that each of them receives a positive liquidity shock. The market clearing price in a double auction mechanism, therefore, is where the demand curve crosses the supply curve or \$0.5. All the bids which are higher than (or equal to) \$0.5, and all the ask prices which are below (or equal to) \$0.5 are executed. The

14 A real world example of such a call auction is opening of the NYSE.

15 In this example, we initialize the endowment to zero just for simplicity. Borrowing and short selling are allowed when it is necessary.

allocation is carried out as far as possible by assigning priority to sellers whose offer prices are smallest and to buyers whose bid prices are largest. Partial execution may exist at the market clearing price. For this example, participant 2 has a partial execution of his ask order, that is, he sold one unit of his risky asset at a price of 0.5. Therefore, his risky asset position is $0.6 - 1$ and his cash position is $\$0.5 \times 1$. Thus the wealth of player 2 is $W_2 = (0.6 - 1) \times 1 + 0.5 \times 1 = 0.1$. The wealth of each participant can be evaluated as shown in the last column.

A market clearing price in a variable order size DA can be formally defined as follows. We are searching for price P such that excess demand is positive (negative) for all the prices that are less (greater) than P. Furthermore, to obtain a unique solution even if multiple prices solve the excess demand, we redefine an excess demand function as the following:

$$XD(p) = \sum_{i=1}^{M} Q_i^B \, I_{(p \le B_i)} - \sum_{i=1}^{M} Q_i^A \, I_{(A_i \le p)} \tag{103}$$

where $I_{(p \le B_i)}$ and $I_{(A_i \le p)}$ are indicator functions which are equal to one if the condition is true and zero. The market price P is the root at which the excess demand is equal to zero otherwise.

The Strategy Function of Customers

Since it is very difficult to assume the functional form of the mapping between a liquidity shock and an action, we use the following flexible form *Artificial Neural Network*

(ANN) for an individual's strategy function:

$$ANN(X_i, \theta_i) = \theta_{i,0} + \sum\nolimits_{k=1}^{H} \theta_{i,k}\, \varphi\left(\theta_{i,(H+1)+k} + \theta_{i,(2H+1)}\, X_i\right) \tag{104}$$

Therefore, the input of this strategy function is liquidity shock X_i. The output of this ANN is a vector of four choice variables $\left(B_i, A_i, Q_i^B, Q_i^A\right)$, and a strategy function is characterized by the vector of parameters θ_i which contains $\theta_{i,0}$, and $\theta_{i,k}$. H is the number of hidden nodes. The *activation function* φ is chosen as the tanh function:

$$\varphi\left(x\right) = \tanh\left(x\right) \tag{105}$$

We slightly modify the first two choice variables, B_i and A_i, to the mid-point of the bid and ask price, $\frac{B_i+A_i}{2}$, and the spread of the bid and ask, $A_i - B_i$. This modification is to guarantee that a ask price is always higher than a bid price. The resulting specification for each market participant's strategy function is:

$$\begin{bmatrix} \frac{B_i+A_i}{2} \\ A_i - B_i \\ Q_i^B \\ Q_i^A \end{bmatrix} = \begin{bmatrix} ANN\left(X_i, \theta_i^{mid}\right) \\ \phi\left(ANN\left(\theta_i^{spread}\right)\right) \\ \phi\left(ANN\left(X_i, \theta_i^{Q^B}\right)\right) \\ \phi\left(ANN\left(X_i, \theta_i^{Q^A}\right)\right) \end{bmatrix} \tag{106}$$

Therefore, θ_i is $\left(\theta_i^{mid}, \theta_i^{spread}, \theta_i^{Q^B} and \theta_i^{Q^A}\right)'$. $\theta_i^{mid}, \theta_i^{spread}, \theta_i^{Q^B}, \theta_i^{Q^A}$ are parameters in vector θ_i that are used to generate the mid point of bid and ask, spread, and bid depth and ask depth, respectively. Notice that ϕ is applied to three outputs of the ANN to guarantee that the bid-ask spread and the bid and ask depth are positive. The function ϕ is given by $\phi\left(x\right) = \log\left(1 + \exp\left(x\right)\right)$.

Figure 1 depicts the bid and ask for a typical customer as a function of his liquidity shock X_i in his risky asset position. At the beginning of each period, customers receive different liquidity shocks, and therefore post different bid and ask prices. The realizations of market clearing prices, which are determined using equation (103), are represented by the dots in the middle part of the graph. The upper and lower curves are ask and bid prices for given liquidity shocks for a representative customer. Since an order can be executed only if the ask (bid) price is greater (lower) than the market clearing price, the figure shows that a customer tends to post a low ask price to try to sell part of his inventory for rebalancing his portfolio if hit by a positive liquidity shock.

4.3.2 Dealership Market

Trading Environment

This section presents a trading model with M customers indexed by j and N dealers indexed by i. They all agree on the distribution of the final value of the risky asset. Trading takes place in two stages. In the first stage, dealers trade with customers by posting a bid and an ask quote simultaneously prior to the arrival of customer orders for any order size. Given a liquidity shock X_j and after observing the dealer's quotes, the customer chooses the quantity to trade with dealers in an attempt to maximize his expected utility. Customer order flow is price elastic in the sense that it decreases with the bid-ask spread that dealers quote. The trade size between the customer and the dealer is private information. Therefore, customer order flow determines the distribution of inventories among dealers at the end of the first stage.

In the second stage, dealers trade strategically in the inter-dealer market to reallocate any excess inventories obtained during the first stage. Dealers can submit a bid quote and an ask quote through a double-auction system. The market clearing price in the second stage is determined in the same manner as for an order-book market. Orders are matched sequentially starting from the highest bid order to the lowest ask order. This matching process continues until no more orders exist for which the bid price is higher than or equal to the ask price. In the end, a fundamental value F is revealed and every dealer's wealth is evaluated according to their trading outcomes and the fundamental value of the risky asset.

In the first stage or public trading round, each customer does not observe quotations from *all* the N competing dealers for the following reason. After observing the quotes, the customer needs to decide whether to split his order between dealers or to trade the whole order as a block to the one dealer who quotes the best price. If all dealers start with the same zero inventory position, allowing customers to split the order equally among all the dealers results in an optimal risk allocation right away. In this case, no dealer has an incentive to trade in the inter-dealer market in the second stage. For most of our experiments, we allow customers to observe quotes from *different subsets* of N competing dealers and to trade only with the dealer who offers the best observed price without splitting his order. In fact, it is a defining characteristic of dealer markets that dealers compete for the whole order from a customer (Franks and Schaefer, 1992). However, allowing for different initial dealer inventories still motivates inter-dealer trading even if the customer splits his order equally. We explore this variation in one of our experiments.

Customer Strategy in Stage 1

In this section we solve for the analytical solution of a customer's optimal strategy. In the first stage, each customer j observes a set of quotes $(B_i^1, A_i^1)_{i\in\Delta_j}$ randomly drawn from a set of $\mathcal{K}$ dealers, denoted as Δ_i, and decides the quantity C_{ij} that he wishes to trade and with which dealer i. The total amount traded by the customer j is denoted as $C_j = \sum_{i\in\mathcal{K}} C_{ij}$. Note that the maximum bid of that subset is $\overline{B}_j^1$ and the minimum ask is $\underline{A}_j^1$, and the corresponding dealers are δ_j^A and δ_j^B.

Each customer j decides upon the quantity C_j that he wishes to trade in order to maximize the expected utility of wealth over C_j, where utility is given by:

$$MaxE(U_j^C) = -\exp\left(-\lambda^C W_j^C\right) \tag{107}$$

Wealth is given by

$$W_j^C = \begin{cases} X_j F + C_j\left(F - \underline{A}_j^1\right) & if \quad C_j > 0 \\ X_j F & if \quad C_j = 0 \\ X_j F + C_j\left(F - \overline{B}_j^1\right) & if \quad C_j < 0 \end{cases} \tag{108}$$

where $C_j > 0$ ($C_j < 0$) represents a purchase (sale). This equation states that customer i's wealth depends on the sign of his order. He buys(sells) units of C_j at the best ask price, $\underline{A}_j^1$ (or the best bid price $\overline{B}_j^1$), or he does not trade if $C_j = 0$.

X_j is known at the moment C_{ij} is decided, and F follows a normal distribution with mean 0 and variance σ_F^2. Therefore, the optimal value of C_{ij} can be solved using

$$C_j = \begin{cases} \dfrac{-\overline{B}_j^1}{\lambda_j^C \sigma_F^2} - X_j & if \quad P_j^R < \overline{B}_j^1 \\ 0 & if \quad \overline{B}_j^1 < P_j^R < \underline{A}_j^1 \\ \dfrac{-\underline{A}_j^1}{\lambda_j^C \sigma_F^2} - X_j & if \quad \underline{A}_j^1 < P_j^R \end{cases} \tag{109}$$

where the reservation price P_j^R of customer j is given by

$$P_j^R = -\lambda_j^C \sigma_F^2 X_j. \tag{110}$$

Therefore, the optimal value of C_{ij} is

$$C_{ij} = \begin{cases} C_j & if \quad i = \delta_j^B \text{ and } C_j < 0 \\ C_j & if \quad i = \delta_j^A \text{ and } C_j > 0 \\ 0 & otherwise \end{cases} \tag{111}$$

This is the analytical solution for the optimal strategy of a customer in a dealership market setting. In later simulations, each customer acts according to the optimal strategy described in this equation.

Dealer's Strategy in Stages 1 and 2

Dealers provide one bid-ask quote (B_i^1, A_i^1) upon a customer's request, and then receive the customer's order flow at the prices they quote. In the second stage, dealers quote a bid and ask price-quantity pair, $\left(B_i^2, A_i^2, Q_i^B, Q_i^A\right)$, and trade in an inter-dealer market to share their inventory risk with each other. The objective of a dealer is to maximize his utility over these six choice variables. Using notation that is similar to that used for the order book market, we write the objective function of dealer i as:

$$MaxU_i^D = -\exp\left(-\lambda^D W_i^D\right) \tag{112}$$

subject to

$$W_i^D = \sum_{j \in \mathcal{M}} C_{ij} \left(I_{(C_{ij}>0)} A_i^1 + I_{(C_{ij}<0)} B_i^1 \right) - Q_i P + F \left(Q_i - \sum_{j \in \mathcal{M}} C_{ij} \right). \quad (113)$$

For the strategy of stage 1,(B_i^1, A_i^1), a single parameter is used for the bid-ask spread, $A_i^1 - B_i^1$. The mid-point of the bid-ask spread, $\frac{B_i^1 + A_i^1}{2}$, is a function of the initial liquidity shock of dealers, which is set at zero in the baseline experiment we examine. This constraint guarantees that we will have $\underline{A}_j^1 > \overline{B}_j^1$for any customer, and therefore that no arbitrage opportunities are available to customers. Since the inter-dealer market is modelled as a double auction limit-order book market in Stage 2, a dealer's trading strategy is identical to the one described earlier in the order book market. A dealer's strategy function is summarized by the following equation:

$$\begin{bmatrix} \frac{B_i^1 + A_i^1}{2} \\ A_i^1 - B_i^1 \\ \frac{B_i^2 + A_i^2}{2} \\ A_i^2 - B_i^2 \\ Q_i^B \\ Q_i^A \end{bmatrix} = \begin{bmatrix} ANN\left(X_i^D, \theta_i^{mid1}\right) \\ \phi\left(ANN\left(\theta_i^{spread1}\right)\right) \\ ANN\left(C_i, \theta_i^{mid2}\right) \\ \phi\left(ANN\left(\theta_i^{spread2}\right)\right) \\ \phi\left(ANN\left(C_i, \theta_i^{Q^B}\right)\right) \\ \phi\left(ANN\left(C_i, \theta_i^{Q^A}\right)\right) \end{bmatrix} \quad (114)$$

Dealers use the inter-dealer market to share the inventory risk they assume in the public trading round. The price that a dealer quotes to a customer in the first stage depends on the risk sharing opportunities available in the second stage.

4.3.3 Numerical Method to Search for a Nash Equilibrium

The market participant's strategy function maps information to an action, such as a bid-offer schedule $\left(B_i, Q_i^B, A_i, Q_i^A\right)$. The information used by participant i is his liquidity shock X_i. Therefore, the strategy function s_i is a mapping from X_i to $\left(B_i, Q_i^B, A_i, Q_i^A\right)$.

A *best response* of participant i is a strategy that maximizes participant i 's utility. Given his initial liquidity shock and other participants' strategies, if we can find a set of strategies where every participant plays his best response, then this set of strategies is called a *Nash Equilibrium* for the auction.

We propose a numerical method to solve for the Nash Equilibrium. The method is described in detail in the Technical Appendix. Here we only discuss one necessary condition to apply this numerical method.

The necessary condition to apply the numerical method is that the utility function is continuously differentiable. The utility function is continuous in wealth but not with respect to the parameters of the strategy function.

For example, in this double-auction game, a customer's order is not executed, when his bid price lies below the market clearing price. As a result, there is no change in his utility function. As he slightly increases his bid, nothing happens to the utility function until the moment his bid is greater than the market clearing price. At that moment, this customer becomes a winner of the auction, and his order is executed. Therefore, this customer observes a discrete change in his utility function.

This discontinuous feature of the payoff function prevents the use of usual optimization methods. We develop a smoothing method to modify the payoff function such that the payoff function is differentiable. When the smoothing parameter approaches a limit, the modified payoff function converges to the original discrete payoff function. In our simulation, the criteria of convergence is that the difference of the two functions is 10^{-5}. Using this smoothing method, each participant receives a proportion of the auction goods, and the proportion decreases as the difference between the participant's bid and the winning bid increases. This proportion is controlled by the smoothing parameter. The greater the parameter, the larger proportion of the goods the winning bid gets. Therefore, the winning bidder gets the whole auction at the limit, since the payoff function converges to the original step payoff function. (See section 4.6.2 of the Technicial Appendix for more details on the smoothing method.)

4.4 Simulation Results

4.4.1 Parameters of trading environments and customer attributes

To determine the preferred market structure of customers, we compare the expected utility that the customer achieves under two market structures, and across trading environments and customer characteristics. We vary the number of customers M as a key parameter for the change of trading environment. The various trading environments that we consider are defined over a set of initial parameter values.

We first discuss the motivation behind the selection of each parameter value given in Table 10. The third column in Table 10 presents the parameter values for the baseline or benchmark trading environment. The parameter values in the last column are those used to generate modified trading environments. Note that in generating a modified trading environment or for changing customer attributes, we adjust only one parameter at a time, while keeping the other parameters at their benchmark values. Based on Table 10, the realized liquidation value of the security is assumed to be drawn from a normal distribution

Table 10. Parameter Configuration

Panel A: Trading Environment Configuration			
Parameters	Definition	Baseline	Modified
σ_F^2	variance of liquidation value	0.7	
γ	risk aversion differential $(\lambda_c - \lambda_D)$	0	0.5
N	number of dealers	6	12
$\sigma_{X^D}^2$	variance dealer inventory	0	0.7
Panel B: Customer Attributes			
Parameters	Definition	Baseline	Modified
$(\sigma_{X^c}^2, \rho)$	variance and correlation of inventory shock of customers	(0.7,0)	(1,0) (0.7,0.3)

with mean zero and variance, σ_F^2, which is equal to 0.7 throughout our analysis. Let γ defines the relative difference between the levels of risk aversion of the customers and the dealer. In the baseline trading environment, we assume that customers and dealers are equally risk averse, with the aversion value from their expected utility functions set to 1, so that $\gamma = 0$. That is, in most trading environments considered, $\gamma = 0$. However, we consider trading environments where γ=0.5, which implies that dealers are less risk averse than customers. This trading environment is meant to capture specialization of dealers in market making, which implies a superior ability to take on inventory risks.

The trading environment also is adjusted by changing the number of market makers, N. By adjusting the number of market makers, we can examine the implication of changing market capacity. Specifically, we set the number of market makers, N, equal to 6 in the baseline configuration, and to 12 in an alternative configuration of the trading environment.

The degree of dealer heterogeneity can also affect the trading environment (Dutta and Madhavan 1997). We allow dealers to be either homogeneous or heterogeneous in terms of their initial inventory positions prior to trading with customers. Dealer inventory heterogeneity is intended to capture inventory heterogeneity documented by Reiss and Werner (1998). It is also intended to proxy for dealer inventory disturbances incurred during the dealers' public trading in prior periods. Initial dealer positions for dealers $i = 1, ..., N$, are drawn from a normal distribution with mean zero and variance $\sigma^2_{X^D}$= 0.7 when dealers are heterogeneous, and $\sigma^2_{X^D}$ = 0 when dealers are assumed to be homogeneous.

To examine how customer attributes affect customer choice of market structure for a given trading environment, we adjust the characteristics of the liquidity shock that customers receive. To model large versus small customers, we adjust the variance of the distribution from which the liquidity shocks are drawn. Recall that customers are motivated to trade because the liquidity shocks they receive requires them to rebalance their portfolios. The larger the liquidity shock, the greater the need for portfolio rebalancing, and the greater the order size that the customer submits to the trading system. As a result, customer liquidity shocks for customer i =1,..., M, are drawn from a normal distribution with mean zero and variance, σ^2_X = 0.7, when customers are characterized as small, and σ^2_X = 1 when customers are assumed to be large. We vary the value of number of customers, M, to represent different thickness of the market.

The assumption that the initial liquidity shock X_i is independent across customers does not appropriately capture the stylized facts observed in fixed-income or foreign exchange markets, where there is a tendency for customer orders to arrive in bunches (Gravelle 2002). Anecdotal evidence indicates that periods of one-sided order flow occur relatively more frequently in fixed income and FX markets than in equity markets. These order-flow dynamics likely reflect the arrival of public information about the fundamental value of the security. Given that the data release is observed by all market participants, customers tend to trade in the same direction to adjust their portfolios. In our one-shot game construct, we proxy this order bunching effect by allowing liquidity shocks, $X_i, i = 1, 2...M$, to be correlated across the customers in some experiments. Given that customer order submission strategies are in part driven by their realized liquidity shocks, making these shocks correlated across customers induces some degree of correlation in customer order flow. We set the correlation between customers liquidity shocks to ρ=0.3 to engender a correlated order flow. The baseline setting is ρ=0, which implies that liquidity shocks are independent across customers and order flow is on average uncorrelated across customers.

4.4.2 Welfare Results

In this section we compare the welfare of customers for two different market structures: a DA order-book market and a dealership market. In an order-book market, without the presence of a dealer, customers receive liquidity shocks and then trade directly among themselves to share the risk. In a dealership market, dealers always stand ready to provide liquidity to customers and charge a bid-ask spread for the service they provide. In a dealership market, customers trade with dealers after receiving their inventory shocks, and then dealers trade among themselves to share the inventory risk resulting from the public trading round.

Tables 3 and 4 report customer welfare for each experiment. Each cell in these tables represents one experiment based on 1000 realizations for each normally distributed parameter, F and X. Customer welfare reported in each cell is the average across M customers, and the mean of the utility level achieved over the 1000 draws of the random variables for each customer. Since we use a negative exponential utility function, which is defined in equation (101), smaller negative values in the cells of the table represent an improvement in welfare.

Figure 2 shows the improvement in a typical customer's welfare through trading in an order-book market. The lower curve is his reservation utility in case of no-trading. The black dots are the realizations of his utility after trading at the given shocks, and the upper curve represents the fitted values for all the realizations of the utility. Since the fitted value lies above the reservation utility curve, we say that on average, customers' welfare is improved after trading.

Each row of these tables corresponds to a different trading environment, one where the trading environment changes with the given number of customers, M. The second and third columns in Table 11 correspond to welfare in the baseline configurations of the order book and the dealership market structures, respectively. All modified trading environments are compared to these benchmark results.

From Table 11 for the baseline configuration, it is clear that increasing the number of customers in the market implies improved customer welfare in the order-book system.

Increasing the number of customers increases competition for order flow, which reduces the bid-ask spread, and increases the probability of an execution. An alternative way to think about this is to recall that large values of M imply a trading environment characterized by a high order arrival rate. Given that customers seek to rebalance their portfolio by trading with a larger set of customers, larger values of M imply an increase in the chance of finding potential counterparts in the order-book setting. This enhances the avenues for risk diversification for customers, which raises their expected utility.[16]

In contrast, in a dealership market structure with a fixed number of market makers, an increase in the number of customers leads to lower customer welfare (see column 3 of Table 11). An increase in the number of customers imposes greater inventory absorbing pressure on dealers. In turn, dealers charge wider bid-ask spreads to compensate for greater inventory risk that they have to bear. As a result customer welfare falls.

The risk sharing feature of inter-dealer markets can be illustrated by Figure 3. The 45 degree line is the position that dealer 1 has after trading with customers. It is the initial position that dealer 1 has when he enters the second stage or inter-dealer trading round. Dots in the graph are the realized final positions after inter-dealer trading, and the curve in the middle is the average of the final positions. The graph shows that the average final position moves away from the initial position and towards the zero line. If dealer 1 starts with a large positive (negative) position, he is able to sell (buy) some of his risky asset in the inter-dealer trading round. Dealers share their inventory risk by trading among themselves.

[16] See Vogler (1997) and Rustichini, Satterthwaite and Williams (1994) who show that the double auction outcome quickly converges to the standard Walrasian equilibrium (i.e., the competitive equilibrium) as the number of players increases. This is the market structure that provides the maximum level of welfare for customers.

The risk sharing capacity of the market can be characterized by the slope of the final-position curve. A flatter line represents a greater risk sharing capacity.

The number of customers is the pivotal parameter. M^*, the critical value of M, is the minimum number of customers required in a market such that customer welfare achieved from trading in the order book systems exceeds the welfare from trading in the dealership system. In this setup, the pivotal value of welfare corresponds to M^*=6 for the baseline configuration. That is, in the baseline configuration (defined in Table 10), when the number of customers reaches 6 or more, customer welfare is greater in the order book system than in the dealership system.

It should be emphasized that the particular value of M^* does not have any empirical significance since the parameter values used in the simulations have not been calibrated in relation to empirical benchmarks. However, how the value of M^* changes across various configurations of the trading environment and customer attributes is important and may provide relevant insights.

Columns 4, 5 and 6 of Table 11 report customer welfare levels achieved in the dealership market under various trading environments. Column 4 reports the results from a trading environment in which dealers hold heterogeneous inventories prior to market entry. The critical number of customers varies between M^*=6 and M^*=9, compared to M^*=6 for the baseline trading environment. A greater number of participants is required for the order-book system to dominate the dealership market. The higher welfare values in column 4 relative to the baseline dealership structure in column 3 indicates that customers are better off when dealers are heterogeneous in terms of their initial inventories. Het-

Table 11. Customer Welfare for Various Trading Environments

Number of customers, M	Order-book Market	Dealership Market			
	Baseline	Baseline	$\sigma^2_{X^D} = 0.7$	$\gamma = 0.5$	$N = 12$
4	-1.1482	-1.0585	-1.0325	-1.0311	-1.0557
6	-1.0936	-1.1017	-1.0717	-1.0480	-1.0599
9	-1.0700	-1.1039	-1.0960	-1.0520	-1.0660
12	-1.0561	-1.1107	-1.1100	-1.0533	-1.0660
18	-1.0442	-1.1194	-1.1182	-1.0544	-1.0687
24	-1.0364	-1.1335	-1.1314	-1.0585	-1.0818

erogeneous dealer inventories increase the liquidity risk sharing capacity of the dealership structure. Since dealers at the extremes of the inventory distribution post the most competitive quotes, customers with opposite trading desires are able to trade at better prices than in the baseline trading environment. In the benchmark dealership environment, customers at either end of the liquidity shock spectrum face a homogeneous set of dealers quoting identical prices. In this modified environment, dealers use the first stage of trading as a way to unwind their initial inventory positions by offering better bid-ask spreads to attract customer orders.

Column 5 of Table 11 reports the simulation results for a trading environment in which the dealers are less risk averse than their customers, and the difference in the risk aversion parameter between customers and dealers, γ, is 0.5. In the baseline trading environment, dealers and customers are assumed to be equally risk averse. The critical number of customers in this trading environment, M^*, rises between 12 and 18. A greater number of customers is required for the order book market to prevail under this trading environment than under the benchmark environment. Given that a dealer's bid-ask spread is a function of his risk aversion, customers benefit because less risk averse dealers quote narrower bid-ask spreads in this trading environment. This result is consistent with those found in the market microstructure literature that considers risk averse market participants (e.g. O'Hara 1995; Madhavan 2000).

Column 6 of Table 11 reports customer welfare resulting from a trading environment in which the number of dealers, N, is set to 12. The critical number of customers M^* rises relative to the baseline environment to 12. A greater number of customers is required in this case than in the baseline trading environment for the order book market to dominant a dealership market. Since more dealers participate in second stage inter-dealer trading which improves the risk sharing capacity of the market, these dealers can quote narrower bid-ask spreads to the customers in the first stage. Therefore, customer welfare is higher than that for the baseline case where there are 6 market makers. Similarly, models of dealership markets that ignore private information show that greater competition among dealers or better risk sharing opportunities reduces the bid ask spreads offered to the public with an increase in the number of dealers (e.g. Grossman and Miller 1988; Vogler 1997).

Figure 4 presents the results using a cubic spline interpolation on the data in Table 11. Two thicker lines represent the baseline trading environments of both the order-book and dealership markets. Their crossing point is the value of M^*. Compared to the baseline case, M^* decreases in the risk aversion of dealers, and increases in the variance of the inventories of dealers prior to market entry and the number of dealers who make the market. Thus, low risk aversion, greater heterogeneity in dealer inventories and a large number of dealers make a dealership market relatively more competitive. A larger number of customers is required for an order-book market to prevail.

Concern has recently been expressed by policy makers about the effects that bank mergers may have on market liquidity. As illustrated here, reducing the number of market makers N from 12 to 6 reduces the market making capacity of the dealership market structure and in turn market liquidity. This makes customers worse off as reduced market liquidity is reflected in wider bid-ask spreads. However, as D'Souza and Lai (2002) show, the effect of bank mergers on market liquidity may be ambiguous. If two business lines are correlated, merger increases the total amount of allocation of the risk capital of the new firm. In turn, this may reduce the risk aversion parameters for the new firm and increase the market making capacity.

Combining the trading environments represented in columns 5 and 6, we can investigate what happens to M^* when the number of dealers N declines from 12 to 6 and dealers also become less risk averse. Column 5 represents a trading environment where dealers' risk aversion is 0.5 and customers' is 1, compared to the baseline case where the risk aversion levels of both dealers and customers are equal to 1. M^* in this modified trading environment is best illustrated in Figure 4. The two dealership curves that lie near the top of the chart depict the results as M is varied. The top dealership curve represents the trading environment with 6 dealers and γ =0.5, and the second curves plots the results for the trading environment with 12 dealers and γ=0. As the number of dealers declines and dealer risk aversion declines, M^* rises from 10 to about 13. This indicates that the dealership system is more likely to prevail even as the number of dealers decline. In this setup, the decline in dealer risk aversion more than offsets the negative impact on the risk sharing capacity of the dealership market arising from a decline in the number of dealers. As such, our frameworks predicts that a reduction in the number of market makers does not unambiguously imply a reduction in the market quality of dealership markets.

Table 12 reports the welfare values from simulations where customers are subject to liquidity shocks with larger variance compared to the baseline case. As a result, customers have greater chance to receive large liquidity shocks, which affect the welfare values for both market structures. First, the critical number of customers M^* rises to 9 from 6 in the baseline setting (Panel B). An increased number of customers M is required for an order book market to prevail when customers are subject to large liquidity shocks. Figure 5 illustrates the results from Table 12. Customer welfare decreases in both the order-book and dealership market structures relative to the baseline environment when customers are subject to large liquidity shocks, but the dealership welfare values are affected less than the order-book system values when M is relatively small.

These results may be explained as follows. In the dealership market structure, dealers are modelled as posting bid-ask prices but are assumed to accept orders of any size at these prices. The dealer's mandate of accepting any order size, C_j, at their posted bid-ask prices, is more beneficial for customers who supply large orders to the market than for customers who are relatively homogeneous and are subject to small liquidity shocks. Moreover, customers with large orders benefit from second stage interdealer trading in the dealership market structure. Dealers compete for large orders that they can then lay off with other dealers in the interdealer setting. Thus, customers subject to large, high-variance shocks benefit from the extra risk sharing capacity offered by the dealership market structure.

Table 12. Altering Customer Attributes: Impact of Large and Correlated Shocks

Panel A: Baseline Environments		
Number of Customers	Order-book Market	Dealership Market
M=4	-1.1482	-1.0585
M=6	-1.0936	-1.1017
M=9	-1.0700	-1.1039
M=12	-1.0561	-1.1107
M=18	-1.0442	-1.1194
Panel B: High Variance of Customer Liquidity Shocks		
Number of Customers	Order-book Market	Dealership Market
M=4	-1.2549	-1.1365
M=6	-1.1783	-1.1428
M=9	-1.1368	-1.1662
M=12	-1.1146	-1.1816
M=18	-1.0985	-1.2041
Panel C: Correlated Customer Liquidity Shocks		
Number of Customers	Order-book Market	Dealership Market
M=4	-1.2787	-1.1764
M=6	-1.2448	-1.2007
M=9	-1.2229	-1.2169
M=12	-1.2224	-1.2484
M=18	-1.2107	-1.2698

When customers are characterized by correlated liquidity shocks (Panel C), we find that the dealership market again is more likely to prevail than is the case for the baseline setting. Relative to the baseline environment, the critical number of customers M^* rises to 12. From Table 12, it is clear that environments characterized by correlated order flow challenges the inventory risk management capabilities of both market structures, as average customer welfare declines relative to the baseline cases. However, again the dealership market structure handles this better than the order book market structure does and so implies a greater M^* value than in the baseline case. In summary, when customers are subject to large, correlated liquidity shocks, dealership markets tend to dominate order-book systems.

4.4.3 Measures of Execution Quality

Given the multidimensional nature of execution quality, it is important for intermarket comparisons to be based on either a broad set of execution quality measures or some measure directly related to customer welfare. Although our approach provides results on customer welfare across market structures, we can also calculate several common measures of execution quality. Thus, for the experiments in which customer attributes are varied, we calculate the bid-ask spreads observed in the order book system and the ones observed in the customer-dealer trading round of dealership system. We also calculate the probability of execution observed in the order book system. Since we do not consider the case of transaction within the bid and ask spread, the dealership system ensures execution of all public orders under all trading environments. Thus the probability of execution is a measure that is pertinent only for limit-order book systems. Customer welfare is negatively related to bid-ask spreads and is positively related to the probability of execution.

Table 13 presents the bid-ask spreads calculated in three different cases: the baseline environment; the case where customers are subject to large liquidity shocks; and the case where customers are subject to correlated liquidity shocks. The bid-ask spreads reported in each cell of Panel A of Table 13 represent the average bid-ask spreads submitted to the trading system by the set of M customers. Each cell in Panel B reports the average bid-ask spread posted by dealers in the public segment of the dealership system.

In the order book market (Panel A) of Table 13, bid-ask spreads for a given number of M are larger than those observed in the baseline case when customers are subject to higher variance liquidity shocks. For the dealership market (Panel B), the rise in bid-ask spreads is less substantial in general. However, when customers are subject to correlated liquidity shocks, the increase in bid-ask spreads relative to the baseline case is greater in the order-book system and is substantially greater in the dealership market. As liquidity shocks increase in size or become correlated, the orders submitted by customers increasingly tax the abilities of both the order book and dealership markets to provide rebalancing services. Thus bid-ask spreads increase to compensate market participants that are more likely to be left with undesired portfolio or inventory positions.

Table 14 reports the probability of execution of an order book market. Each number is calculated as the average percentage of the bid and ask order execution in each period per person. Formally, it is defined as $(1/M)(1/T)\sum_{i=1}^{M}\sum_{t=1}^{T}(I_{(B_{it}>P_t)} + I_{(A_{it}<P_t)})$. When customer liquidity shocks are correlated, the simulation results reported in Table 14 indicate that the average probability of limit order execution across customers decreases from those observed in the baseline case. The decline in the probability of execution reflects the

Table 13. Bid-ask Spreads in Order-Book and Dealership Markets

Panel A: Bid-ask Spreads in Order Book Market			
Number of customers	Baseline: variance of the shock is 0.7	Variance of the shock is 1.0	Correlation of the shock is 0.3
M=4	0.4977	0.5902	0.4962
M=6	0.4193	0.4873	0.4316
M=9	0.3789	0.4079	0.4282
M=12	0.2911	0.3550	0.4267
M=18	0.2802	0.3347	0.4204
M=24	0.2490	0.2976	0.4194
Panel B: Bid-ask Spreads in Dealership Market			
Number of customers	Baseline: variance of the shock is 0.7	Variance of the shock is 1.0	Correlation of the shock is 0.3
M=6	0.3298	0.3718	0.8870
M=9	0.3388	0.4382	1.0035
M=12	0.3612	0.4533	1.0357
M=18	0.4213	0.4832	1.0766
M=24	0.4638	0.5012	1.0978

smaller number of offsetting customers orders being submitted to the market when orders are correlated. However, the presence of customers in the order book market who are subject to larger liquidity shocks increases the probability of execution. This occurs because customers are more heterogeneous in this case and thus are more likely to be matched with customers submitting offsetting orders. Given that the probability of execution in the dealership market structure is always one, any increase in the order-execution probability observed in the order book market tends to increase the likelihood of this market prevailing, other things being equal.

The effects on the execution quality of the order-book market seem ambiguous when one combines these higher execution probability figures with the wider bid-ask spread measures for large customers, as reported in Panel A of Table 13. However, as shown in Table 12, customer welfare is lower compared to that in the baseline environment when the liquidity shock variance increases. The effect on customer welfare from wider bid-ask spreads tends to outweight the beneficial effect of the higher probability of order execution. This illustrates the danger of gauging market quality solely on a single measure. The reliance on one measure of execution quality, probability of execution in this case, would have led to the wrong conclusion about the impact on customer welfare.

4.5 Conclusion

In this chapter, we examine the factors that affect a liquidity motivated investor's preference of market structure. We compare the welfare associated with customers in a dealership market structure and a limit-order book structure. We find that the customer's

Table 14. Probability of Execution on Order Book Market

Number of customers	Baseline: variance of the shock is 0.7	Variance of the shock is 1.0	Correlation of the shock is 0.3
M=4	0.5328	0.5563	0.5011
M=6	0.5754	0.6503	0.5712
M=9	0.6254	0.6683	0.5734
M=12	0.7317	0.7239	0.5898
M=18	0.7502	0.7511	0.5953
M=24	0.7697	0.7785	0.6023

choice of trading venue varies depending on the trading environment in which they trade in and the customer's own characteristics. This study builds on the theoretical market microstructure literature in that it develops a theoretical model for both the dealership and order book markets. However, to capture the variety in trading environments and customers characteristics observed in actual financial markets, we use an agent-based computational approach that allows us to examine the equilibrium properties of models that are richer than those in the extant literature in terms of the heterogeneity of trading and customer attributes.

The main predictions of our model can be summarized as follows. First, an increase in the thickness of the market, measured by the number of customers submitting orders at any point in time, causes a rise in probability of an order book system prevailing. Second, customers who submit large-size order flow and customers whose order flow tends to be correlated, are more likely to trade on dealership trading systems than those customers who are smaller and whose order flow is hardly correlate. Third, decreases in the number of dealers increase the likelihood that the order-book market will prevail as the preferred trading venue of risk averse customers. Fourth, an increase in the degree of risk averse dealer heterogeneity or lower dealer risk aversion levels (relative to customer risk aversion) increases the likelihood that customers will choose to trade on the dealership system.

The findings of the paper can be understood in terms of a market structure's capacity to supply liquidity-risk-sharing services under different trading environments, combined with the idea that the varying demand for these services depends on various customer attributes. As trading environments are changed or as customer attributes are varied, a particular market structure will be more robust in terms of its capacity to provide risk sharing services to market participant and in turn become the preferred trading venue. Overall, our study shows that policymakers should be leery of adopting regulations that explicitly or implicitly force competing trade venues (i.e., competing trading systems)to adopt identical or similar trading architectures. This type of "one shoe (market structure) fits all " policy would in the end be detrimental to public investors whom regulators seek to protect.

The results also help explain several of the empirical regularities observed in financial markets. First, these results explain the existence of multiple trade venues each with different trading structures in various financial market segments. Given the heterogeneity of customer characteristics, different market structures appeal to different customers. To the extent that some types of customers lack access to their preferred market trading system, the entry of new trade venues that offer trading structures that appeal to this underserved customer explain their success in competing with incumbent trade venues. Second, our model sheds some light on why the public trading segments of the government bond and FX markets are predominately organized as dealership markets. One explanation could be that these markets are driven mostly by large institutional investors with large-size and likely correlated order flow. Under this trading environment, dealership structure is a better choice for these customers.

There are several questions that lie outside of the scope of this study which we leave for future research. First, asymmetric information does not play a role in the dealer's or the customer's optimizing behavior. Therefore, the spread that a dealer quotes is only a function of inventory risk. To be more realistic, these spreads should be a function of both inventory risk and the asymmetric information once we allow customer order flows to carry pay-off relevant information. Given that our traders do not act on private information, we cannot examine how the public disclosure of trade information may impact customer welfare and execution quality in various trading environments. However, our approach is sufficiently flexible to allow our market participants access to private signals about the liquidation value of the security. This should be pursued in future work.

Second, we do not examine the degree of market efficiency offered by the two market structures examined. A key question raised in much of the agent-based literature is the extent to which the market efficiency of any given market mechanism is attributable to the trading agents' behavior versus the inherent design of the market structure. It would be interesting to measure the degree of market efficiency achieved in the trading mechanisms examined here, and then to assess the extent to which market efficiency is attributable to these various factors.

Finally, Canadian regulators recently considered rules that would, in effect, have allowed customers in the government securities dealership market to submit orders to the interdealer broker system rather than trade with dealers. This initiative would have changed the nature of the decision-making process of both the customers and dealers, since customers would have direct access to the inventory risk-sharing services provided by the interdealer broker system rather than indirect access to these services through dealers. Future research on this issue could examine three market trading mechanisms: the two studied in this chapter, plus a third hybrid dealership structure. In the third one, some customers submit orders to the interdealer broker system (modelled as an limit-order system) while other customers continue to trade with dealers.

4.6 Technical Appendix

In this section, we present the numerical procedure used throughout this chapter to find the optimal strategies for each dealer and the resulting equilibrium. The procedure we develop is usable to solve a wide class of one-period simultaneous-move games. Under certain conditions, one can also use this method to solve multiperiod games, such as in a dealership market and an order book system, which is considered in this chapter.

Before describing the numerical procedure we note the following two caveats. First, even though a wide range of games can be solved using our method, we do not seek to define the class of games that can be solved by our method. Second, if a game admits multiple equilibria, there is no guarantee that the method we propose herein can numerically solve all of them. Moreover, depending on their stability characteristics, certain types of equilibria cannot be attained using our method.

The following section describes the type of game that our method is intended to solve and the method itself. The two subsequent sections describe their application to order book and dealership markets.

4.6.1 Description of the algorithm

Necessary Conditions for Nash Equilibrium

Let Γ be a simultaneous move game described by the following elements. Let $\mathcal{N} = \{1, \ldots, N\}$ be a set of N players. Each player observes a particular random event ω_i drawn from a distribution with support $\Omega_i \subset \mathcal{R}^M$. Each player also has a set of strategy $S_i : \omega_i \longmapsto a_i \in \mathcal{A}_i$, which maps a random event ω_i to an action a_i, where $\mathcal{A}_i$ is a set of actions. The game also defines for each player a payoff function $U_i : \left((a_i)_{i\in\mathcal{N}}, \varepsilon\right) \longmapsto u_i$, a real number depending on the realization of a random variable ε. Therefore, the problem faced by each player is to find a strategy $s_i^* \in S_i$ such that

$$s_i^*\left((s_j)_{j\in\mathcal{N}^{i-}}\right) = \arg\max_{s_i\in S_i} EU_i\left((s_j(\omega_j))_{j\in\mathcal{N}}, \varepsilon\right) \tag{115}$$

where $\mathcal{N}^{i-} = \{1, ...i-1, i+1, ...N\}$. The function s_i^* is the best-response function of player i. In the case where

$$s_i^{**}\left(\left(s_j^{**}\right)_{j\in\mathcal{N}^{i-}}\right) = s_i^{**} \qquad \forall i \in \mathcal{N}, \tag{116}$$

we refer to the set of strategies $(s_i^{**})_{i\in\mathcal{N}}$ as the Nash equilibrium of the game Γ. As (116) indicates, a Nash equilibrium is a fixed point of the best response function of all the players taken simultaneously.

This section describes how this fixed point problem is solved numerically. As explained later, this procedure optimizes numerically the utility function over the possible strategies of each player. In order to guarantee that a numerical optimization is possible we require that the game respects the following two conditions:

1. Each strategy function can be parametrized by a real-valued vector of parameters θ_i. We want the strategy function to have a parametric form since the numerical optimization of each player's utility is substantially simplified if we optimize over a set of parameters rather than a set of functions.

2. The utility function $U_i\left(\left(s_j\left(\omega_j,\theta_j\right)\right)_{j\in\mathcal{N}},\varepsilon\right)$ is continuously differentiable on θ_i. If this condition does not hold, the characterization of the optimal response of a player is much more complicated to define and to obtain numerically.

If these requirements are met, a necessary condition for a set of strategies $(s_i^{**})_{i\in\mathcal{N}}$ to be a Nash equilibrium is

$$\frac{\partial EU_i}{\partial\theta_i}\left(\left(s_j^{**}\left(\omega_j;\theta_j\right)\right)_{j\in\mathcal{N}^{i-}},\varepsilon\right)=0 \qquad \forall i\in\mathcal{N} \tag{117}$$

This condition states that no player has an incentive to deviate, at least locally, from s_i^{**}.

The game in this chapter does not satisfy the function given by the equation (117).

Nevertheless, it is possible to modify the game in such a way that both conditions are met. Doing so allow us to apply our numerical procedure on the modified game. More specifically, each aspect is treated in the following way:

1. With respect to the first condition, suppose that we know very little about the form of s^*. It is still possible to use a flexible function that is parametrized by a vector of parameters θ that can be arbitrarily close to the real optimal strategy function s^*. Numerous examples of such functions exist, such as splines or kernel-based functions.

For the particular games we analyze in this chapter, we use a flexible form of Artificial Neural Network (ANN) that has a form given by

$$\begin{aligned} a_i &= \tilde{s}_i(\omega_i;\theta) \\ &= ANN(\omega_i;\theta) \\ &= \theta_{i,0} + \sum_{k=1}^{H} \theta_{i,k}\, \varphi\left(\theta_{i,(H+1)+k} + \sum_{j=1}^{J} \theta_{i,(2H+1)+j}\, \omega_{i,j}\right), \end{aligned} \tag{118}$$

where a_i is the action performed by player i with given information ω_i and a strategy described by the vector of parameters θ. J is the total number of inputs. The function φ is generally chosen as a twice continuously differentiable function from $\mathcal{R}$ to a bounded interval of $\mathcal{R}$, generally $(0,1)$ or $(-1,-1)$. A few standard choices for the activation function φ are tanh, sigmoid or a bell-shaped Gaussian function.

Often, this form is described in terms of its number of input nodes, which is equal to the dimensionality of ω_i and its number of hidden nodes, given by the parameter H. This terminology refers to the layered organization of the ANN specification. More explicitly, the value of the argument $\omega_{i,j}$ is referred to as the activation level of the input node j, the value of $\varphi(...)$ is called the activation level of a hidden (or middle) node, and a_i is called the activation level of the output node.

2. Since U_i is generally a continuously differentiable function of θ_i, we first define an increasing sequence of real numbers $\{\eta_k\}_{k\in\mathcal{R}}$ s.t. $\lim_{k\to\infty} \eta_k = \infty$, and then look for a sequence of functions $\left(\tilde{U}_i^{\eta_k}\right)_{\eta\in\mathcal{R}}$ that are each continuously differentiable on θ_i and tend to U_i as $k\to\infty$. We refer to the variable η_k as the smoothing parameter. The particular choice of $\tilde{U}_i^{\eta_k}$ we used in this study is explained in the following section,

where we describe more explicitly the application of our procedure to the games analyzed herein. At this point, one should think of $\tilde{U}_i^{\eta_k}$ as given.

Numerical Procedures for Nash Equilibrium

Once a parametric form of the strategy function and a continuously differentiable utility function are defined, it is possible to numerically find an equilibrium of the modified game using the strategy and utility function.

The method we use here is the following: for a given value of η we have a set of functions $\left(\tilde{U}_i^{\eta_0}\right)_{i\in\mathcal{N}}$ that define the game $\tilde{\Gamma}^{\eta_0}$ which has an equilibrium $\left(\tilde{s}_i^{**\eta_0}(\omega_i;\theta_i)\right)_{i\in\mathcal{N}}$ characterized by (117). To get the equilibrium for Γ, we first set a relatively low value η_0 for the smoothing parameter and search the equilibrium strategies $\left(\tilde{s}_i^{**\eta_0}(\omega_i;\theta_i)\right)_{i\in\mathcal{N}}$. Then we increase the smoothing parameter to η_1, and track the movement of the equilibrium. This procedure is continued using a sequence $(\eta_2,\eta_3,...)$ that tends to infinity. This procedure yields a sequence of equilibria $\left(\left(\tilde{s}_i^{**\eta_k}(\omega_i;\theta_i)\right)_{i\in\mathcal{N}}\right)_{k=1,2,...K}$ of the sequence of games $\left(\tilde{\Gamma}^{\eta_k}\right)_{k=1,2,...K}$ that tends to the equilibrium of Γ.

More explicitly, for a given η_k, it is possible to approximate numerically the equilibrium of the game $\tilde{\Gamma}^{\eta_k}$ by using the fixed point characterization given in (116) or (117). The procedure is as follows:

1. Draw a large number, T, of realizations of the random variables $(\omega_i)_{i\in\mathcal{N}}$ where ε should be drawn from the appropriate distribution. These variables are noted as $(\omega_i^t)_{i\in\mathcal{N}}$ and ε^t, $t=1,....,T$

2. For $i = 1$, and given $(\theta_j)_{j\in\mathcal{N}}$, we numerically search for

$$\arg\max_{\theta_i} \frac{1}{T}\sum_{t=1}^{T} \widetilde{U}_i^{\eta_k}\left(\left(s_j\left(\omega_j^t;\theta_j\right)\right)_{j\in\mathcal{N}}, \varepsilon^t\right) \tag{119}$$

and assign that value to θ_i. This procedure corresponds to finding the optimal response of player i to the strategies $\left(s_j\left(\omega_j^t;\theta_j\right)\right)_{j\in\mathcal{N}}$. This is equivalent to solving the equations given in (117) for θ.

3. The variable i is increased incrementally to $i+1$, and step 2 is repeated until convergence of all the $\left(s_j\left(\omega_j^t;\theta_j\right)\right)_{j\in\mathcal{N}}$.[17]

After applying this procedure an equilibrium is obtained for the game $\tilde{\Gamma}^{\eta_k}$. The smoothing parameter is then increased from η_k to η_{k+1}, and a search for an equilibrium of the game $\tilde{\Gamma}^{\eta_{k+1}}$ using the values obtained at the equilibrium of $\tilde{\Gamma}^{\eta_k}$ as starting values for θ. Repeating this procedure until η is large enough, we obtain the equilibrium of a game that is an arbitrarily close approximation of the game Γ.

4.6.2 Application A: Order Book Market – A One Stage Game

In an order book market, we have a set $\mathcal{M}$ of M customers. Each of them receives a liquidity shock X_i and participates in a double auction by posting a bid quote $\left(B_i, Q_i^B\right)$. If the bid B_i is greater than the price P, the customer agrees to buy Q_i^B units of the risky asset at the price P. Also, the customer posts an ask quote $\left(A_i, Q_i^A\right)$. This indicates that if the price is above his ask A_i, he agrees to sell Q_i^A units at the price P. This price in a double auction (DA) is given by Eq.(103).

[17] The formal demonstration of the convergence of the smoothing method is beyond the scope of this study, and we leave that for our future research.

After the auction, the value of the risky asset F is revealed. The random variables F and $(X_i)_{i\in\mathcal{M}}$are independent normal variables with mean 0 and variance σ_F^2 and σ_X^2, respectively.

The information used by customer i is $\omega_i = (X_i)$. Therefore, the strategy function s_i is a mapping from X_i to $\left(B_i, Q_i^B, A_i, Q_i^A\right)$, where $B_i \leq A_i$and $Q_i^B, Q_i^A \geq 0$. We can define the wealth of player i by

$$W_i = X_i F + Q_i (F - P) \tag{120}$$

where Q_i is the quantity executed of player i's order and payoff to player i is in turn given by

$$U_i = -\exp(-\lambda_i W_i) \tag{121}$$

where $\lambda_i$is the risk aversion of customer i.

Parameterization of the strategy function

As mentioned in Section 1 of the Appendix, there are two necessary conditions for finding the Nash Equilibrium. The first deals with the parameterization of the strategy functions. In this application we use the following specification for the strategy functions of market participants:

$$\begin{aligned} \frac{B_i + A_i}{2} &= ANN\left(X_i, \theta_i^{mid}\right) \\ A_i - B_i &= \phi\left(\theta_i^{spread}\right) \\ Q_i^B &= \phi\left(ANN\left(X_i, \theta_i^{Q^B}\right)\right) \\ Q_i^A &= \phi\left(ANN\left(X_i, \theta_i^{Q^A}\right)\right) \end{aligned} \tag{122}$$

Therefore, θ_i is $\left(\theta_i^{mid}, \theta_i^{spread}, \theta_i^{Q^B}, \theta_i^{Q^A}\right)'$. ANN is defined by Eq. (118). The function ϕ is given by

$$\phi(x) = \log(1 + \exp(x))$$

and is used to give non-negative actions.

Modification of the Utility Function

It can easily be noted that the function U_i does not satisfy the smoothness conditions mentioned earlier. To make the utility function differentiable, a smoothing technique is required to modify the utility function. In this application we show one example for the smoothing method that can be used in this game.

Given a smoothing parameter η, we rewrite the definition of P, now noted $\tilde{P}^\eta$, as the root of:

$$\sum_{i=1}^{M} Q_i^B \tilde{I}^\eta_{\left(B_i - \tilde{P}^\eta\right)} - \sum_{i=1}^{M} Q_i^A \tilde{I}^\eta_{\left(\tilde{P}^\eta - A_i\right)} = 0 \tag{123}$$

where $\tilde{I}_\eta$ is defined as

$$\tilde{I}^\eta_{(x)} = \begin{cases} 0 & if \quad x < -\dfrac{1}{\eta} \\ \dfrac{(x\eta)^2}{2} + x\,\eta + \dfrac{1}{2} & if \quad -\dfrac{1}{\eta} < x < 0 \\ -\dfrac{(x\,\eta)^2}{2} + x\,\eta + \dfrac{1}{2} & if \quad 0 < x < \dfrac{1}{\eta} \\ 1 & if \quad x > \dfrac{1}{\eta} \end{cases} \tag{124}$$

Hence, $\tilde{I}_\eta$ is a second degree polynomial defined by components having continuous first derivatives everywhere. In the limit, (as $\eta \to \infty$), it is equivalent to $I\ (x > 0)$. It offers the advantage that the space of polynomials so defined by components is closed under addition. This implies that (123) also is a second degree polynomial defined by parts having continuous first derivatives. Therefore, its root easily are calculated.

The use of the function $\tilde{I}_\eta$ is illustrated in the excess demand function given in Figure 6. As the smoothing parameter η increases from 1 to 200, this smoothed excess demand function becomes steeper and approaches the step function.

Since the polynomial indicator function is closed under addition, the excess demand function (123) also is a second degree polynomial defined by components. Therefore, its root, $\tilde{P}_\eta$, can be calculated. Now given $\tilde{P}_\eta$, the executed quantity for the player i is

$$\tilde{Q}_i^\eta = Q_i^B \, \tilde{I}^\eta_{\left(B_i - \tilde{P}^\eta\right)} - Q_i^A \, \tilde{I}^\eta_{\left(\tilde{P}^\eta - A_i\right)}, \tag{125}$$

The wealth of player i in equation (120) becomes

$$\tilde{W}_i^\eta = X_i \, F + \tilde{Q}_i^\eta \left(F - \tilde{P}^\eta\right). \tag{126}$$

Accordingly, the utility function in equation (1) becomes

$$\tilde{U}_i^\eta = -\exp\left(-\lambda_i \, \tilde{W}_i^\eta\right). \tag{127}$$

The objective function of each player is approximated by simulation. This method begins by drawing T realizations of the vector $(X_i)_{i\in\mathcal{M}}$. Given (X_{it}, θ_i), each player submits a bid and ask quote $\left(B_{it}, Q_{it}^B, A_{it}, Q_{it}^A\right)$, from which we can find $\left(\tilde{P}_t^\eta, \tilde{Q}_{it}^\eta\right)$. Conditional on $(X_{jt})_{j\in\mathcal{M}}$, the wealth of player i has a normal distribution with mean $-\tilde{Q}_{it}^\eta \, \tilde{P}_t^\eta$ and variance $\sigma_F^2 \left(X_{it} + \tilde{Q}_{it}^\eta\right)^2$. Therefore, the utility of player i conditional on $(X_{it})_{i\in\mathcal{M}}$ is written as

$$E\left(\tilde{U}_i^\eta \mid (X_{jt})_{j\in\mathcal{M}}\right) = -\exp\left(\frac{\lambda_i^2 \, \sigma_F^2}{2} \left(X_{it} + \tilde{Q}_{it}^\eta\right)^2 + \lambda_i \, \tilde{Q}_{it}^\eta \, \tilde{P}_t^\eta\right)$$

The objective function for player i is approximated by

$$E\tilde{U}_i^\eta \cong \frac{1}{T}\sum_{t=1}^{T} E\left(\tilde{U}_{it}^\eta \mid (X_{jt})_{j\in\mathcal{M}}\right) \tag{128}$$

For our experiments, we consider $\sigma_F^2 = 0.7$, $\sigma_X^2 = 0.7$ and $\lambda_i = 1$. These values guarantee that the mean and the variance of the unconditional expected utility are still finite in the case where no trade happens in the double auction. Also, we set $T = 1000$ for the number of random draws. The smoothing parameter η starts with an initial value of 1 and is multiplied by 1.2 on each iteration, until it reaches a value of approximately 1000. The ANN structure is one input-one hidden node and one output. The activation function φ takes the form of $\varphi(x) = \tanh(x)$.

4.6.3 Application B: Dealership Market– A Two Stage Game

Recalling that in a dealership market, we have a set $\mathcal{M}$ of M customers indicated by j, and a set $\mathcal{N}$ of N dealers indicated by i. Each customer receives a liquidity shock X_j. Each dealer i submits a pair of quotes (B_i^1, A_i^1) that indicates that the dealer agrees to buy or sell any number of units of the risky asset at a price B_i^1 or A_i^1, respectively. Each customer j then is allowed to observe a randomly chosen set of k dealers denoted Δ_j and determine the quantity C_{ij} that he wishes to trade and with which dealer. The total amount traded by the customer j is denoted $C_j = \sum_{i\in\mathcal{N}} C_{ij}$. The dealers then are allowed to trade among each other by participating in a double auction, where they submit a bid quote $\left(B_i^2, Q_i^{B2}\right)$ and an ask quote $\left(A_i^2, Q_i^{A2}\right)$. In a similar way as in the order book market, a quantity Q_i is executed at a price P. Finally, a fundamental value F is revealed.

In the first stage of the game, each customer observes a set of quotes $(B_i^1, A_i^1)_{i\in\Delta_j}$. Note that the maximum bid of that subset $\overline{B}_j^1$and the minimum ask $\underline{A}_j^1$, specifically, are:

$$
\begin{aligned}
\overline{B}_j^1 &= \max\{B_i^1 : i \in \Delta_i\} \\
\underline{A}_j^1 &= \min\{A_i^1 : i \in \Delta_i\}
\end{aligned}
$$

Similarly, as in an order book market, the values for $P$and $Q_i$can be smoothed according to equation (123), (124) and (125).

C_{ij} still has to be written as a smooth function of B^1 and A^1. First, we rewrite the best quotes $\overline{B}_j^1$ and $\underline{A}_j^1$ as

$$
\begin{aligned}
\widetilde{\overline{B}}_j^{1\eta} &= \frac{\sum_{i\in\Delta_j} B_i^1 \exp(\eta\, B_i^1)}{\sum_{i\in\Delta_j} \exp(\eta\, B_i^1)} \\
\widetilde{\underline{A}}_j^{1\eta} &= \frac{\sum_{i\in\Delta_j} A_i^1 \exp(-\eta\, A_i^1)}{\sum_{i\in\Delta_j} \exp(-\eta\, A_i^1)}
\end{aligned}
\tag{129}
$$

These functions are weighted averages of the B_i^1 and A_i^1 and converge toward the maximum and minimum elements of $(B_i^1)_{i\in\Delta_j}$ and $(A_i^1)_{i\in\Delta_j}$, respectively.

Consequently, customer's order C_j in function (109) can be rewritten as

$$
\widetilde{C}_j^\eta = \left(\frac{-\widetilde{\overline{B}}_j^{1\eta}}{\lambda_j^C\,\sigma_F^2} - X_j\right)\widetilde{I}^\eta_{\left(\widetilde{\overline{B}}_j^{1\eta} - P_j^R\right)} + \left(\frac{-\widetilde{\underline{A}}_j^{1\eta}}{\lambda_j^C\,\sigma_F^2} - X_j\right)\widetilde{I}^\eta_{\left(P_j^R - \widetilde{\underline{A}}_j^{1\eta}\right)}.
$$

The share of C_j in function (111) that is passed to dealer i is given by

$$
\widetilde{C}_{ij}^\eta = \widetilde{C}_j^\eta\left(\widetilde{I}^\eta_{\left(\widetilde{\overline{B}}_j^{1\eta} - P_j^R\right)}\frac{\exp(\eta\, B_i^1)}{\sum_{i\in\Delta_j}\exp(\eta\, B_i^1)} + \widetilde{I}^\eta_{\left(P_j^R - \widetilde{\underline{A}}_j^{1\eta}\right)}\frac{\exp(-\eta\, A_i^1)}{\sum_{i\in\Delta_j}\exp(-\eta\, A_i^1)}\right)
\tag{130}
$$

Finally, a smoothed version of dealer i's wealth (113) is written as

$$
\widetilde{W}_i^{D\eta} = \sum_{j\in\mathcal{M}}\widetilde{C}_{ij}^\eta\left(\widetilde{I}^\eta_{(\widetilde{C}_{ij}^\eta>0)}A_i^1 + \widetilde{I}^\eta_{(\widetilde{C}_{ij}^\eta<0)}\,B_i^1\right) - \tilde{Q}_i^\eta\,\tilde{P}^\eta + F\left(\tilde{Q}_i^\eta - \sum_{j\in\mathcal{M}}\widetilde{C}_{ij}^\eta\right)
\tag{131}
$$

The objective function of each dealer is approximated by simulation. First, T realizations of the vector $(X_j)_{j\in\mathcal{M}}$ and $(\Delta_j)_{j\in\mathcal{M}}$ are drawn. Conditionally on $(X_j)_{j\in\mathcal{M}}$ and $(\Delta_j)_{j\in\mathcal{M}}$, dealer i's wealth $\widetilde{W}_i^{D\eta}$ in eq. (131) has a normal distribution with mean and variance as:

$$E(\widetilde{W}_{it}^{D\eta}|\,(X_{jt})_{j\in\mathcal{M}}, (\Delta_{jt})_{j\in\mathcal{M}}) = \sum\nolimits_{j\in\mathcal{M}} \widetilde{C}_{ijt}^{\eta} \left(\widetilde{I}^{\eta}_{(\widetilde{C}^{\eta}_{ijt}>0)} A^1_{it} + \widetilde{I}^{\eta}_{(\widetilde{C}^{\eta}_{ijt}<0)} B^1_{it} \right) - \tilde{Q}^{\eta}_{it} \tilde{P}^{\eta}_{t} \tag{132}$$

$$var(\widetilde{W}_{it}^{D\eta}|\,(X_{jt})_{j\in\mathcal{M}}, (\Delta_{jt})_{j\in\mathcal{M}}) = \sigma_F^2 \left(\tilde{Q}^{\eta}_{it} - \sum\nolimits_{j\in\mathcal{M}} \widetilde{C}^{\eta}_{ijt} \right)^2 v \tag{133}$$

Therefore, the utility of dealer i conditional on $(X_{jt})_{j\in\mathcal{M}}$ and $(\Delta_{jt})_{j\in\mathcal{M}}$, is written as

$$\begin{aligned} & E(\widetilde{U}_{it}^{D\eta}|\,(X_{jt})_{j\in\mathcal{M}}, (\Delta_{jt})_{j\in\mathcal{M}}) & (134) \\ = & -\exp[-\lambda^D(E(\widetilde{W}_{it}^{D\eta}|\,(X_{jt})_{j\in\mathcal{M}}, (\Delta_{jt})_{j\in\mathcal{M}})) + \frac{(\lambda^D)^2}{2} var(\widetilde{W}_{it}^{D\eta}|\,(X_{jt})_{j\in\mathcal{M}}, (\Delta_{jt})_{j\in\mathcal{M}})] \end{aligned}$$

Using a method identical to the one used in the order book market, the unconditional expected utility is approximated by

$$E(\widetilde{U}_i^{D\eta}) \cong \frac{1}{T}\sum_{t=1}^{T} E(\widetilde{U}_{it}^{D\eta}|\,(X_{jt})_{j\in\mathcal{M}}, (\Delta_{jt})_{j\in\mathcal{M}}) \tag{135}$$

For the baseline experiments, a setup similar to the one used for the order book system is chosen in order to facilitate the comparison: $\sigma_F^2 = 0.7$, $\sigma_X^2 = 0.7$, $\lambda^D = 1$, $\lambda^C = 1$. We use $T = 1000$ random draws. The smoothing parameter starts with an initial value of 1 and is multiplied by 1.2 on each iteration, until it reaches a value of approximately 1000.

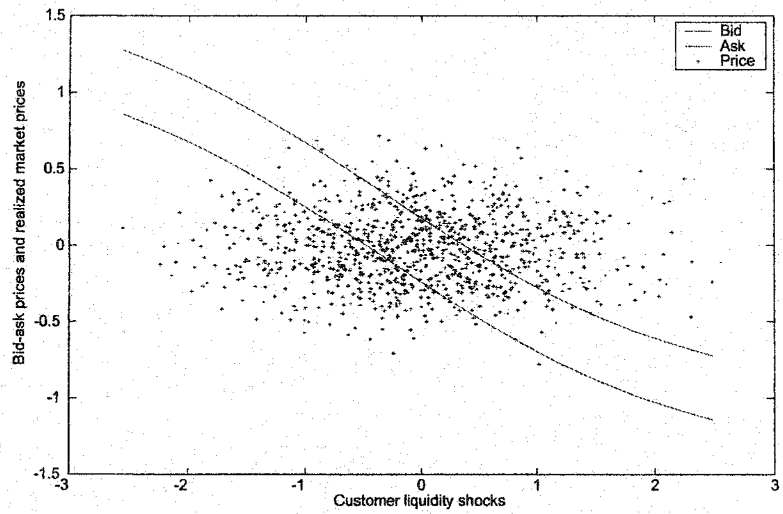

Figure 1: Customer's Best Response Function

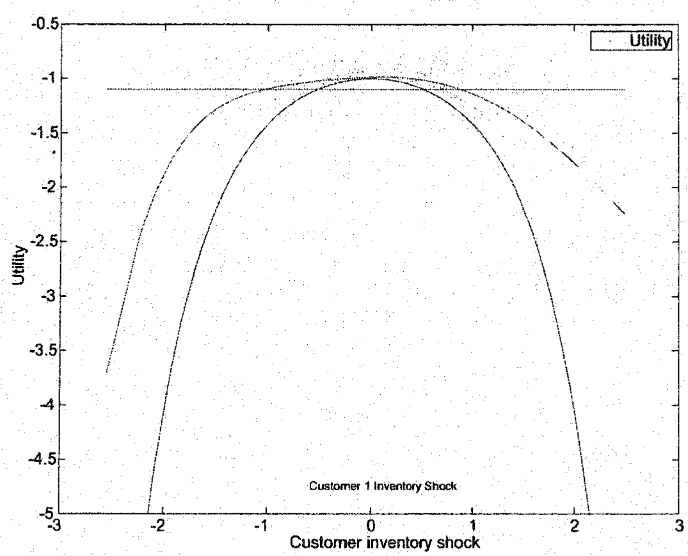

Figure 2: Reservation Utility vs. Realized Utility

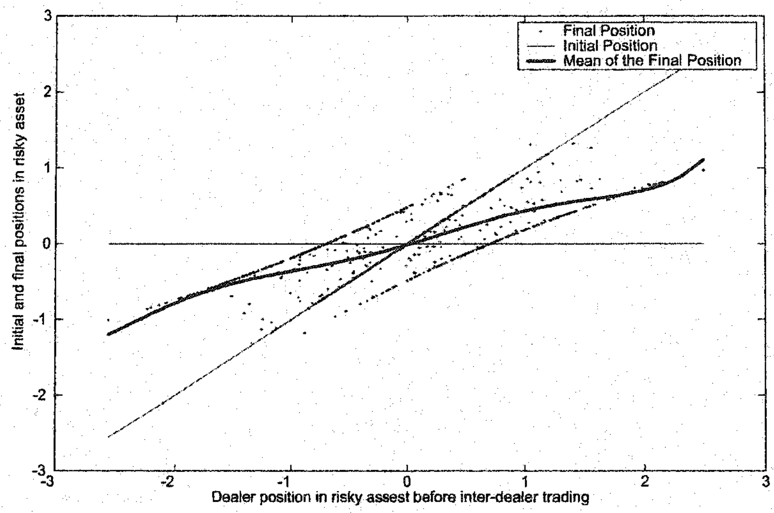

Figure 3. Risk Sharing

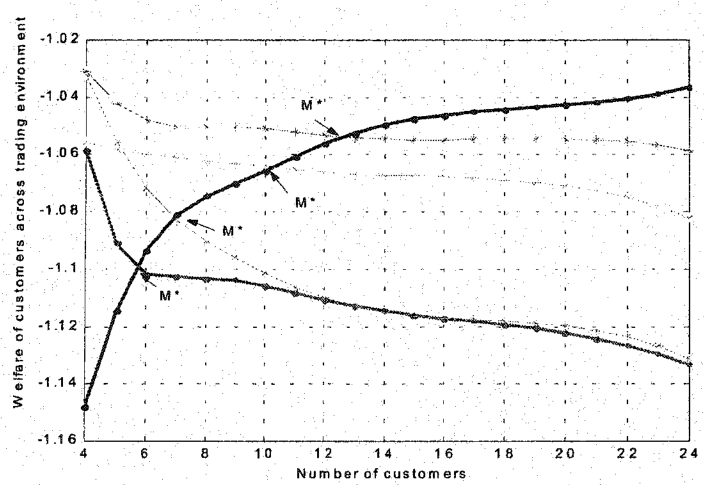

Figure 4: Customer Welfare Across Trading Environment

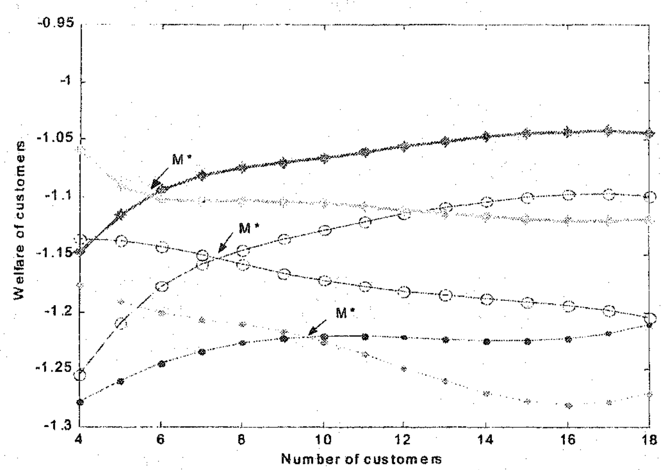

Figure 5: Impact of the Features of Customers Order Flow

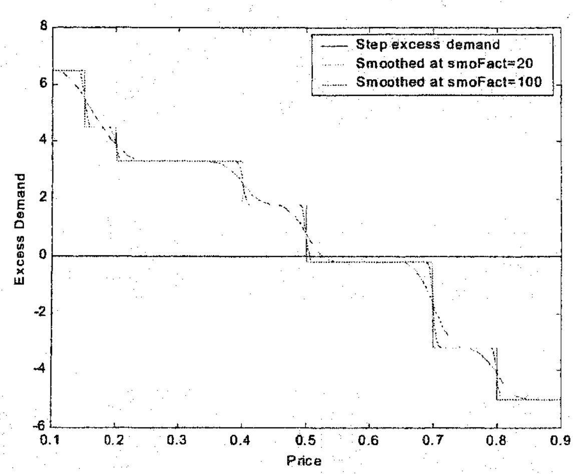

Figure 6: Smoothed Excess Demand in a Double Auction

Chapter 5

Conclusions and Future Research

5.1 Major Findings

This thesis bridges agent-based simulations and the market microstructure literature. The first chapter provides a road map for agent-based research and a selective survey of market microstructure research on institutional design. Despite the many insights generated from agent-based studies, most of this literature fails to address the importance of trading institutional design in financial markets. The three following chapters investigate issues regarding institutional design from various perspectives.

Chapter 2 advances the agent-based literature by introducing a more realistic continuous double-auction trading mechanism in a simulated market. It studies the impact of market mechanisms on the learning behaviors of agents and, in turn, market dynamics compared with a benchmark case involving a Walrasian auction. In most of the agent-based literature, the time is irrelevant. For instance, in a Walrasian auction, trades are aggregated and the market prices are determined by the net trading imbalance. The time when orders are submitted is not relevant as transactions always occur at the equilibrium price. In a continuous double auction market, however, the time of order submission has an impact on the transaction price. In turn, prices become path dependent. We find that market dynamics under double-auction converge to the Rational Expectation Equilibrium in some experiments. However this convergence is sensitive to deviations from rationality. In fact, in the experiments where noise trading is introduced, convergence becomes unattainable. Therefore, minimal rationality is not sufficient to generate convergence in a double-auction market when market prices are endogenous. When the liquidation value is endogenously generated, the convergence to REE is difficult to attain. Because market prices depend on each participant's expectation for the asset. However, whether it is caused by the learning behavior of agents or by the double auction trading system is still a question that remains unanswered.

Chapter 3 builds a theoretical model to investigate the determinants of the dealer's choice of market mechanism in the context of inter-dealer trading. The choice set includes two inter-dealer markets (IDM); namely, an order-driven brokered IDM and a quote-driven direct IDM. We show that the brokered IDM dominates the direct IDM under certain conditions. Under these conditions, the brokered inter-dealer market provides a better risk-sharing opportunities in the sense that the trading dealer can unwind a greater proportion of his inventory at a lower cost.

In the absence of information asymmetry, we show that a sufficient increase in volatility of the asset and dealer risk aversion tends to shifts IDM trading from being direct to being brokered. Moreover, our findings suggest a link between the choice of inter-dealer trading venue with the dynamics of customer order flow. We show that an increase in the volatility of customer order flow tends to shift the IDM trading to be brokered. This finding is particularly interesting for markets, where trading on private information is not significant.

We also demonstrate that the dealer's choice of inter-dealer trading venue has an impact on customer welfare. When dealers have a choice of inter-dealer markets, we show that a customer can negotiate a better or at least not-worse price than when dealers do not have. In turn, this implies that the welfare of customers improve or at least is not worse-off when dealers have a choice of inter-dealer trading venues. This result may explain the widespread use of the hybrid of two inter-dealer trading mechanisms in a variety of dealership markets that are populated by institutional investors.

In Chapter 2, agent-based models with more realistic trading system and strategies are are used to capture the complex learning behaviors and dynamics in financial markets. Under this framework, financial agents do not necessarily interact in an optimal fashion. Chapter 3 builds a theoretical framework to study the equilibrium behaviors of agents in a highly idealized setting. By solving the Nash equilibrium of the game numerically, we ensure that agents interact based on their optimal strategies in Chapter 4.

Chapter 4 studies the customers choice of an order-book market versus a dealership market. Using agent-based simulation, this chapter develops a numerical method to solve for a Nash equilibrium. We find the presence of dealers increases the welfare of public investors in economies with less opportunities for improving risk sharing. More precisely, customers who submit large-size order flow and customers whose order flows tend to be correlated, are more likely to trade on dealership trading systems. A decrease in the thickness of the market, measured by the number of customers submitting orders at any point in time, causes a rise in the probability that order book system prevails. An increase in the heterogeneity of the degree of risk averse dealers or lower dealer risk aversion levels (relative to customer risk aversion) increases the likelihood that customers choose to trade on the dealership system.

The findings of this chapter can be understood in terms of a market structure's capacity to supply liquidity and risk-sharing services under different trading environments and varying demand for these services. As trading environments are changed or as customer attributes are varied, a particular market structure will be more robust in terms of its capacity to provide risk sharing services to market participants. In turn, this trading venue will become the preferred trading venue.

5.2 Future Research

Although agent-based simulations have already demonstrated an ability to understand many aspects of financial markets, more formal modelling is required for the field to progress. Current agent-based simulations have generally involved little formal modelling. It is hard to pin down the relationships among variables without a structured model. One direction for this field is to search for an appropriate tool, such as game theory, to set-up the experiment to model the strategic behavior among agents. We made the very first attempt to use game theory to formalize an agent-based simulation in Chapter 4 where a numerical method was developed to solve for a Nash equilibrium. The generalization of this numerical method, the characteristic of the class of games to which this method can apply, and the convergence properties of the numerical method remain interesting topics for future research.

Theoretical research on microstructure features of fixed income and FX markets are needed. Goodhart and O'Hara (1997) in their survey article state "... a serious omission in the work survey is the lack of research on microstructure issues in fixed income market. Despite the size and the importance of these markets, there are virtually little empirical and theoretical microstructure analysis." (page 105). The model typically employed in the analysis of equity markets emphasizes the role of private information, and it models the price process by informed traders, uninformed traders and a market maker. Analyses of fixed income and FX markets have focused more on public information. In this approach, inventory control plays an important role. But is this the only motivation for trading? What is the role of dealers, and to what extent can this role be replaced by a centralized order-driven electronic trading system? We have attempted to answer some of these questions in Chapter 4 assuming the absence of asymmetric information. However, if asymmetric information exists on these markets, does it exist among dealers, between dealers and customers, or between different types of customers? Where this asymmetric information originates and how dealers use this information to extract rents from customer also are of interest. All these aspects are not well understood yet.

Although empirical research on the microstructure of fixed income markets is needed, the decentralized feature of these markets makes centralized data collection very difficult. The situation is changing with the rapid growth of ATS or ECNs. CanPX in Canada and GovPX in the US, as centralized inter-dealer broker systems, collect much valuable data. Given the rudimentary state of current work, a small-scale study would be of great interest to the field. Empirical work and even stylized facts are needed to establish the difference in microstructure features between equity and fixed income markets.

Policy issues call for more research on market transparency. Market microstructure is naturally well placed for providing policy implications of market regulation. One of the questions is regulation of trading disclosures. For example, Canadian regulators recently proposed a regulation under which dealers in fixed income markets are required to disclose on a real-time basis to the market all of their post-trade prices and volumes. This can be a natural extension for the work reported in Chapter 4. One can introduce trading disclosure after the first round of customer-dealer trades and then investigate the impact that this rule has on bid-ask spreads, risk sharing and the choice of the inter-dealer market for trading. It is interesting to determine whether greater transparency shifts the distribution of return from dealers to small customers, and whether incomplete transparency allows for the optimal level of risk sharing among market participants.

References

Admati, A. and P. Pfleiderer, 1988, "A Theory of Intraday Trading Patterns," *Review of Financial Studies*, 1, 3-40.

Amihud, Y. and H. Mendelson, 1986, "Asset Pricing and the Bid-Ask Spread," *Journal of Financial Economics*, 17: 223-249.

Andersen, T.G., 1996, "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance*, 51, 1, 169-204.

Arifovic, J., 1996, "The Behavior of the Exchange Rate in the Genetic Algorithm and Experimental Economics," *Journal of Political Economy*, 104, 510-541.

Arthur, B., J. Holland, B. LeBaron, R. Palmer and P. Tayler, 1997, "Asset Pricing Under Endogenous Expectations in an Artificial Stock Market," in Arthur, W.B., S. Durlauf, and D. Lane, (eds.), *The Economy as an Evolving Complex System II*, Addison-Wesley, Reading, MA, 15-40.

Bank for International Settlements, BIS, 2001, "The Implications of Electronic Trading in Financial Markets," CGFS Publication No. 16, January.

Bloomfield, R. and M. O'Hara, 1999, "Market Transparency: Who Wins and Who Loses?", *The Review of Financial Studies*, Spring 1999, Vol. 12, 5-35.

Bloomfield, R and M. O'Hara, 2000, "Can Transparent Markets Survive?", *Journal of Financial Economics*, 55: 425-459.

Belttratti, A., S. Margarita and P. Terna, 1996, "Neural Networks for Economic and Financial Modeling," International Thomson Computer Press, London, UK.

Bray, M., 1982, "Learning, Estimation, and the Stability of Rational Expectations," *Journal of Economic Theory*, 26, 318-339.

Brock, W., J. Lakonishok and B. LeBaron, 1992, "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns," *Journal of Finance* 47,1731-1764.

Brown, S., W. Goetzmann and A. Kumar, 1998, "The Dow Theory: William Peter Hamilton's Track Record Reconsidered," *Journal of Finance*.

Cao, H., and Lyons, R., 1998, "Inventory Information," Working Paper, Haas School of Business, U.C. Berkeley, August.

Camerer, C., and Weigelt, K., 1992, "Convergence in Experimental Double Auctions for Stochastically Lived Assets," in Friedman, D. and Rust, J., (eds.), *The Double Auction Market: Institutions, Theories, and Evidence*, Addison-Wesley, Reading, MA, 355-390.

Campbell, J., A. Lo and A. MacKinlay,1996, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

Chan, LeBaron, Lo and Poggio, 2001, "Information Dissemination and Aggregation in Asset Markets with Simple Intelligent Traders," Working Paper, MIT.

Chen, S. H., and C.H. Yeh, 2002, "Evolving Traders and the Faculty of the Business School: A New Architecture of the Artificial Stock Market," in Chen S. H. (ed.), *Evolutionary Computation in Economics and Finance*, Springer-Verlag.

Cochrane, J., 2001, *Asset Pricing*, Princeton Press, US.

Dattels, P., 1995, "The Microstructure of Government Securities Markets," IMF Working Paper 95/117.

DeLong, J., A. Schleifer, L. Summers and R. Waldman, 1995, "The Survival of Noise Traders in Financial Markets," *Journal of Business*, 64, 1-20.

DeLong, J., A. Shleifer and L. Summers, 1991, "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, 98, 703-738.

Donaldson, G. and M. Kamstra, 1996, "A New Dividend Forecasting Procedure that Rejects Bubbles in Asset Prices: The Case of 1929 Stock Crash," Review of Financial Studies 9, Summer, 333-83.

D'Souza C. and A. Lai, 2002, "The Effects of Bank Consolidation on Risk Capital Allocation and Market Liquidity," Bank of Canada Working Paper.

Dutta, P. and A. Madhavan, 1997, "Competition and Collusion in Dealer Markets," *Journal of Fiance*, Vol. 52(1), 245-276.

Evans, M. and R. Lyons, 2002, "Order Flow and Excahnge Rate Dynamics," *Journal of Political Economy*, February, 170-180.

Financial Services Authority, 2000, "The FSA's approach to Regulation of Market Infrastructure," FSA Discussion Paper, January, 2000.

Fleming, M., and E. Remolona, 1999, "Price Formation and Liquidity in the U.S. Treasury Market: The Response to Public Information," *Journal of Finance*, 54, 5, 1901-1916.

Flood, M., R. Huisman, K. Keodijk, R. Mahieu, and A. Röell, 1999, "Quote Disclosure and Price Discovery in Multiple-dealer Financial Markets," *Review of Financial Studies*, Volume 12, 1, 37-59.

Flood, M., 1994, "Market Structure and Inefficiency in the Foreign Exchange Market," *Journal of International Money and Finance*, 1994, 13, 2., 131-158.

Forsythe, R., T.R. Palfrey and C.R. Plott, 1982, "Asset Valuation in an Experimental Market," *Econometrica*, 50, 537-567.

Friedman, D and J. Rustin, 1992, Preface, in Friedman, D. and Rust, J., (eds.), *"The Double Auction Market: Institutions, Theories, and Evidence*," Addison-Wesley, Reading, MA, 155-198.

Ganley, J., A. Holland, V. Saporta and A. Vila, 1998, "Transparency and the Design of Securities Markets," Bank of England Financial Stability Review, No. 4, Spring.

Garcia, R. and R. Gencay, 1999, "Pricing and Hedging Derivative Securities with Neural Networks and Homogeneity Hint," *Journal of Econometrics*.

Gemmill, G., 1996, "Transparency and Liquidity: A Study of Block Trades on the London Stock Exchange under Different Publication Rules," *Journal of Finance*, 51, 1765-1790.

Glosten, L., and P. Milgrom, 1985, "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Agents," *Journal of Financial Economics*, 14, 71-100.

Glosten, L., 1994, "Is the Electronic Open Limit Order Book Inevitable?" *Journal of Finance*, Vol. XLIX, No. 4, September 1994.

Gode, D. and S. Sunder, 1993, "Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality," *Journal of Political Economy*, Vol.101.No.1.

Gravelle, T. 1999, "Liquidity of the Government of Canada Securities Market: Stylised Facts and Some Market Microstructure Comparisons to the United States Treasury Market," Bank of Canada Working Paper No. 99-11.

Gravelle, T., 2002, "The Market Microstructure of Equity and Government Securities Markets: How They Differ," Bank of Canada Working Paper, 2002-9.

Grossman, S. and Miller, M., 1988, "Liquidity and Market Structure," *Journal of Finance*, Vol. 43(3), 617-637.

Grossman, S. and T. Stiglitz, 1980, "On the Impossibility of Informationally Efficient Markets," *American Economic Review,* 70, 393-408.

Hansch, O., N. Naik, and S. Viswanathan, 1998, "Do Inventories Matter in Dealership Markets? Evidence from the London Stock Exchange," *Journal of Finance*, Vol. LIII, No. 5, 1998.

Harris, L., 1993, "Consolidation, Fragmentation, Segmentation and Regulation," *Financial Institutions and Instruments,* Volume 2, 5, 1-28.

Hertz, J., A. Krogh, and R. Palmer, 1991, *An Introduction to the Theory of Neural Computation,* Addison-Wesley Publishing Company, Reading, Mass.

Hirshleifer, J., 1971, "The Private and Social Value of Information and Reward to Incentive Activity," *American Economic Review*, 61, 561-574.

Ho, T. and H. Stoll, 1983, "The Dynamics of Dealer Markets under Competition," *Journal of Finance,* 38, 1053-1074.

Hussman, 1992, "Market Efficiency and Inefficiency in Rational Expectations Equilibria," *Journal of Economic Dynamics and Control,* 16.

Hutchinson, J., A. Lo and T. Poggio, 1994, "A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks," *Journal of Finance*, 49, 3, 851-889.

Kyle, A., 1985, "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315-1335.

Kryzanowski, L. and H. Nemiroff, 2001, "Market Quote and Spread Component Cost Behavior Around Trading Halts for Stocks Interlisted on the Montreal and Toronto Stock Exchanges?", *The Financial Review*, Vol. 36, No. 2, 2001, 115-138.

Lamoureux, C and C. Schnitzlein, 1997, "When It's Not the Only Game in Town: the Effect of Bilateral Search on the Quality of a Dealer Market," *Journal of Finance*, Vol. LII, No. 2, 1997.

Leach, J and A. Madhavan, 1993, "Price Experimentation and Security Market Structure," *Review of Financial Studies*, 6, 375-404.

LeBaron, B., B. Arthur and R. Palmer, 1999, "Time Series Properties of an Artificial Stock Market," *Journal of Economic Dynamics and Control*, 23, 1999, 1487-1516.

LeBaron, B., 2000, "Agent Based Computational Finance: Suggested Reading and Early Research," *Journal of Economic Dynamics and Control*, Vol. 24, June 2000.

LeBaron, B., 2001, "A Builders Guide to Agent-based Financial Markets," *Quantative Finance*, Vol. 1, 254-261.

LeBaron, B., 2002, "Building Financial Markets with Artificial Agents: Desired Goals, and Present Techniques," in Karakoulas, G. (ed.), *Computational Markets*, MIT Press, forthcoming.

Lee, C. 1993, "Market Integration and Price Execution for NYSE-listed Securities," *Journal of Finance,* 48, 1009-1038.

Lyons, R., 2001, *"The Microstructure Approach to Exchange Rates,"* MIT Press.

Lyons, R., 1997, "A Simultaneous Trade Model of the Foreign Exchange Hot Potato," *Journal of International Economics,* Vol. 42 ,1997, 275-298.

Lyons, R., 1996, "Optimal Transparency in a Dealer Market with an Application to Foreign Exchange," *Journal of Financial Intermediation,* 5, 225-254.

Lyons, R., 1995, "Tests of Microstructural Hypotheses in the Foreign Exchange Market," *Joural of Financial Economics,* Vol. 39, 1995, 321-351.

Madhavan, A. and S. Smidt, 1991, "A Bayesian Model of Intraday Specialist Pricing," *Journal of Financial Economics,* 30, 99-134.

Madhavan, A. and S. Smidt, 1993, "An Analysis of Changes in Specialist Inventories and Quotations," *Journal of Finance,* 48, 1595-1628.

Madhavan, A., 1992, "Trading Mechanisms in Securities Markets," *Journal of Finance*, Vol. XLVII, No. 2, June 1992.

Madhavan, A., 1995, "Consolidation, Fragmentation, and the Disclosure of Trading Information" *Review of Financial Studies*, 8, 579-603.

Madhavan, A., 2000, "Market Microstructure: A survey," *Journal of Financial Markets,* Vol. 3, 205-258.

Madhavan, A., D. Porter and D. Weaver, 2001, "Should Securities Markets be More Transparent?", Bank of Canada 2001 Conference Paper.

Naik, N., A. Neuberger and S. Viswanathan, 1999, "Trade Disclosure regulation in markets with negotiated trades," *Review of Financial Studies*, Volume 12, 4, 873-900.

O'Hara, M., 1995, *Market Microstructure Theory,* Blackwell Publishers, Cambridge USA.

O'Hara, M. and G. Oldfield, 1986, "The Microeconomics of Market Making," *Journal of Financial and Quantitative Analysis*, 21, 361-376.

Pagano, M. and A. Röell, 1996, "Transparency and Liquidity: A Comparison of Auction and Dealer Markets with Informed Trading," *Journal of Finance*, 51, 579-611.

Plott, C. R. and S. Sunder, 1982, "Efficiency of Experimental Security Markets with Insider Information: An Application of Rational Expectations Models," *Journal of Political Economy,* 90, 663-698.

Plott, C. R. and S. Sunder, 1988, "Rational Expectations and the Aggregation of Diverse Information in Laboratory Settings," *Econometrica*, 56, 1085-1118.

Porter and Weaver, 1998, "Post -Trade Transparency on Nasdaq's National Market System," *Journal of Financial Economics,* 50, 231-252.

Proudman, J., 1995, "The Microstructure of the U.K. Gilt Market," Bank of England Working Paper 38.

Ready, M., 1999, "The Specialist's Decision: Stopped Orders and Price Improvement," *Review of Financial Studies,* 12, 1075-1112.

Reiss, P. and D.Werner, 1998, "Does Risk Sharing Motivate Interdealer Trading?", *Journal of Finance*, 53, 5, 1657-1703.

Rust, J., J. Miller and R. Palmer, 1992, "Behavior of Trading Automata in a Computerized Double Auction Market," in Friedman, D. and J. Rust, (eds.), *The Double Auction Market: Institutions, Theories, and Evidence,* Addison-Wesley, Reading, MA, 155-198.

Rustichini, A., M. Satterthwaite and S. Williams, 1994, "Convergence to Efficiency in a Simple Market with Incomplete Information," *Econometrica,* Vol. 62, 1041-1063.

Sallach, D. and T. Wolsko, (eds.), *Proceedings of the Workshop on Simulation of Social Agents: Architecture and Institutions,* University of Chicago, 2000.

Saporta, V., 1997, "Which Interdealer Market Privails? An Analysis of Interdealer Trading in Opaque Markets," Bank of England Working Paper.

Sargent, T., 1993, *Bounded Rationality in Macroeconomics,* Oxford University Press, Oxford, UK.

Smith, M., 1993, *Neural Networks for Statisticians*, Van Nostrand Reinhold, NewYork, N.Y.

Smith, V., G. Suchanek, and A.Williams, 1988, "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica,* 56(6), 1119-1152.

Stoll, H., 2001, *Market Microstructure*, Working Paper, 01-16, Vanderbilt University.

Tirole, J., 1982, "On the Possibility of Speculation under Rational Expectations," *Econometrica,* 50, 1163-1181.

Viswanathan, S. and J. Wang, 2000, "Inter-Dealer Trading in Financial Markets?" Working Paper, Duke University.

Viswanathan, S. and J. Wang, 2002, "Market Architecture: Limit-Order Books versus Dealership Markets," *Journal of Financial Markets*, 5, 127-168.

Vitale, P., 1998, "Two Months in the Life of Several Gilt-Edged Market Makers on the London Stock Exchange," *Journal of International Financial Markets, Institutions and Money*, 8, 301-326.

Vogler, K., 1997, "Risk Allocation and Inter-Dealer Trading," *European Economic Review,* 41, 417-441.

Wener, I., 1997, "A Double Auction Model of Inter-Dealer Trading," Stanford University Working Paper.

White, H., 1989, "Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings," *Neural Networks,* 3, 535-549.

White, H., 1992, *Artificial Neural Networks: Approximation and Learning Theory*, Black Well Publishers, Cambridge, MA.

Yang, J., 2002, "The Efficiency of an Artificial Double Auction Stock Market with Neural Learning Agents," in Chen S. H. (eds.), *Evolutionary Computation in Economics and Finance*, Springer-Verlag, 85-106.

Zhang, H. and L. Kryzanowski, 2002, "Intraday Market Price Integration for Shares Cross-listed Internationally," *Journal of Financial and Quantitative Analysis*, forthcoming.