

UNIVERSITY OF CALIFORNIA  
Los Angeles

# Essays on Trading and Contract Theory

A dissertation submitted in partial satisfaction  
of the requirements for the degree  
Doctor of Philosophy in Economics

by

Linlin Ye

2016

© Copyright by  
Linlin Ye  
2016

ABSTRACT OF THE DISSERTATION

**Essays on Trading and Contract Theory**

by

Linlin Ye

Doctor of Philosophy in Economics

University of California, Los Angeles, 2016

Professor Pierre-Olivier Weill, Chair

This dissertation provides a study of optimal trading and contracting decisions, and their impacts on modern market structures. The dissertation is composed of three chapters.

Chapter 1 investigates the impact of dark pools on the informational efficiency of prices (price discovery). Traders trade an asset in either an exchange or a dark pool, with informed traders having heterogeneous private signals whose distribution is determined by an information precision level. We find that dark pools have an amplification effect on price discovery. That is, when information precision is high, adding a dark pool enhances price discovery, whereas when information precision is low, adding a dark pool impairs price discovery. The main force behind this result is a sorting effect: in equilibrium, traders with strong signals trade in exchanges, traders with modest signals trade in dark pools, and traders with weak signals do not trade. These results produce novel empirical predictions regarding dark pools that reconcile the empirical evidence. The results also provide regulatory suggestions on enhancing the informational efficiency of pricing in equity markets and in emerging markets.

Chapter 2 provides a framework to study information diffusion and interaction between a centralized and a bilateral market. In the model, traders trade to hedge their positions with some agents possessing private information. All agents can trade in the bilateral market before trading with a monopolistic market maker in a centralized market. We show that an active bilateral market functions as a channel to disseminate information. Moreover,

information diffusion depends on the centralized market liquidity: both overly liquid and overly illiquid centralized markets discourage bilateral trading, and only reasonable centralized market liquidity activates bilateral trading and hence information diffusion. We also find that information diffusion is conducted in an asymmetric way. Which type of news spreads faster depends on the conjecture of the uninformed traders. Lastly, when prices in the centralized market contain information, it may “squeeze out” trades and information diffusion in the bilateral market.

Chapter 3, which is co-authored with Jen-wen Chang, analyzes the optimal screening decision for a monopolistic firm when consumers have time-inconsistent preferences and unobservable multiple degrees of naiveté (unawareness of their time-consistency). The firm can contract to screen their degrees of sophistication, subject to profit maximization. We characterize the optimal contracts and show that the firm offers a non-screening contract when consumers have deterministic costs and offers a screening contract when consumers have random costs. We argue that the uncertainty of the consumption costs lowers the firm’s screening costs, and a discount per-usage price serves as a commitment device for the more sophisticated types. The results explain the phenomenon of the variation of contracts in sports club memberships, saving plans, and retirement programs.

The dissertation of Linlin Ye is approved.

Ichiro Obara

Joseph M. Ostroy

Tomasz Marek Sadzik

Avanidhar Subrahmanyam

Pierre-Olivier Weill, Committee Chair

University of California, Los Angeles

2016

*To the ones that I love.*

## TABLE OF CONTENTS

<b>1</b>	<b>Understanding the Impacts of Dark Pools on Price Discovery</b>	<b>1</b>
1.1	Introduction	1
1.2	Dark Pools: An Overview	7
1.3	The Model	14
1.3.1	Transaction rules in the exchange (Lit market)	16
1.3.2	Transaction rules in the dark pool	18
1.3.3	The informed speculators' problem	20
1.3.4	The uninformed liquidity traders' problem	22
1.4	The Equilibrium	24
1.4.1	Benchmark model: without a dark pool	24
1.4.2	Multi-venue model: with a dark pool	28
1.5	Dark Pool Trading and Information Structure	35
1.5.1	Information Precision and Market Characteristics	35
1.5.2	Dark Pool Impacts on Market Characteristics	39
1.5.3	Dark Pool Impacts on Price Discovery	41
1.6	Discussion: Empirical and Regulatory	47
1.6.1	A Summary of Testable Empirical Predictions	48
1.6.2	Regulatory Considerations	54
1.7	Conclusion	56
1.8	Appendix	57
1.8.1	Proof of Lemma 1	57
1.8.2	Proof of Lemma 2	58

1.8.3	Proof of Theorem 1 . . . . .	59
1.8.4	Proof of Theorem 2 . . . . .	60
1.8.5	Proof of Proposition 1 and Proposition 2 . . . . .	65
1.8.6	Proof of Proposition 3 . . . . .	72
1.8.7	Proof of Proposition 4 . . . . .	78
1.8.8	Proof of Proposition 5 . . . . .	79
<b>2</b>	<b>Information Diffusion With Centralized and Bilateral Trading . . . . .</b>	<b>81</b>
2.1	Introduction . . . . .	81
2.2	Benchmark: A Liquid Centralized Market with An Uninformed Market Maker	84
2.2.1	Setup . . . . .	84
2.2.2	Equilibrium Definition . . . . .	89
2.2.3	Equilibrium Characterization . . . . .	91
2.2.4	The Market Maker's Profit . . . . .	95
2.3	Illiquid Centralized Market and An Uninformed Market Maker . . . . .	98
2.3.1	Equilibrium Characterization . . . . .	99
2.3.2	Information diffusion and Market Interaction . . . . .	103
2.3.3	Market Maker's Profit and Optimal Trading Limits . . . . .	107
2.4	Illiquid Centralized Market and An Informed Market Maker . . . . .	108
2.5	Conclusion . . . . .	111
<b>3</b>	<b>Screening Decisions with Time-Inconsistent Agents . . . . .</b>	<b>113</b>
3.1	Introduction . . . . .	113
3.2	The Model . . . . .	116
3.2.1	Model Set-up . . . . .	116



3.3	Deterministic Cost . . . . .	120
3.3.1	Deterministic Cost: screening Contract . . . . .	121
3.3.2	Deterministic Cost: non-screening contract . . . . .	124
3.3.3	Deterministic Cost: to screen or not to screen? . . . . .	125
3.4	Random Cost . . . . .	126
3.4.1	Random Cost: to screen or not to screen . . . . .	131
3.4.2	Random Cost: a numerical example . . . . .	132
3.5	An Extension: The Multi-period Model . . . . .	137
3.6	Conclusion . . . . .	142

## LIST OF FIGURES

1.1	<b>Dark Pool Market Share.</b> The plot shows the annual data of dark pool volume as a percentage of the total consolidated volume in the US, Europe, and Canada. <b>Data source:</b> US data (2008 - 2015) is from “Rosenblatt Securities: <i>Let There Be Light</i> , January 2016 Issue” and Europe data (2010 - 2015) is from “Rosenblatt Securities: <i>Let There Be Light – European Edition</i> , January 2016 Issue”. Figures in Canada (2007 - 2015) are derived from reports of IROC. . . . .	9
1.2	<b>Average Trade Size.</b> The plot shows the annual average trade size of US dark pools, NYSE and NASDAQ, from 2009 to 2015. <b>Data source:</b> Rosenblatt Securities. . . . .	11
1.3	Trading Timeline . . . . .	24
1.4	Payoffs For Traders, Single-venue . . . . .	26
1.5	Payoffs For Traders, Multi-venue . . . . .	29
1.6	Strategy of Informed Traders . . . . .	30
1.7	Strategy of Uninformed Traders . . . . .	32
1.8	<b>Transaction Costs.</b> The left-hand figure shows the normalized spreads on the exchange and how they vary with $\log(\sigma_e)$ ; the right-hand figure shows the adverse selection cost in the dark pool and how it vary with $\log(\sigma_e)$ . In both figures, $\log(\sigma_v) = 0$ . . . . .	37
1.9	<b>Participation Rates.</b> The left figure plots the expected participation rates of the uninformed and how they vary with $\log(\sigma_e)$ . The right one shows the participation rates for the informed traders how they vary with $\log(\sigma_e)$ . In both plots, $\log(\sigma_v) = 0$ , $\mu_z = 60$ , $\mu = 30$ . . . . .	38
1.10	<b>Price Discovery.</b> The left plots the dark pool impact on price discovery with 2-dimension: $\sigma_v$ and $\sigma_e$ . The right plots the “Signal-to-noise” ratio $I(\overline{\gamma_e}, \underline{\gamma_e}, \alpha_e)$ as a function of $\sigma = \frac{\sigma_v}{\sigma_e}$ . . . . .	45

1.11	<b>The Likelihood DPs Harm Price Discovery <math>\bar{\sigma}_v</math>.</b> The left-hand figure plots the threshold $\bar{\sigma}_v$ as a function of the relative size of informed traders, $\frac{\mu}{\mu_z}$ . The right-hand shows the threshold $\bar{\sigma}_v$ as a function of the information precision, $-\log(\sigma_e)$ . On the left, $\log(\sigma_e) = 0$ . On the right, $\frac{\mu}{\mu_z} = .2$ . . . . .	48
1.12	<b>Execution Probability and Trade-off of A Dark Pool.</b> The left figure plots the non-execution probability as a function of $\log(\sigma_e)$ . The right-hand figure plots the non-execution probability as a function of the exchange spread $A/\sigma_v$ . In both plots, $\log(\sigma_v) = 0$ . . . . .	49
1.13	<b>Dark Volumes and Market Share.</b> The left figure plots the dark pool volume, total volume and dark pool market share as a function of $\log(\sigma_e)$ . The right-hand figure shows the dark pool market share as a function of the exchange spread $A/\sigma_v$ . In both plots $\log(\sigma_v) = 0$ . . . . .	51
1.14	<b>Predicability of Dark Pool Trades.</b> The left figure plots the dark pool “Predictive Fraction” as a function of spread $A/\sigma_v$ . The right-hand plots the dark pool adverse selection costs as a function of spread $A/\sigma_v$ . . . . .	52
1.15	Equilibrium Existence . . . . .	63
1.16	$h(s)$ and $\hat{h}(s)$ . . . . .	74
2.1	Timeline (Benchmark Model) . . . . .	87
2.2	<b>The Market Maker’s profit As <math>\alpha</math> Varies</b> (no trading limitation, no information for the market maker). $M0$ represents the point where the market maker breaks even. The parameters are $\gamma = 0.8, \bar{z} = 0.3$ . . . . .	97
2.3	<b>The Market Maker’s Profit As <math>[-\bar{y}, \bar{y}]</math> Varies.</b> The graph plots how the market maker’s profit varies as $[-\bar{y}, \bar{y}]$ becomes wider. . . . .	107
3.1	Timeline . . . . .	116
3.2	<b>An Example of Decreasing Density Functions</b> . . . . .	133

3.3	$\phi(p)$ . The plot shows the function $\phi(p) = \int_{\hat{\beta}_1 \delta b - p}^{\hat{\beta}_2 \delta b - p} -p - c + \delta b dF(c)$ . The distribution is an exponential distribution with parameter $\lambda$ . The other parameters assigned are: $\hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1$ . . . . .	134
3.4	<b>Indifference Curves.</b> The plot shows the indifferent curves for both type 1 and type 2 consumers. The parameters are $\lambda = 0.1, \alpha = .4, \beta = 0.3, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1$ . . . . .	135
3.5	<b>Indifference Curves (Multi-period Model).</b> Parameters are $\lambda = 0.1, \beta = 0.3, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1$ . . . . .	141

## LIST OF TABLES

<p>2.1 <b>Markets Interaction.</b> This table studies the relationship between the trading limits, <math>\underline{y}, \bar{y}</math>, and bilateral trading activities. Parameters are <math>\gamma = 0.8, \alpha = 0.1, \bar{z} = 0.3, \delta_I = 0.7</math>. Each column represents a trading limit, <math>[\underline{y}, \bar{y}]</math>. <math>\Pi</math> is the profit of the Market maker, while <math>p</math> is the equilibrium market maker price in the centralized market. <math>q^U, q^H, q^L</math> are, respectively, the equilibrium bilateral prices that will be proposed by an uninformed trader, high-sigaled informed trader and low-sigaled informed trader. “Bi.T” indicates whether or not there exists trades in bilateral market, and “Info” represents the information diffusion in the bilateral market. . . . .</p>	105
<p>2.2 <b>Information Diffusion When Uninformed Proposers Are “Biased” Towards Learning Good News.</b> This table studies the result of information diffusion after an active bilateral trading period, in a case when the uninformed proposers are “biased” towards learning signal <math>H</math>. “Fraction” is the population fractions of all matching scenarios. Yes(resp.) and Yes(prop.) represent the case that the uninformed agents learn the full signal as a responder and as a proposer, respectively. Partial(prop.) presents the case that the uninformed agents learn the signal partially as a proposer. . . . .</p>	106
<p>3.1 <b>Optimal contracts and profits as <math>\lambda</math> varies.</b> Parameters are <math>\alpha = .4, \beta = 0.3, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1</math>. . . . .</p>	136
<p>3.2 <b>Optimal Contracts and Profits As <math>\alpha</math> Varies.</b> Parameters are <math>\lambda = 0.1, \beta = 0.3, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1</math>. . . . .</p>	137
<p>3.3 <b>Optimal Contracts and Profits As <math>\lambda</math> Varies (Multi-period Model).</b> <math>u_i</math> represents the perceived utility of a type <math>i</math> consumer and <math>\tilde{u}_i</math> the actual utility. Parameters are <math>\beta = 0.3, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1</math>. . . . .</p>	142
<p>3.4 <b>Optimal Contracts and Profits As <math>\alpha</math> Varies (Multi-period Model).</b> <math>u_i</math> represents the perceived utility of a type <math>i</math> consumer and <math>\tilde{u}_i</math> the actual utility. Parameters are <math>\lambda = 0.1, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1</math>. . . . .</p>	142

## ACKNOWLEDGMENTS

I will be always be indebted to my advisor, Pierre-Olivier Weill, for his great encouragement through my years as a graduate student and his invaluable inputs on all my research projects. I am also extremely grateful to Avanidhar Subrahmanyam, Tomasz Sadzik, and Mark Garmaise for their immense support in my studies by going through many versions of my papers and giving me incredible advice. I am particularly obliged to Joseph Ostroy, John Wiley, Obara Ichiro, Simon Board, and Ivo Welch for their extraordinary encouragement and feedbacks on my research during my presentations and our bilateral meetings. I would like to thank William Zame for his great initial guidance in my Chapter Two. Also, special thanks go to Antonio Bernardo, Francis Longstaff, Bernard Herskovic, William Mann, Daniel Andrei, Shuyang Sheng, and Jernej Copic for their wonderful support and feedbacks on my research. I also thank Alex Kemmsies in Rosenblatt Rosenblatt Securities and Sayena Mostowfi in TABB Group for providing the industry data and analysis in Chapter One. I would really like to thank my parents for their precious and unconditional love that shines through all my life. Finally, I thank my great friends in the UCLA Economics Department. The friendship we have means tremendously for me. Special gratitude goes to Simpson Zhang, Jen-Wen Chang, and Ming Yan, who are always there for me as my coauthors, my colleagues, and my best friends.

## VITA

- 2006 B.A. (Business Management), University of International Business and Economics.
- 2010 M.S. (Mathematical Finance), Central University of Finance and Economics.
- 2012 M.A. (Economics), University of California Los Angeles.

# CHAPTER 1

## Understanding the Impacts of Dark Pools on Price Discovery

### 1.1 Introduction

Over the years, the world financial system has experienced a widening of equity trading venues, among which dark pools have rapidly grown in popularity. The market share of dark pools in the US has grown from 7.51% in 2008 to 16.57% in 2015.<sup>1</sup> In contrast with a traditional exchange, dark pools do not publicize information about their orders and best price quotations before trade. Unlike a stock exchange in which prices are formed to clear the buy and sell orders, a typical dark pool executes orders using a price derived from the stock exchanges. Those dark pools do not contribute to the process of information aggregation, and hence they do not offer price discovery (i.e., the process and efficiency of prices aggregating information about asset value). Price discovery is essential to achieving the confidence of a broad community of market participants and ensuring the efficiency of capital markets. Therefore, the question of whether dark pool trading will harm price discovery has become a rising concern and matter of debate for regulators, and industry practitioners.<sup>2</sup> Academic

---

<sup>1</sup>Rosenblatt Securities: *Let There Be Light*, January 2016 Issue.

<sup>2</sup>For example, as remarked by the SEC Commissioner Kara M. Stein before the Securities Traders Association's 82nd Annual Market Structure Conference in Sep. 2015, "As more and more trading is routed to dark venues that have restricted access and limited reporting, I am concerned that overall market price discovery may be distorted rather than enhanced." According to "An objective look at high-frequency trading and dark pools," a report released by PricewaterhouseCoopers (2015), "Dark pools may harm the overall price discovery process, particularly in a security in which a significant portion of that security's trade volume is in the pools."



research, for its part, has yielded conflicting results. [Ye \(2011\)](#) predicts that, in theoretical studies, the addition of a dark pool strictly harms price discovery. By contrast, [Zhu \(2014\)](#) predicts that dark pools strictly improve price discovery. Empirically, there are findings that support each of the different predictions.

This paper investigates the question whether dark pool trading will harm price discovery. We find a novel *amplification effect* of dark pools on price discovery: price discovery will be enhanced when price discovery is high and will be impaired when price discovery is low. We model the decisions of traders who enter the markets to trade an asset: a) trade in an exchange, b) trade in a dark pool, or c) do not trade. Informed traders have heterogeneous private signals, and the distribution of these signals' strengths is determined by the information precision level. Uninformed liquidity traders have heterogeneous demands for liquidity. In equilibrium, there is a *sorting effect*: for informed traders, those with strong signals trade in exchanges, those with modest signals trade in dark pools, and those with weak signals do not trade. For uninformed liquidity traders, those with high liquidity demand trade in the exchange, those with modest liquidity demand trade in the dark pool, and those with low liquidity demand delay trade. We show that price discovery is an increasing function of the information precision level, and information precision determines a dark pool's impact on price discovery. When information precision is high (meaning price discovery is high), adding a dark pool enhances price discovery, whereas when information precision is low (meaning price discovery is low), adding a dark pool impairs price discovery.

The results highlight the importance of the information structure (information precision) in pricing efficiency when markets are fragmented by dark pools. We show that the results have immediate policy implications for enhancing price discovery in equity markets and dark pool usage in emerging economies. The results also help to reconcile the seemingly contradictory empirical findings and generate novel empirical predictions regarding the information content of dark pool trades, dark pool market share, and their relationships with exchange spread.

The *sorting effect* exists in a *noisy* information structure (imprecise information) which

is absent in the current literature. We assume that informed traders receive noisy signals, and the distribution of the signals is determined by the information precision level. More specifically, in the model, there are informed speculators and uninformed liquidity traders. Informed speculators receive heterogeneous signals, whereas uninformed liquidity traders have various degrees of liquidity demand. Both types of traders choose among three options: a) to trade an asset in an exchange, b) to trade an asset in a dark pool, or c) to not trade (delay trade for liquidity traders). The exchange is modeled as market makers posting bid-ask prices and guaranteeing execution, whereas the dark pool is modeled as a crossing-mechanism that uses the average of bid and ask (mid-price) in the exchange to execute orders (if there are more buy orders than sell orders, buy orders are executed probabilistically, with some buy orders not executed, and vice versa). Trading in dark pools has a trade-off: compared with the exchange, dark pools provide better prices, but this is offset by a non-execution probability. Therefore, amongst informed traders, those with strong signals prefer an exchange because they are very confident about making profits and desire a guaranteed execution more than a better price; those with modest signals prefer a dark pool because they are less confident about making profits and desire a better price more than execution; and finally, those with weak signals prefer not to trade because they are unconfident about making profits. A similar argument holds for liquidity traders.

The *amplification effect* is a result of the sorting effect. In equilibrium, the choice of an informed trader is determined by the strength of his or her signal, whereas the choice of an uninformed liquidity trader depends on the degree of his or her liquidity demand. Different information precision levels result in different distributions in the strengths of signals and hence different choices for the majority of the informed traders, which result in different dark pool impacts on price discovery. When information precision is high, the majority of informed traders have strong signals and prefer an exchange. Adding a dark pool attracts only a small fraction of informed traders, compared with the liquidity traders, leaving a higher ratio of informed to uninformed traders in the exchange and improving price discovery. In contrast, when information precision is low, the majority of informed traders have modest signals and

prefer the dark pool, and adding a dark pool attracts a higher fraction of informed traders compared with liquidity traders, lowering the ratio of informed to uninformed traders in the exchange and decreasing price discovery.

This paper points out an important function of dark pools not yet discussed in the existing literature: dark pools help informed traders mitigate their information risk, that is, the loss that is attributable to wrong information. When traders' information is relatively weak (meaning there is a higher probability that it is wrong), they face a high risk of losing money in trading. Dark pools provide those traders a perfect "buffer zone" – a place that strictly lowers their information risk. This function of dark pools is only present, however, when traders have a noisy information structure.

To the best of our knowledge, this paper is the first to introduce a noisy information structure in a fragmented market to study dark pools and price discovery. Examining the noisiness in information is of essential importance, not only because it is much more realistic than assuming perfect information, but also because it reveals the process of price discovery by identifying the motivations of traders' choices. As a result, our predictions are more robust in the sense that the sorting and amplification effects hold in every equilibrium. In contrast, the current theoretical literature assumes that all informed traders have perfectly precise information. This obscures trading motivations and induces instability in the results. For example, [Zhu \(2014\)](#) studies some equilibria in which dark pools improve price discovery, but there may exist other equilibria in his model in which dark pools harm discovery. Yet, [Zhu \(2014\)](#) does not discuss these equilibria.

Our findings have immediate policy implications for the ongoing debate over dark pool usage. Our findings imply that, in contrast with current literature, there is no uniform impact that dark pools have on price discovery and other measures of market quality. Dark pool activity and its impacts display significant cross-sectional variation and should be evaluated differently across various economic environments. Concrete suggestions for regulators to enhance pricing efficiency include: (i) identifying firm characteristics and monitoring dark pool trades in firms that are likely to have a negative dark pool impact, such as high R&D firms,

young firms, small firms, and less analyzed firms, (ii) facilitating information transmission and processing, enhancing accounting and reporting disclosure systems, and improving the efficiency of the judicial systems and law enforcement against insider trading, and (iii) being cautious in emerging markets with regards to dark pool trading, given that most emerging markets are regulated by poor legal systems that lack implemental power against insider trading and have a low precision in information disclosure. A more detailed discussion is provided in Section [1.6.2](#).

Our study also produces testable predictions and helps to reconcile the seemingly contradictory results in the current empirical literature. One of the predictions that could motivate empirical and regulatory concerns is how much dark pool trades can forecast price movements. We predict that the information content of dark pool trades has an inverted U-shape relationship with the liquidity level (exchange spread), implying that assets with modest liquidity have the highest information content in their dark pool trades, whereas the most liquid and illiquid assets have the lowest information content in their dark pool trades. There are also some predictions which coincide with current theoretical literature. For example, dark pool usage also has an inverted U-shape association with exchange spread. Dark pools create additional liquidity for the market. A more detailed discussion is in Section [1.6.1](#).

**Related Work:** There is a large collection of studies that examines information asymmetry and price discovery in financial markets, in both the theoretical and empirical fields. In theoretical studies, a large set of papers analyze non-fragmented markets, including the two pioneering works in price discovery, [Glosten and Milgrom \(1985\)](#), and [Kyle \(1985\)](#). Other studies examine fragmented lit markets, for example [Viswanathan and Wang \(2002\)](#), [Chowdhry and Nanda \(1991\)](#), and [Hasbrouck \(1995\)](#). There are a handful of papers that study information asymmetry in a market fragmented by lit and dark venues (see, e.g., [Hendershott and Mendelson 2000](#), [Degryse et al. 2009](#), [Buti et al. 2011a](#)). Yet, these models assume either non-freedom of choice for traders or exogenous prices. Our study, on the other hand, considers free venue selection for traders and endogenous prices. This paper is closely related to [Zhu \(2014\)](#) whose trading protocols are the same as ours. But unlike [Zhu \(2014\)](#)

who considers an exact information structure, we examine a noisy information structure. Under this noisy information structure, we predict different results in price discovery and other measures from [Zhu \(2014\)](#). When the informational noise is absent in our model (i.e., information noise converges to zero), our prediction of price discovery coincides with [Zhu \(2014\)](#)'s. Our paper is also related but divergent from [Ye \(2011\)](#). Whereas our model considers free selection of traders, [Ye \(2011\)](#) assumes that uninformed traders are not subject to free-choice between different venues, and hence the corresponding piece of the pricing mechanism is missing. In our model, if we fix the choices of uninformed traders and only allow informed traders to choose between venues, our prediction also coincides with [Ye \(2011\)](#).

Empirical works report conflicting results regarding dark pool impact on price discovery. These results are within the predictions of our study. For example, [Buti et al. \(2011b\)](#), [Jiang et al. \(2012\)](#), and [Fleming and Nguyen \(2013\)](#) support an improvement for price discovery with dark trading, while [Hatheway et al. \(2013\)](#), and [Weaver \(2014\)](#) discover a diminishment in price informativeness. Also, [Hendershott and Jones \(2005\)](#) find a negative impact for dark trading on price discovery, while [Comerton-Forde and Putniņš \(2015\)](#) find that, cross-sectionally, dark pool trading improves price discovery when the proportion of non-block dark trades are low (below 10%, suggesting a low fraction of informational content) and harms price discovery when the proportion of non-block dark trades is high.

There are also other empirical studies that focus on dark pool operation and other measures of market quality. Some papers analyze the information content of dark pool trades. For example [Peretti and Tapiero \(2014\)](#) find that dark trades can predict price movement. Some study the trade-offs of dark trading. For example, [Gresse \(2006\)](#), [Conrad et al. \(2003\)](#), [Næs and Ødegaard \(2006\)](#) , and [Ye \(2010\)](#) study the execution probability in dark pools. Another category studies the association between dark trading and the exchange spread. My study predicts the same inverted U-shape as [Ray \(2010\)](#) and [Preece \(2012\)](#). My study also suggests a cross-sectional variance and provides insights in explaining the contradictory results reported in other papers. For example [ASIC \(2013\)](#), [Comerton-Forde and Putniņš \(2015\)](#), [Degryse et al. \(2015\)](#), [Hatheway et al. \(2013\)](#), and [Weaver \(2014\)](#) find a positive

association while [O’Hara and Ye \(2011\)](#) and [Ready \(2014\)](#) find a negative association between dark pool market share and exchange spread. Others find cross-sectional differences (see, e.g., [Nimalendran and Ray 2014](#), [Buti et al. 2011b](#)). A more detailed discussion of the relationship between our predictions and the current empirical literature is provided in [Section 1.6.1](#).

## 1.2 Dark Pools: An Overview

Over the last decade, numerous trading platforms have emerged to compete with the incumbent exchanges. Today, in the U.S. investors can trade equities in approximately 300 different venues. According to TABB (Oct. 2015),<sup>3</sup> as of June 2015, there are 11 exchanges, 40 active dark pools, a handful of ECNs, and numerous broker-dealer platforms that are operating as equity trading venues in the U.S.<sup>4</sup>

Among those venues, dark pools are a type of equity trading venue that does not publicly disseminate the information about their orders, best price quotations, and identities of trading parties before and during the execution.<sup>567</sup> The term “dark” is so named for this lack of transparency. Dark pools emerged as early as the 1970s as private phone-based networks between buy-side traders (See [Degryse et al. \(2013\)](#)). In the early days, the success of these

---

<sup>3</sup> “US Equity Market Structure: Q2-2015 TABB Equity Digest,” TABB Group, Oct. 2015.

<sup>4</sup>In Europe, according to [Gomber and Pierron \(2010\)](#) there are around 32 dark pools operating in equity markets. In Australia, from [ASIC \(2013\)](#), there are 20 dark trading venues operating.

<sup>5</sup>Although the information about orders are hidden before trade, the after executed trades are not: executed trades are recorded to the consolidate tape right after the trade. SEC requires reporting of OTC trades in equity securities within 30 seconds of execution. Also, dark pools are required to report weekly aggregate volume information on a security-by-security basis to FINRA.

<sup>6</sup>SEC Reg NMS Rule 301 (b) (3) requires all alternative trading systems (ATSs) that execute more than 5% of the volume in a stock to publish its best-priced orders to the consolidated quote system. However, it only applies if the ATS distributes its orders to more than one participant. If it does not provide information about its orders to any participants, it is exempt from the quote rule.

<sup>7</sup>Electronic Communication Networks (ECNs) are registered as a type of ATS. But unlike dark pools, ECNs display orders in the consolidated quote stream.

trading venues was limited, but this has changed substantially in the last decade. Dark pools have experienced a rapid growth of trading activity in the U.S., Europe and Asia-Pacific area. Figure 1.1 shows the annual data on the market share of dark pool trading as of the consolidated volume in the U.S., Europe, and Canada, updated to 2015.<sup>8</sup> According to the data, the U.S. market share of dark pools increased from about 7.51% in 2008 to 16.57% in 2015. The dark pool market shares in Europe and Canada are less, but they exhibit the same growth trend. In Australia, according to the Australian Securities & Investments Commission (ASIC 2013), as of June 2015 dark liquidity consists of 26.2% of total value that traded in Australian equity market.<sup>9</sup>

One reason behind the rapid growth of dark pool trading is the technology development in electronic trading algorithms. Advances in technology have made it easier to automatically optimize routing and execution according to different sets of considerations and trading protocols. Another reason for the proliferation is the regulation changes that have been made to encourage competition between trading venues. For example, in the U.S., Regulation NMS (National Market System) was revised and reformed in 2005 to encourage the operation of various platforms, and as a consequence, a wide variety of trading centers have been established since then. Another example is the introduction of the Market in Financial Instruments Directive (MiFID) in the European Union in 2007, which spurred the creation of new trading venues, including dark pools.<sup>10</sup>

---

<sup>8</sup>We estimated Canadian dark pool market share from “Report of Market Share by Marketplace—Historical (2007-2014),” IIROC, Aug 2015, “Report of Market Share by Marketplace (historical 2015–Present),” IIROC, May, 2016. Precisely, we estimate the market share of the following 4 dark pools operating in Canada: Liquidnet, Matchnow, Instinet, and SigmaX Canada.

<sup>9</sup>Australian Securities & Investments Commission, “Equidity Market Data,” June 2015. The number contains 12% block size dark liquidity and 14.1% non-block size dark liquidity. It describes all the hidden orders in the markets including those in exchanges and dark pools.

<sup>10</sup>In recent years, however, as the debate about dark pool usage has escalated, many countries have started to consider restrictions on dark trading. For example, Canada and Australia have required dark liquidity to provide a “meaningful price improvement” of at least one trading increment (i.e., one cent in most major markets), and US regulators have also been contemplating imposing such restrictions. In recent years, US regulators start to strengthen law enforcement against dark pools and urged their upgrading in operation. These cases include UBS Securities (Jan 2015), Goldman Sachs Execution & Clearing, L.P (SIGMA X, July 2015), and Barclays (Jan 2015).

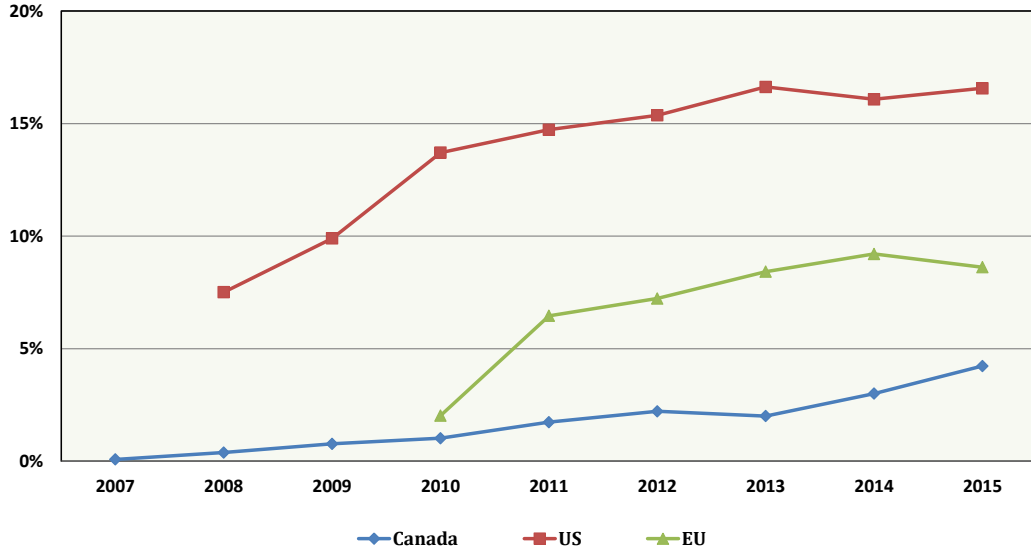


Figure 1.1: **Dark Pool Market Share.** The plot shows the annual data of dark pool volume as a percentage of the total consolidated volume in the US, Europe, and Canada.

**Data source:** US data (2008 - 2015) is from “Rosenblatt Securities: *Let There Be Light*, January 2016 Issue” and Europe data (2010 - 2015) is from “Rosenblatt Securities: *Let There Be Light – European Edition*, January 2016 Issue”. Figures in Canada (2007 - 2015) are derived from reports of IIROC.

There are two key commonalities in dark pools’ operating protocols: the pricing mechanism and execution mechanism. First, dark pools generally do not provide price discovery. Instead, they typically use a price derived from an existing primary market as their transaction price. The most commonly used pricing mechanism is the mid-point mechanism: a pricing method to cross orders at the concurrent mid-point of the National Best Bid and Offer (NBBO).<sup>11</sup> Second, unlike exchanges where orders are cleared at the exchange price, in most of the dark pools, orders don’t clear. Instead, dark pools adopt a “rationing” mechanism to execute orders. That is, traders anonymously place unpriced orders to the pool, and the orders are matched and executed probabilistically – orders in the shorter side are executed for sure, whereas orders in the longer side are rationed probabilistically.

<sup>11</sup>Nimalendran and Ray (2014) document the usage of such a pricing mechanism in their dark trading sample and find that not all trades are at the midpoint of NBBO, but about 57% transactions are within .01% of the price around the midpoint. In this paper, we follow the majority and adopt the mid-point pricing mechanism.



The pricing and execution mechanisms of dark pools' operation reflect the trade-off of trading in a dark pool for an individual trader. On the one hand, dark pools have lower transaction costs than exchanges (typically because orders are executed within the NBBO, with the "trade-at rule" further enhancing such price improvement), and lessen the price impact for big orders. On the other hand, investors suffer a lower execution rate compared with the exchange. Gresse (2006) found that the execution probability in the two dark pools in his dataset was only 2-4 percent, while Ye (2010) documents a dark pool execution probability of 4.11% (NYSE listed) and 2.17% (NASDAQ listed) in his dataset, in comparison with a probability of 31.47% and 26.48% for their exchange counterparts.<sup>12</sup>.

The dark pools' participating constituent base has evolved over time. In the early years, dark pools were designed as venues where large, uninformed traders transact blocks of shares to reduce price impact. This is possible because dark pools are not subject to NMS fair access requirements and can thus prohibit or limit access to their services (see Reg ATS Rule 301(b)(5)). In recent years, however, this has changed greatly. According to an industry insider in Rosenblatt Securities Inc., "it can be assumed that most pools are open to most investors connecting to the pool, provided the investors do not violate any codes of conduct." A measure of such a change is reflected in the trading sizes of dark pools. Figure 1.2 shows the average trading size in the U.S. According to the data, the US average trading size in dark pools and exchanges (NYSE and NASDAQ) have been started to converge since 2011, highlighting the fact that the participating constituents in these venues have become more and more similar. It implies that the exclusivity of a dark pool to informed traders has been weakened . As a result, more prominence has been attached to the issue of the potential impact of dark pools on price discovery, because as more informed traders obtain access to dark pools, their migration to dark pools may hurt the information aggregation process in the exchange,<sup>13</sup>.

---

<sup>12</sup>Nowadays, a rising concern of dark pools is their vulnerability to predatory trading by High Frequency Traders (HFTs) (See Mittal (2008), Nimalendran and Ray (2014), ASIC (2013) for instance.)

<sup>13</sup>This paper, as well as Zhu (2014) considers full access for informed traders.

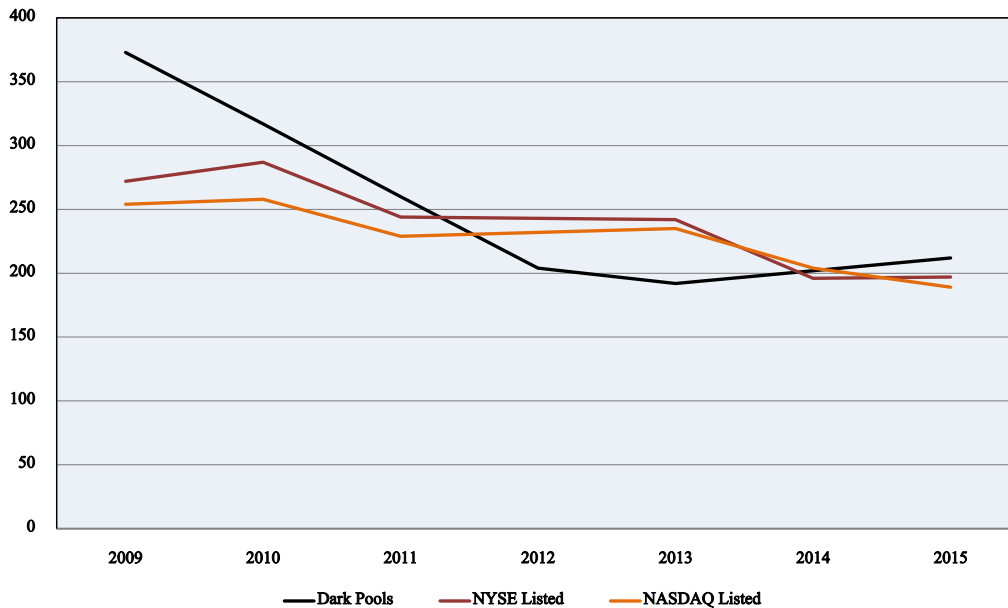


Figure 1.2: **Average Trade Size.** The plot shows the annual average trade size of US dark pools, NYSE and NASDAQ, from 2009 to 2015.

**Data source:** Rosenblatt Securities.

Dark pools are heterogeneous. The types of dark pools can be classified according to different characteristics based on their ownership structure, pricing access, operation mechanism, constituency and other factors. All of these categories are in constant flux for the dark pools. Most of the pools also overlap in one or more categories as well, only the owner types remain constant overtime. We provide a discussion on some characteristics and their examples.

**(i) Pricing.** Dark pools use three primary pricing mechanisms. The execution will take place once two sides of a suitable trade are matched. The three pricing mechanisms are automatic pricing (usually at the midpoint of the best bid and offer), derived pricing (for example, average price during the last five minutes), and negotiated pricing (for example, Liquidnet Negotiatoin offers availability of one-to-one negotiation of price and size).

**(ii) Order Type.** There are primarily three types of order that prevails in dark pools: limit orders (to buy or to sell a security at a desired price or better), peg orders (peg to

the NBBO, for example midpoint or alternate midpoint,<sup>14</sup>) and immediate or cancel order (IOC). A dark pool may accept a subset of these order types. Pools that accept limit orders may offer some price discovery (usually within the NBBO). These pools include, for example, Credit Suisse’s CrossFinder, Goldman Sachs’ Sigma X, Citi’s Citi Cross, and Morgan Stanley’s MS Pool. Pools that execute peg orders do not provide price discovery. These include, for example, Instinet, Liquidnet, and ITG Posit. Pools accepting only IOC orders are single dealer platforms (SDP), where the operator works as market makers and customers interact solely with the operator’s own desk (for example, Citadel Connect and Knight Link by KCG<sup>15</sup>).

**(iii) Execution Frequency and Order Information.** There are three modes of execution: scheduled crossing, continuous blind crossing, and indicated market.<sup>16</sup> The scheduled crossing networks include BIDS, ITG POSIT Match, and Instinet US Crossing. In scheduled crossing networks, the two sides of a trade cross during a set period. These networks typically do not display quotes but may have an order imbalance indicator. Continuous blind crossing networks continuously cross orders for which no quotes are given. Indicated markets cross orders using participants’ indications of interest (IOIs) and provide some level of transparency in order to attract liquidity. Liquidnet and Merrill Lynch offer variations on this theme.<sup>17</sup>

**(iv) Customer Base and Exclusivity.** There are dark pools which design their rules and monitor trading in an attempt to limit access to buy-side (natural contra-side) institutional investors. According to [Boni et al. \(2013\)](#), Liquidnet “Classic” is one of those. A

---

<sup>14</sup>Traders are able to specify premiums or discounts vis-à-vis the mid when placing a trade. For example, a motivated buyer may specify an order that promises to pay the mid plus a penny. This would give this trade priority over all other buy orders.

<sup>15</sup>Getco LLC once operated an SDP called GetMatched. Following the 2013 merger of Knight Capital Group and Getco LLC, GetMatched was decommissioned.

<sup>16</sup>See [DeCovny \(2008\)](#).

<sup>17</sup>Pipeline, a well-known dark pool using IOIs, settled allegations that it misled customers and was shut in May 2012.

measure of the exclusivity is the average trading size of a dark pool. In May 2015, among the 40 active dark pools operating in the US, there are 5 dark pools in which over 50% of their Average Daily Volumes are block volume (larger than 10k per trade). Those pools can be regarded as “Institutional dark pools,” and they include Liquidnet Negotiated, Barclays Directx, Citi Liquifi, Liquidnet H20, Instinet VWAP Cross, and BIDS Trading. Other dark pools have percentages of block volumes less than 15%, with most of them lower than 2%.<sup>18</sup>

**(v) Ownership Structure.** According to Rosenblatt (2015), dark pools can be classified into four categories according to their ownership structure. This is the only classification that does not fluctuate over time. The four categories include the Bulge Bracket/Investment bank, Independent agency, Market maker, and Consortium-sponsored. In May 2015, The market shares of the four categories are, respectively, 55.28%, 24.11%, 13.79%, and 6.82%. Examples of the Bulge Bracket/Investment bank-owned dark pools are CS Crossfinder, UBS ATS, DB SuperX, and MS Pool. Independent agency owned pools include, for example, ITG POSIT, Instinet CBX, ConvergEx Millennium. Market maker owned pools include Citadel Connect and Knight Link by KCG, and Consortium-sponsored pools include Level and BIDS.<sup>19</sup>

Finally, “dark pools liquidity” is not equivalent to “dark liquidity.” Dark liquidity, or dark volume, is a broader concept since it measures the total non-displayed market volume. Exchanges, for example, can contain “dark” volumes, which are applied through iceberg orders and workup processes. According to the TABB group’ classification, dark volume can break down into retail-wholesaler, dark pool volume, and hidden exchange volume. As of Q2-2015, the percentages of each are 40.1%, 39.7%, and 20.2% respectively. In total, the dark volume was 43.9% of the consolidated volume.<sup>20</sup>

---

<sup>18</sup>“Let There Be Light , Jun 2015,” Rosenblatt Securities, Inc.

<sup>19</sup>“Let There Be Light , Jun 2015,” Rosenblatt Securities, Inc.

<sup>20</sup>“US Equity Market Structure: Q2-2015 TABB Equity Digest,” TABB Group, Oct. 2015.

### 1.3 The Model

The model considers an economy that lasts for three periods. We index the periods by 0, 1, and 2. There is one risky asset that is traded during the two periods with an uncertain fundamental value

$$\tilde{v} = \begin{cases} -\sigma_v, & \text{with probability } \frac{1}{2}, \\ \sigma_v, & \text{with probability } \frac{1}{2}. \end{cases}$$

That is to say, the risky asset has an unconditional mean zero and standard deviation  $\sigma_v$ . In period 0,  $\tilde{v}$  is realized, but this information is not revealed to the public.

There are two types of traders who are potentially interested in the risky asset: informed speculators and uninformed liquidity traders. We assume that they are all risk-neutral. There is a continuum of informed speculators with measure  $\mu$ , a continuum of uninformed liquidity buyers with measure  $Z^+$ , and a continuum of liquidity sellers with measure  $Z^-$ . We assume that  $Z^+, Z^-$  are identical and continuously distributed random variables on  $[0, +\infty)$ , with mean  $\frac{1}{2}\mu_z$ .  $Z^+, Z^-$  are also realized at period 0 so that liquidity buyers and liquidity sellers arrive at the market at the same time. The realizations of  $Z^+, Z^-$  are not observed by any market participants.

In period 0, each informed speculator receives his or her own private signal regarding the value of the asset,  $s_i = \tilde{v} + e_i$ , where  $i$  is the index of informed traders and  $e_i$  represents the noise of the signal.<sup>21</sup> We assume that  $e_i$  are identically independently distributed normal random variables, with mean 0 and standard deviation  $\sigma_e$ . Therefore, in the first period, they trade on both their private information and public information (if there is any). They can trade (either buy or sell) up to 1 unit of the asset. If there are more than one venue to trade, they can split their orders. Without loss of generality, we assume that informed

---

<sup>21</sup>According to [Gyntelberg et al. \(2010\)](#), there are various types of private information that stock market investors may have about the fundamental determinants of a firm's value, including knowledge of the firm's products and innovation prospect, management quality, and the strength and likely strategies of the firm's competitors. Private information may also include passively collected information about macro-variables and other fundamentals which may be dispersed among customers. Equity market order flow to a large degree reflects transactions by investors who are very active in collecting private information. A more detailed discussion is in section [1.6.2](#).

speculators only trade in period 1.<sup>22</sup> The model is distinctive to [Zhu \(2014\)](#) in the information structure. [Zhu \(2014\)](#) assumes that all informed traders receive exact signals about the asset, whereas we consider a noisy information structure.<sup>23</sup> The introduction of a richer information structure is crucial to our analysis, not only because it is more realistic, but also because it reveals a sorting effect of market fragmentation on information. That is, in equilibrium, traders with strong signals trade in the exchange, traders with modest signals trade in the dark pool, and traders with weak signals do not trade. This sorting effect is the major economic force in the trader’s venue-selection and the process of price discovery. The absence of such an effect will likely cause instability of predictions in multiple equilibria, such as discussed in [Zhu \(2014\)](#). A more detailed discussion is in Section [1.4.2](#).

A liquidity buyer (seller) comes to the market to buy (sell) 1 unit of the risky asset. Similarly, one can split their orders if there exist multiple transaction venues. The uninformed liquidity traders, however, do not have any private information. They enter the market to meet their liquidity demands. The level of their liquidity demand is measured by a delay cost, a cost that reflect how urgent one needs his or her order to be fulfilled in period 1. More precisely, if a liquidity trader, buyer or seller, cannot have his or her order executed in period 1, a delay cost is incurred. The delay cost (per unit) is represented by  $\sigma_v d_j$ , where  $j$  is the index for the liquidity traders.  $d_j$ s are i.i.d random variables with a Cumulative Distribution Function  $G(x) : [0, \bar{d}] \rightarrow [0, 1]$ , where  $G(x) \in C^2$ ,  $1 \leq \bar{d} < \infty$  and  $G'(x) + xG''(x) \geq 0, \forall x \in [0, 1]$ .<sup>24</sup> Again,  $d_j$ s are realized at period 0.

There are two venues for traders to trade: an exchange (the Lit market) and a dark pool. We will then consider a benchmark model where there is only one trading venue for the agents – the exchange only. By comparing our model with the benchmark model, we are

---

<sup>22</sup>In period 2 when the informed traders’ private information becomes public, they lost their information advantage. Since the informed agents are risk neutral and they only enter the market for profit, they will not actively place orders in the second period.

<sup>23</sup>We do not consider information acquisition cost because it is modeled as a sunk cost in this paper.

<sup>24</sup>This additional assumption is for the uniqueness of the equilibrium. It is satisfied by many commonly used distributions. For example, a uniform distribution.

able to study the impact of a dark pool to the public exchange, and the interaction between the two venues. We now specify the transaction rules in the two venues and the problems of each type of traders.

Finally, the distributions of  $\tilde{v}, Z^+, Z^-, \{e_i\}, \{d_j\}$  are all publicly known information.

### 1.3.1 Transaction rules in the exchange (Lit market)

A lit market is an exchange for the asset. The exchange is modeled in the spirit of [Glosten and Milgrom \(1985\)](#). Precisely, in the lit market, there is a risk neutral market maker who facilitate transactions. The objective of the market maker is to balance his or her budget. The market maker has no private information. Therefore, at period 0, the market maker announces a bid and an ask price for the risky asset, based only on public information. The announced bid and ask price will be the prices for any order submitted to the exchange in period 1, and will be committed by the market maker. Because of symmetry of  $\tilde{v}$  and the fact that the unconditional mean of  $\tilde{v}$  is zero, the midpoint of the market maker’s bid and ask is zero. Therefore, the ask price in the lit market is some  $A > 0$ , and bid price in the lit market is  $-A$ . That is, the half-spread is represented by  $A$ . We normalize  $A$  by the standard deviation of  $\tilde{v}$ ,  $\frac{A}{\sigma_v}$ , and get the normalized half-spread. For simplicity, we refer to  $A$  as the “spread,” and  $\frac{A}{\sigma_v}$  as the “normalized spread.” The spread represents a transaction cost in the lit market, because all traders, buyers or sellers, lose  $A$  dollars (per unit) to the market maker whenever they trade on the exchange. Thus, alternatively, we also refer to  $A$  as the (per unit) “exchange transaction cost” and  $\frac{A}{\sigma_v}$  as the (per unit) “normalized exchange transaction cost.”

In period 1, since informed speculators hold some information advantage about the asset, the market maker may lose money to the informed traders ex post. For example, if the realized value of the asset is  $\sigma_v$ , then the market maker loses money if he is trading against a “Buy” order. Precisely, let  $\overline{\gamma}_e, \underline{\gamma}_e$  be the respective fraction of informed speculators who place “Buy” and who place “Sell” orders on exchange, and let  $\alpha_e$  be fraction of uninformed

liquidity traders who trade in the exchange. For now we assume that they do not split orders among venues, then WLOG if the realized value of  $\tilde{v}$  is  $\sigma_v$ , the ex post payoff of the market maker is

$$\text{MM payoff} = \sigma_v[(\underline{\gamma}_e\mu - \overline{\gamma}_e\mu) + (\alpha_e Z^- - \alpha_e Z^+)] + A[\overline{\gamma}_e\mu + \underline{\gamma}_e\mu + \alpha_e Z^+ + \alpha_e Z^-],$$

where the first term is the market maker's profit on the asset. It is composed of the net gain from the informed traders,  $\underline{\gamma}_e\mu - \overline{\gamma}_e\mu$ , and the net gain from the uninformed traders,  $\alpha_e Z^- - \alpha_e Z^+$ . The second term is the gains obtained from the transaction fee (spread) per every exchange order. If the realized value of the asset is  $-\sigma_v$ , by symmetry, the market maker's payoff shall be the same as above. In this way, we also refer to  $\overline{\gamma}_e$  as the fraction of informed who "make money" (trade in the "right direction"), and  $\underline{\gamma}_e$  the fraction of informed who "lose money" (trade in the "wrong direction").

A market maker's objective is to break even on average.<sup>25</sup> That is,

$$0 = \mathbb{E} \left\{ \sigma_v[(\underline{\gamma}_e\mu - \overline{\gamma}_e\mu) + (\alpha_e Z^- - \alpha_e Z^+)] + A[\overline{\gamma}_e\mu + \underline{\gamma}_e\mu + \alpha_e Z^+ + \alpha_e Z^-] \right\}.$$

Since  $\mathbb{E}Z^+ = \mathbb{E}Z^- = \frac{1}{2}\mu_z$ , the market maker's objective becomes

$$0 = \sigma_v(-\overline{\gamma}_e + \underline{\gamma}_e)\mu + A[(\overline{\gamma}_e + \underline{\gamma}_e)\mu + \alpha_e\mu_z].$$

It implies that,

$$A = \frac{\overline{\gamma}_e - \underline{\gamma}_e}{\overline{\gamma}_e + \underline{\gamma}_e + \alpha_e \frac{\mu_z}{\mu}} \sigma_v. \quad (1.1)$$

If  $\overline{\gamma}_e \geq \underline{\gamma}_e (\geq 0)$  and  $\alpha_e \geq 0$ , then the normalized spread  $0 \leq \frac{A}{\sigma_v} \leq 1$ . In the next sections, we will show that in equilibrium,  $\overline{\gamma}_e > \underline{\gamma}_e$ . In other words, informed traders are more likely to "make money" (trade in the "right direction"). Intuitively this is true because of their information advantage. Therefore, on average, the market maker loses money to the informed.

---

<sup>25</sup>One can think of this as a result of the competition among market makers. For simplicity, we assume that there is one market maker operating.



At the end of period 1, the market maker observe the exchange volumes  $V_b, V_s$  for “Buy” volume and “Sell” volume respectively. Based on such information, the market maker then announces a closing price  $P_1 = \mathbb{E}[\tilde{v}|V_b, V_s]$ , which we consider as a proxy for the fundamental value of the asset  $\tilde{v}$ . This is because  $\mathbb{E}[\tilde{v}|P_1, V_b, V_s] = \mathbb{E}[\mathbb{E}[\tilde{v}|P_1, V_b, V_s]|P_1] = P_1$ . We are interested in how much the price  $P_1$  can aggregate information in the market (price discovery), that is, how close  $P_1$  is to the true value of the asset.

In period 2, since the realization of  $\tilde{v}$  has already been revealed, all trades will be made at the price that is equal to that realization. Thus, the payoff of the market maker in period 2 is automatically zero.

The reason we model the exchange as a market maker instead of other trading protocols such as limit order books is for the same reason as [Zhu \(2014\)](#). It is a simple but tractable way to capture the basic trade-off of dark pools. These trade-offs include lower transaction costs (lower spread) and higher execution risks, which is common to most trading protocols.

### 1.3.2 Transaction rules in the dark pool

We consider the operational costs of the dark pool as a sunk cost, and hence not considered in the model. Also, we normalize the entry fee of a dark pool as zero. The trading protocols in the dark pool we consider, include the pricing mechanism, which refers to on what price the dark pool execute orders, and the execution mechanism, which refers to how to match the buying and selling orders.<sup>26</sup>

We restrict our attention to dark pools of a particular pricing mechanism: the midpoint pricing. That is, the orders in the dark pool are crossed at the midpoint of the bid-ask in the exchange. Since the midpoint of the exchange price is 0, the transaction price in the dark pool is 0. The midpoint pricing mechanism is a reflection of an advantage trading in the dark pool: price improvement. As we point out previously, a trader has to pay a transaction

---

<sup>26</sup>As in section 1.2, we point out that not all dark pools are equal. There might be other features that investors concern. But for simplicity we focus on the two major aspects of a dark pool.

cost (the spread)  $A$  on the exchange, no matter at which direction he or she is trading. But in the dark pool, such cost is reduced to 0.

The execution mechanism we consider in this paper is a rationing mechanism. That is, orders in the shorter side are executed with probability one, whereas orders in the longer side are executed probabilistically to balance the market. For example, suppose the realization of  $\tilde{v}$  is  $\sigma_v$  (the case when  $\tilde{v} = -\sigma_v$  is symmetric). Let  $\overline{\gamma}_d, \underline{\gamma}_d$  be the fractions of informed speculators who trade in the “right direction” and “wrong direction” respectively,  $\alpha_d$  be the fraction of uninformed liquidity traders who trade in the dark pool in period 1, then the respective expected execution rates (taken with respect to  $Z^-, Z^+$ ) for trading in the “wrong direction” and in the “right direction” are: ,

$$\bar{R} = \mathbb{E} \left[ \min \left\{ 1, \frac{\overline{\gamma}_d \mu + \alpha_d Z^+}{\underline{\gamma}_d \mu + \alpha_d Z^-} \right\} \right], \quad (1.2)$$

$$\underline{R} = \mathbb{E} \left[ \min \left\{ 1, \frac{\underline{\gamma}_d \mu + \alpha_d Z^-}{\overline{\gamma}_d \mu + \alpha_d Z^+} \right\} \right]. \quad (1.3)$$

Therefore,  $\underline{R}, \bar{R} \in [0, 1]$ . The execution mechanism in the dark pool reflects a disadvantage of trading in the dark pool: execution risk. On average, one cannot expect that his or her orders be executed with probability 1 in a dark pool. In contrast, the market maker in the exchange is able to provide such certainty.

Moreover, as we will show in the next section,  $\overline{\gamma}_d > \underline{\gamma}_d$ . This means that the information asymmetry exists in the dark pool and informed traders are more likely to trade in the “right direction.” Therefore,  $\underline{R} \leq \bar{R}$ . That is to say, orders that are in the “right direction” are less likely to be executed than orders that are in the “wrong direction.” In this way, we obtain a measure of dark pool adverse selection cost in the dark pool by

$$(\bar{R} - \underline{R})\sigma_v$$

We therefore refer to  $\bar{R} - \underline{R}$  as the “Normalized dark pool adverse selection cost.”

Without loss of generality, we assume that the dark pool operates only in period 1. In period 2, since the realization of  $\tilde{v}$  is revealed, orders in the exchange are executed at that realized value. The dark pool loses its advantage and becomes redundant as nobody is willing

to trade there. Therefore, unless cancelled, orders that failed to execute in period 1 will be routed to the exchange and executed there in period 2.

### 1.3.3 The informed speculators' problem

As we point out, the informed traders only participate in period 1, when they can use their private information to their advantage. Upon the reception of a signal, the informed speculators update their beliefs about the asset fundamental value using Bayes' rule. Let  $B(s)$  be the probability that the realization is high ( $\sigma_v$ ), conditional on signal  $s$ , then by Bayes' rule,

$$B(s) = \Pr(\tilde{v} = \sigma_v | s) = \frac{\phi\left(\frac{s-\sigma_v}{\sigma_e}\right)}{\phi\left(\frac{s-\sigma_v}{\sigma_e}\right) + \phi\left(\frac{s+\sigma_v}{\sigma_e}\right)}, \quad (1.4)$$

where  $\phi(x)$  is the pdf of a standard normal distribution function.  $B(s) \in (0, 1)$  and  $B(s)$  is strictly increasing in  $s$ .

Consider an informed trader with signal  $s$ , given the exchange spread,  $A$ , and the dark pool execution probabilities,  $\bar{R}, \underline{R}$ , the expected (per unit) "Buy" and "Sell" profit in each venue, or do not trade, are respectively,

$$\begin{aligned} \text{Exchange(Lit): "Buy": } & B(s)\sigma_v - (1 - B(s))\sigma_v - A, \\ \text{"Sell": } & - [B(s)\sigma_v - (1 - B(s))\sigma_v] - A. \\ \text{Dark pool: "Buy": } & B(s)\underline{R}\sigma_v - (1 - B(s))\bar{R}\sigma_v, \\ \text{"Sell": } & - [B(s)\underline{R}\sigma_v - (1 - B(s))\bar{R}\sigma_v]. \\ \text{Not trade: } & 0. \end{aligned}$$

An informed speculator's problem is then, given his or her signal  $s$ , to choose a trading direction in  $\{\text{"Buy"}, \text{"Sell"}\}$ , the quantity in each venue  $\{\text{Exchange(Lit), Dark pool, Do not trade}\}$  to maximize his or her total expected payoff, such that total quantity does not exceed 1 unit.<sup>27</sup>

---

<sup>27</sup>The case that the informed speculator simultaneously place "Buy" and "Sell" orders in each venue is

We argue that, in equilibrium, whenever he or she decides to trade, an informed trader will place a “Buy” order if his or her signal is positive, and a “Sell” order if his or her signal is negative. Moreover, almost surely it is optimal for him to send the entire order to one of the two venues, or not trade at all. The argument is summarized in Lemma 1.

**Lemma 1. (Trading direction and non-split orders, informed)**<sup>28</sup> If an informed trader decide to trade, it is strictly optimal to “Buy” if his or her signal  $s > 0$  and to “Sell” if  $s < 0$ . Moreover, with probability one, an informed trader strictly prefers to send the entire order to one of the two venues, or do not trade at all.

The trading direction is rather straightforward since a positive signal indicates that the asset’s fundamental value is more likely to be high (i.e.,  $\sigma_v$ ), and hence more profitable in a “Buy” direction, whereas a negative signals indicates a low value (i.e.,  $-\sigma_v$ ) and hence more profitable in a “Sell” direction. And, since each trader’s signal is drawn from the same continuous distribution, and there is a continuum of informed traders, by law of large numbers, the realization of signals among them are continuously distributed. Therefore, the beliefs are distributed continuously. Since no individual has impact on the market, and the expected profit in each venue is linear in the agents’ beliefs, it is with probability 1 that, for any informed trader with signal  $s$ , one venue (or not trade) is strictly better than others.

By Lemma 1, the potential trading direction is determined once an agent receives his or her signal. Moreover, the magnitude of  $B(|s|)$  reflects the probability that this trading direction is “right.” Thus  $|s|$  can be regarded as the strength of one’s signal, and  $B(|s|)$ , can be regarded as the agent’s confidence level in their information. A strong signal (i.e., a high  $|s|$ ) represents a strong belief that the trading direction is “right,” whereas weak signals (i.e., low  $|s|$ ) represents a weak belief in the trading direction. We will show in the next

---

not considered, because the agents have no individual impact to the market. By the linearity of the per unit profit in each venue, it is never optimal to do so.

<sup>28</sup>A non-split order is strictly preferred in this model. This is a stronger result than [Zhu \(2014\)](#), in which it is only weakly optimal to not split orders for the informed because they are all indifferent between the two venues.

section, how much credit an informed trader gives to his or her private information is crucial in determining his or her strategies of venue selection.

Based on an informed traders' signal strength,  $B(|s|)$ , the payoffs of trading in each venue and no trade are, respectively,

$$\text{Exchange(Lit)} : B(|s|)\sigma_v - (1 - B(|s|))\sigma_v - A, \quad (1.5)$$

$$\text{Dark pool} : B(|s|)\underline{R}\sigma_v - (1 - B(|s|))\bar{R}\sigma_v, \quad (1.6)$$

$$\text{Not trade} : 0. \quad (1.7)$$

An informed agent's problem is then reduced to choosing one of the two venues and sending the entire 1 unit to it, with a trading direction specified in Lemma 1, or not trade at all, to yield the maximum payoff, based on his or her confidence level  $B(|s|)$ .

Finally, we define the strategy of an informed speculator who receives a signal  $s$  by a mapping

$$h_I(s) : (\infty, \infty) \rightarrow \{\text{"Buy"}, \text{"Sell"}\} \times \{\text{Exchange(Lit)}, \text{Dark pool}, \text{Not trade}\}.$$

### 1.3.4 The uninformed liquidity traders' problem

Liquidity buyer or seller types are specified by the level of their liquidity demand – the (per unit) delay cost  $d$ . If the agent fails to have his or her order executed in period 1, he or she will bear a (per unit) cost of  $\sigma_v d$ . Therefore a higher delay cost implies a higher demand for liquidity, and a higher devaluation on execution risk for the traders.

More precisely, a type  $d$  uninformed liquidity buyer's (seller's) per unit payoffs of trading in the exchange, in the dark pool, or delaying trade are, respectively,

$$\text{Exchange(Lit)} : -A, \quad (1.8)$$

$$\text{Dark pool} : -\frac{(\bar{R} - \underline{R})}{2}\sigma_v - \left(1 - \frac{\bar{R} + \underline{R}}{2}\right)\sigma_v d, \quad (1.9)$$

$$\text{Delay trade} : -\sigma_v d. \quad (1.10)$$

Similarly, we argue that in period 1, it is strictly optimal for any liquidity trader to send the entire order to one of the two venues, or delay the trade, almost surely. The argument is summarized in Lemma 2.

**Lemma 2. (No split orders, uninformed)** A liquidity trader (buyer or seller) strictly prefers to send the entire order to one of the venues, or delay trade.

The intuition of Lemma 2 is similar. Since all individuals are infinitesimal, no single trader has an impact on the market. For any liquidity trader, he or she either strictly prefers one venue over the other or is indifferent between two venues (or do not trade). Since the distribution of the delay cost  $d$  is continuous, it is with probability one that one venue (or delay) is strictly better than the other.

By Lemma 2, a type  $d$  liquidity buyer's (or seller's) problem is to maximize his or her payoff (i.e., minimize the costs), by choosing one of the venues in which trade the entire order in period 1, or to delay trade to period 2.

Moreover, we define the strategy of a type  $d$  uninformed liquidity trader by a mapping:

$$h_{U,\iota}(d) : [0, \bar{d}] \rightarrow \{\text{Exchange(Lit), Dark pool, Delay trade}\},$$

where  $\iota \in \{\text{Buyer, Seller}\}$

Finally, the trading timeline of the model is summarized in Figure 1.3. At period 0, the asset fundamental value  $\tilde{v}$ , the measure of liquidity buyers  $Z^+$  and liquidity sellers  $Z^-$ , the signal for each informed trader  $s_i$ , the per unit delay cost for each uninformed trader  $d_j$  are realized. But none of this information is public. Also, at period 0, the market maker announces the bid-ask prices with the spread  $A$ . After that, traders select venues in which place orders, which are executed according to the transaction rules in each venue. At the end of period 1, before the revelation of the value of the asset, the market maker announces a closing price of period 1, based on the volumes he observes in the exchange during that period. Then after the revelation of  $\tilde{v}$ , orders that failed to execute in period 1 are routed to the exchange (unless cancelled) and execute at the revealed value of  $\tilde{v}$ . The market is then closed.

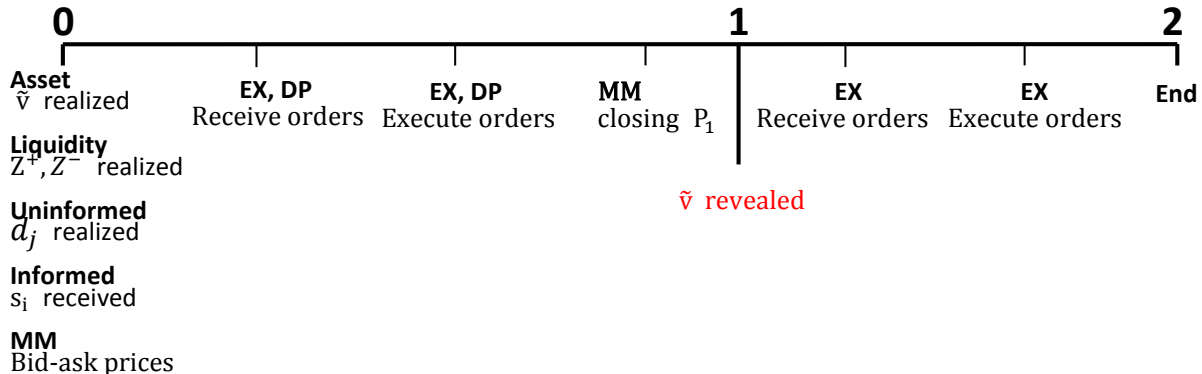


Figure 1.3: Trading Timeline

## 1.4 The Equilibrium

The model we describe in Section 1.3 assumes that both the exchange (Lit), and the dark pool are available to traders. We refer to it as the “Multi-venue” Model. We now introduce a benchmark in which there is only one venue that is operating: the exchange (Lit market). We refer to it as the “Single-venue” Model. The comparison between the two model in Section 1.5 gives us insights into the impacts of dark pools to market behaviors.

### 1.4.1 Benchmark model: without a dark pool

In the benchmark model, all else are the same except that the exchange (the lit market) is the only trading venue available for traders. Lemma 1 and Lemma 2 also hold in this model, i.e., traders do not split their orders. We use the superscription “S” to denote the “single venue” model. The equilibrium is defined as follows:

**Definition 1. (Benchmark: without a dark pool)** An equilibrium of the “Single-venue” model is a strategy for the informed speculators,  $h_I^S(s)$ , a strategy for the uninformed liquidity traders,  $h_{U,\iota}^S(d)$ ,  $\iota \in \{\text{Buyer, Seller}\}$ , an exchange spread  $A^S$ , a set of participation fractions  $\overline{\gamma}_e^S, \underline{\gamma}_e^S, \alpha_e^S$ , such that

(i) given  $A^S$ ,  $h_I^S(s)$  and  $h_{U,\iota}^S(d)$  are optimal, respectively, for an informed speculator with

signal  $s$  and for an uninformed liquidity trader with per unit delay cost  $d$ ;

(ii) given  $\overline{\gamma}_e^{\mathbb{S}}, \underline{\gamma}_e^{\mathbb{S}}$ , and  $\alpha_e^{\mathbb{S}}$ , the exchange spread  $A^{\mathbb{S}}$  makes a market maker in the exchange break-even on average;

(iii)  $\overline{\gamma}_e^{\mathbb{S}}, \underline{\gamma}_e^{\mathbb{S}}$  measure the respective fractions of informed traders who trade in the “right” and “wrong” direction in the exchange, and  $\alpha_e^{\mathbb{S}}$  measures the period 1 exchange fraction of uninformed traders.

Given  $\overline{\gamma}_e^{\mathbb{S}}, \underline{\gamma}_e^{\mathbb{S}}$ , and  $\alpha_e^{\mathbb{S}}$ , an exchange spread  $A^{\mathbb{S}}$  that makes the market maker break even on average satisfies (1.1). That is,

$$A^{\mathbb{S}} = \frac{\overline{\gamma}_e^{\mathbb{S}} - \underline{\gamma}_e^{\mathbb{S}}}{\overline{\gamma}_e^{\mathbb{S}} + \underline{\gamma}_e^{\mathbb{S}} + \alpha_e^{\mathbb{S}} \frac{\mu_z}{\mu}} \sigma_v. \quad (1.11)$$

Equation (1.11) implies that if  $\overline{\gamma}_e^{\mathbb{S}} \geq \underline{\gamma}_e^{\mathbb{S}} \geq 0$ , and  $\alpha_e^{\mathbb{S}} > 0$ , then  $\sigma_v \geq A^{\mathbb{S}} \geq 0$ . Considering an informed trader with signal “ $s$ ,” by Lemma 1, the optimal trading direction is to “Buy” if  $s \geq 0$  and to “Sell” if  $s < 0$ . Then given  $A^{\mathbb{S}}$ , The expected payoffs of trading in the exchange and do not trade are, respectively:

$$\begin{aligned} \text{Exchange(Lit)} &: B(|s|)\sigma_v - (1 - B(|s|))\sigma_v - A^{\mathbb{S}}, \\ \text{Not trade} &: 0. \end{aligned}$$

Suppose  $\sigma_v \geq A^{\mathbb{S}} \geq 0$ , then if the signal is extremely weak, i.e.,  $B(|s|) = \frac{1}{2}$ , or,  $s = 0$ , the expected payoff of trading in the exchange is strictly negative, and it is strictly optimal not to trade. In contrast, if the signal is extremely strong, i.e.,  $B(s) = 1$ , or,  $s = \pm\infty$ , the expected payoff of trading in the exchange is strictly positive, and it is strictly optimal to trade in the exchange. This is illustrated in Figure 1.4a. Therefore there must exist some cut-off point  $\widehat{s} > 0$  such that the  $\widehat{s}$  type informed traders are indifferent between trading in the exchange and do not trade. That is,

$$B(\widehat{s})\sigma_v - (1 - B(\widehat{s}))\sigma_v - A^{\mathbb{S}} = 0, \quad (1.12)$$



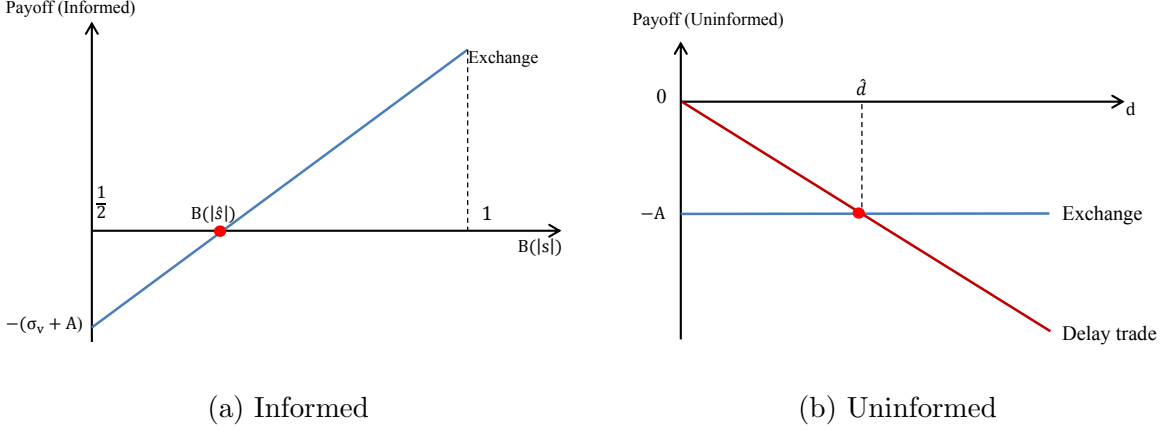


Figure 1.4: Payoffs For Traders, Single-venue

and the optimal choice for an informed trader with signal  $s$  is then

$$h_I^{\mathbb{S}}(s) = \begin{cases} (\text{"Buy"}, \text{Exchange(Lit)}) & \text{if } s \geq \hat{s}, \\ (\text{"Sell"}, \text{Exchange(Lit)}) & \text{if } s < -\hat{s}, \\ \text{Do not trade} & \text{others.} \end{cases} \quad (1.13)$$

Without loss of generality, we assume that the realization of  $\tilde{v}$  is  $\sigma_v$ . If all informed speculators follow the same optimal strategy, then the fraction of informed traders who will trade in the “right” and “wrong” directions across the population are, respectively,

$$\bar{\gamma}_e^{\mathbb{S}} = Pr(s \geq \hat{s} | \tilde{v} = \sigma_v) = Pr(s \leq -\hat{s} | \tilde{v} = -\sigma_v) = 1 - \Phi\left(\frac{\hat{s} - \sigma_v}{\sigma_e}\right), \quad (1.14)$$

$$\underline{\gamma}_e^{\mathbb{S}} = Pr(s < -\hat{s} | \tilde{v} = \sigma_v) = Pr(s > \hat{s} | \tilde{v} = -\sigma_v) = 1 - \Phi\left(\frac{\hat{s} + \sigma_v}{\sigma_e}\right). \quad (1.15)$$

(1.14),(1.15) imply that  $\bar{\gamma}_e^{\mathbb{S}} \geq \underline{\gamma}_e^{\mathbb{S}} > 0$ .

Now, we consider an uninformed liquidity trader with a (per unit) delay cost “ $d$ .” Similarly, his or he payoffs of trading in the exchange and delaying trade are, respectively:

$$\text{Exchange(Lit)} : -A^{\mathbb{S}},$$

$$\text{Delay trade} : -\sigma_v d.$$

Since  $d \in [0, \bar{d}]$  with  $\bar{d} \geq 1$ , and  $\sigma_v \geq A^{\mathbb{S}} \geq 0$ , if the liquidity trader is extremely patient, i.e.,  $d = 0$ , it is strictly optimal to delay trade to period 2. In contrast, if the liquidity trader is

extremely impatient, i.e.,  $d = \bar{d} > 1$ , it is strictly optimal to trade in the exchange. This is shown in Figure 1.4b. Therefore, there also exists a cut-off  $\hat{d}$  such that the type “ $\hat{d}$ ” liquidity trader is indifferent between trading in the exchange and delaying trade to the next period. That is,

$$-A^{\mathbb{S}} = -\sigma_v \hat{d}. \quad (1.16)$$

To combine (1.12) with (1.16), we derive that

$$\hat{d} = 2B(\hat{s}) - 1.$$

The optimal strategy for uninformed liquidity traders is then,

$$h_{U,\iota}^{\mathbb{S}}(d) = \begin{cases} (\text{“Buy” if } \iota=\text{Buyer, or “Sell” if } \iota=\text{Seller, Exchange(Lit)}) & \text{if } d \geq 2B(\hat{s}) - 1, \\ \text{Delay trade} & \text{others.} \end{cases} \quad (1.17)$$

The period 1 exchange participation rate for the uninformed traders is then

$$\alpha_e^{\mathbb{S}} = Pr(d \geq \hat{d}) = 1 - G(2B(\hat{s}) - 1), \quad (1.18)$$

and  $0 \leq \alpha_e^{\mathbb{S}} \leq 1$ .

We then find a cut-off equilibrium. Theorem 1 summarizes the existence and uniqueness.

**Theorem 1. (Existence and Uniqueness, benchmark)** For any  $\sigma_e, \sigma_v \geq 0$ , there exists an equilibrium in which traders follow cut-off strategies. That is, the respective optimal strategies for informed speculators and uninformed liquidity traders,  $h_I^{\mathbb{S}}(s)$  and  $h_{U,\iota}^{\mathbb{S}}(d)$ , are defined as (1.13) and (1.17), with the cut-off  $\hat{s}$  determined by (1.12). The exchange spread  $A^{\mathbb{S}}$  satisfies (1.11), and the participation fractions  $\overline{\gamma}_e^{\mathbb{S}}, \underline{\gamma}_e^{\mathbb{S}}, \alpha_e^{\mathbb{S}}$  are determined respectively by (1.14), (1.15), (1.18), (1.11).

Moreover, every equilibrium is a cut-off equilibrium, and the equilibrium is unique if  $\sigma_e, \sigma_v > 0$ ,  $G'(x) + xG''(x) \geq 0, \forall x \in [0, 1]$ .

The benchmark clearly gives us some insight regarding the *sorting effect* on types of traders. In equilibrium, it is strictly optimal for informed traders with relatively strong signals to trade in the exchange and for those with weak signals not to trade (avoid trading). Similarly, it is strictly optimal for uninformed liquidity traders who are relatively patient to trade in the exchange and for those who are relatively impatient to delay trade. The exchange provides functions to separate certain types of traders from others. As we will point out later, such a sorting effect is even strengthened in the presence of a dark pool.

#### 1.4.2 Multi-venue model: with a dark pool

Two trading venues are available in the multi-venue model: an exchange (Lit) and a dark pool. To differentiate from the single-venue model, we do not use the superscription  $\mathbb{S}$  in the multi-venue model. The equilibrium of the multi-venue is defined as follows:

**Definition 2. (Multi-venue, with a dark pool)** An equilibrium is a strategy for the informed speculators,  $h_I(s)$ , a strategy for and for the uninformed liquidity traders,  $h_{U,\iota}(d)$ , an exchange spread,  $A$ , two expected execution rate in the dark pool  $\bar{R}, \underline{R}$ , and a set of participation fractions  $\bar{\gamma}_e, \underline{\gamma}_e, \bar{\gamma}_d, \underline{\gamma}_d, \alpha_e, \alpha_d$ , s.t.

(i)  $h_I(s)$  is optimal for informed speculators with signal  $s$ , whereas  $h_{U,\iota}(d)$  is optimal for uninformed liquidity traders with (per unit) delay cost  $d$ , given  $A, \bar{R}$ , and  $\underline{R}$ .

(ii) the exchange spread  $A$  makes a market maker in the exchange break-even on average, given  $\bar{\gamma}_e, \underline{\gamma}_e, \bar{\gamma}_d, \underline{\gamma}_d, \alpha_e$ , and  $\alpha_d$ ;

(iii) the dark pool operates using a mid-pricing and a rationing execution mechanism.  $\underline{R}$  and  $\bar{R}$  are the respective expected execution probability for orders that are in the “right” and in the “wrong” directions;

(iv)  $\bar{\gamma}_e$  and  $\underline{\gamma}_e$  measure the respective fractions of informed traders in the exchange who trade in the “right” and “wrong” directions.  $\bar{\gamma}_d$  and  $\underline{\gamma}_d$  measure the respective fractions of informed traders in the dark pool who trade in the “right” and “wrong” directions.  $\alpha_e$  and

$\alpha_d$  measure the respective fraction of uninformed traders who trade in the exchange and in the dark pool in period 1.

Consider an informed speculator with signal “ $s$ .” Based on the strength of his or her signal  $B(|s|)$ , the payoffs of trading in the exchange, the dark pool and do not trade are summarized in (1.5), (1.6), (1.7). These payoffs are shown in Figure 1.5a.

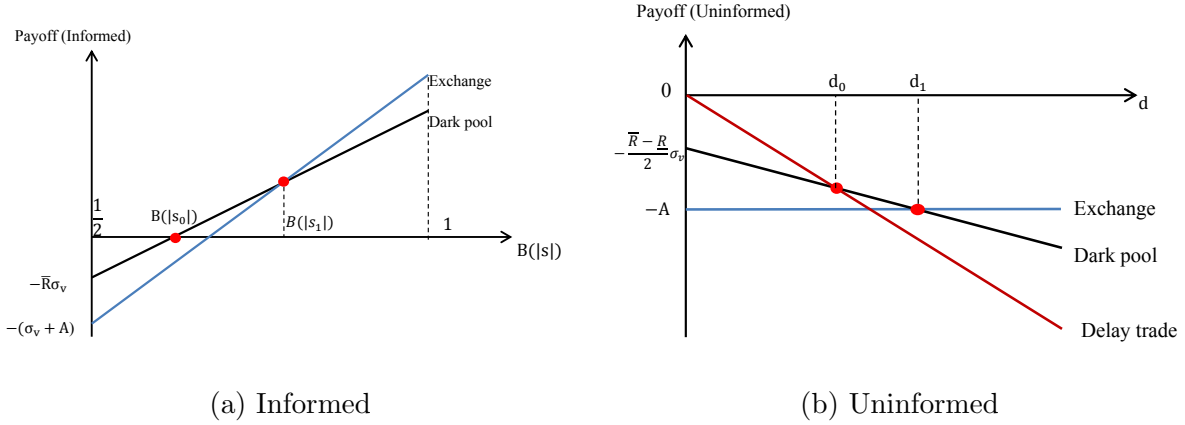


Figure 1.5: Payoffs For Traders, Multi-venue

Suppose  $1 \geq \bar{R} \geq \underline{R} > 0$  and  $\sigma_v \geq A \geq 0$ . As is shown in Figure 1.5a, if a trader receives extremely weak signals ( $s = 0$  for example), it is never profitable to trade, since trading is costly. However, whenever an informed trader decides to trade, he faces a trade-off between execution certainty in the exchange and price improvement in the dark pool. When  $|s|$  is low, the need for price improvement overwhelms the need for execution, in which case, trading in a dark pool is better. But as the signals becomes stronger, the need for execution grows faster than the need for price improvement. This can be observed from the fact that the exchange payoff has a higher slope with respect to  $B(|s|)$  than the dark pool payoff. Therefore, when  $s$  is extremely high, it is possible that the two intersect. Suppose an informed trader with signal  $s_0 > 0$  is indifferent between trading in a dark pool and not trade, an informed with signal  $s_1 > 0$  is indifferent between trading in a dark pool and in the exchange, then by

(1.5), (1.6), and (1.7),  $s_0, s_1$  satisfies:

$$B(s_0)(\bar{R} + \underline{R}) = \bar{R} \quad (1.19)$$

$$B(s_1) [(1 - \bar{R}) + (1 - \underline{R})] \sigma_v = A + (1 - \bar{R})\sigma_v. \quad (1.20)$$

At this point, the existence and relationship of  $s_0$  and  $s_1$  is not established yet. For now, we suppose that  $(s_0, s_1)$  exists and  $s_0 < s_1 < +\infty$  (we will prove that this is true in every equilibrium), the optimal strategy for an informed trader with signal  $s$  is then

$$h_I^S(s) = \begin{cases} (\text{“Buy”}, \text{Exchange(Lit)}) & \text{if } s \geq s_1, \\ (\text{“Buy”}, \text{Dark pool}) & \text{if } s_0 \leq s < s_1, \\ (\text{“Sell”}, \text{Dark pool}) & \text{if } -s_1 \leq s < -s_0, \\ (\text{“Sell”}, \text{Exchange(Lit)}) & \text{if } s < -s_1, \\ \text{Do not trade} & \text{others.} \end{cases} \quad (1.21)$$

This is illustrated in Figure 1.6. That is, it is strictly optimal that informed traders with strong signals to trade in the exchange, informed traders with modest signals to trade in the dark pool, and informed traders with weak signals to not trade.

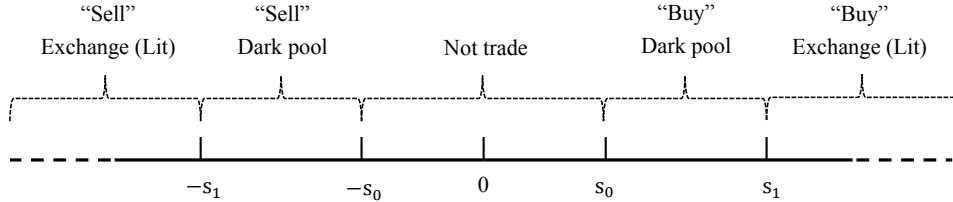


Figure 1.6: Strategy of Informed Traders

If all informed traders follow such strategy, the exchange fraction of informed who trade in the “right” and “wrong” directions are, respectively,

$$\bar{\gamma}_e = \Pr(s \geq s_1 | \tilde{v} = \sigma_v) = \Pr(s \leq -s_1 | \tilde{v} = -\sigma_v) = 1 - \Phi\left(\frac{s_1 - \sigma_v}{\sigma_e}\right), \quad (1.22)$$

$$\underline{\gamma}_e = \Pr(s < -s_1 | \tilde{v} = \sigma_v) = \Pr(s > s_1 | \tilde{v} = -\sigma_v) = 1 - \Phi\left(\frac{s_1 + \sigma_v}{\sigma_e}\right). \quad (1.23)$$

And the dark pool fraction of informed who trade in the “right” and “wrong” directions are, respectively,

$$\bar{\gamma}_d = \Pr(s_0 \leq s < s_1 | \tilde{v} = \sigma_v) = \Pr(-s_1 \leq s < -s_0 | \tilde{v} = -\sigma_v) = \Phi\left(\frac{s_1 - \sigma_v}{\sigma_e}\right) - \Phi\left(\frac{s_0 - \sigma_v}{\sigma_e}\right), \quad (1.24)$$

$$\underline{\gamma}_d = \Pr(-s_1 \leq s < -s_0 | \tilde{v} = \sigma_v) = \Pr(s_0 \leq s < s_1 | \tilde{v} = -\sigma_v) = \Phi\left(\frac{s_1 + \sigma_v}{\sigma_e}\right) - \Phi\left(\frac{s_0 + \sigma_v}{\sigma_e}\right). \quad (1.25)$$

Similarly, for the uninformed, the payoffs of trading in the exchange, in the dark pool, and delaying trade are respectively given in (1.8), (1.9), and (1.10), as illustrated in Figure 1.5b. Again, a liquidity trader with extremely low liquidity demands would find it optimal to delay trade. However, if he decides to trade in period 1, only those with extremely high liquidity demands (i.e., extremely impatient) are willing to trade, for the similar reason as the informed traders. Let  $d_0$  and  $d_1$  respectively represent the type of liquidity traders who are indifferent between delaying trade and trading in a dark pool, and the type who are indifferent between delaying trading in a dark pool and in the exchange, then by (1.8), (1.9), and (1.10) we have

$$\begin{aligned} -\frac{(\bar{R} - \underline{R})}{2}\sigma_v - \left(1 - \frac{\bar{R} + \underline{R}}{2}\right)\sigma_v d_0 &= -\sigma_v d_0, \\ -\frac{(\bar{R} - \underline{R})}{2}\sigma_v - \left(1 - \frac{\bar{R} + \underline{R}}{2}\right)\sigma_v d_1 &= -A. \end{aligned}$$

Combine this with (1.19) and (1.20), we derive that

$$\begin{aligned} d_0 &= 2B(s_0) - 1, \\ d_1 &= 2B(s_1) - 1. \end{aligned}$$

By a similar argument, the optimal strategy for an uninformed trader is also a cut-off

strategy:

$$h_{U,t}^S(d) = \begin{cases} \text{("Buy" if } \iota=\text{Buyer, or "Sell" if } \iota=\text{Seller, Exchange(Lit))} & \text{if } d \geq 2B(s_1) - 1, \\ \text{("Buy" if } \iota=\text{Buyer, or "Sell" if } \iota=\text{Seller, Dark pool)} & \text{if } 2B(s_0) - 1 \\ & \leq d < 2B(s_1) - 1, \\ \text{Delay trade} & \text{otherwise.} \end{cases} \quad (1.26)$$

This is described in Figure 1.7. The exchange fraction,  $\alpha_e$ , and dark pool fraction,  $\alpha_d$ , of uninformed liquidity traders, are, respectively,

$$\alpha_e = 1 - G(2B(s_1) - 1), \quad (1.27)$$

$$\alpha_d = G(2B(s_1) - 1) - G(2B(s_0) - 1). \quad (1.28)$$

The fact that the traders the cut-off of uninformed traders' are functions of the cut-off

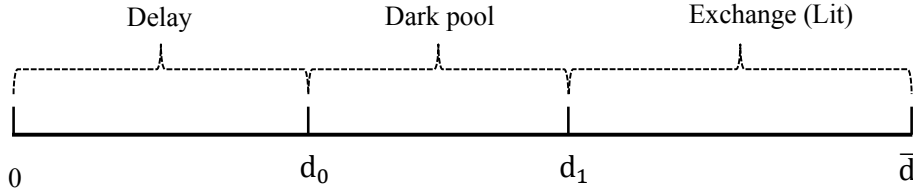


Figure 1.7: Strategy of Uninformed Traders

of informed traders' reveals that, in equilibrium, uninformed and informed traders always move together. It cannot happen that uninformed traders move collectively from one venue to another, forming a new equilibrium without influencing the behavior of the informed traders. This is in contrast with [Zhu \(2014\)](#).

Given  $\bar{\gamma}_e, \underline{\gamma}_e, \alpha_e$ , the exchange spread  $A$  captured in (1.1) makes the market maker break even. Also, given  $\bar{\gamma}_d, \underline{\gamma}_d, \alpha_d$ , and given the distribution of  $Z^+$  and  $Z^-$ , the expected execution rates in the dark pool,  $\bar{R}$  and  $\underline{R}$ , are respectively determined by (1.2) and (1.3).

If such  $s_0, s_1$  exists, we find a cut-off equilibrium. But the existence is not obvious. The difficulty arises from two aspects. First, we cannot simply apply a fixed point theorem

because it cannot distinguish the trivial equilibrium from others: a trivial equilibrium is one in which all trades happen in one venue, for example, the exchange. Second, the equilibrium involves a very complicated equation system and these equations are non-linear and are not likely to exhibit monotonicity. Nevertheless, we are able to show in Theorem 2 that the equilibrium exists. Moreover, all equilibria are cut-off equilibria, and all equilibria are non-trivial.

**Theorem 2. (Equilibrium with DP)** For any  $\sigma_v, \sigma_e > 0$ , an equilibrium exists in which traders follow cut-off strategies. That is, the respective optimal strategies for informed and uninformed traders,  $h_I(s)$  and  $h_v^U(d)$ , are defined as in (1.21) and (1.26), with cut-offs  $(s_0, s_1)$  solving (1.19) and (1.20),  $0 < s_0 < s_1$ . Moreover, every equilibrium is a cut-off equilibrium, and every equilibrium is non trivial (meaning positive participation for both informed and uninformed traders in both venues).

The exchange spread,  $A$ , the expected execution rates,  $\bar{R}, \underline{R}$ , are determined, respectively, by (1.1), (1.2), and (1.3). The set of participation fractions,  $\{\bar{\gamma}_e, \underline{\gamma}_e, \bar{\gamma}_d, \underline{\gamma}_d, \alpha_e, \alpha_d\}$  are determined by (1.22), (1.23), (1.24), (1.25), (1.27), and (1.28).

**Corollary 1. (Liquidity begets liquidity)**  $\alpha_d > 0$  if and only if  $\bar{\gamma}_d - \underline{\gamma}_d > 0$ .

The equilibrium characterized in Theorem 2 is distinctive to Zhu (2014) in the following aspects. First, in contrast with Zhu (2014), in equilibrium in our model, there is a *sorting effect* of market fragmentation, and uninformed and informed traders always move together. It is respectively optimal for informed traders with strong signals, modest signals, and weak signals to trade in the exchange, in the dark pool, and do not trade, whereas it is respectively optimal for uninformed traders with high, modest, and low degrees of impatience to trade in the exchange, in the dark pool, and delay trade. In Zhu (2014), however, such a sorting effect is absent for informed traders. In his model, informed traders are homogeneous and indifferently between venues. This may cause the instability of its prediction. For example, uninformed traders can collectively move from the dark pool to the exchange. This movement may increase the adverse selection cost in the dark pool so much so that they will stay in



the exchange, and price discovery is strictly decreased. These equilibria are not discussed in [Zhu \(2014\)](#). Our prediction is more robust in the sense that traders always move together and this sorting effect exists in every equilibrium. The same predictions on price discovery hold in every equilibrium.

Second, unlike [Zhu \(2014\)](#), in which there exists some cases where informed traders do not participate in the dark pool, we predict that all equilibrium is non-trivial. That is, informed and uninformed participate in both venues in all equilibria, as captured in [Corollary 1](#). This casts light on the dynamics of liquidity creation in a dark pool: informed and uninformed traders tend to arrive the dark pool in a clustered fashion, which in turn attract more liquidity to the dark pool, as documented in the literature.<sup>29</sup> One explanation why [Zhu \(2014\)](#) predicts a different result is that he assumes exact signals for traders. As we have pointed out, traders with strong signals tend to prefer an exchange. It is possible that, in some cases, they all crowd in the exchange and are absent in the dark pool. But again, this might be subject to an unstable status. In our model, this will not happen because with a noisy information structure, the dark pool will always be attractive to some informed traders. This is related to the following aspect.

The equilibrium described in [Theorem 2](#) also disclose one important function of dark pools: a function that cannot be captured without a noisy information structure. That is, dark pools help to mitigate traders' information risk, i.e., the loss attributable to bad information. Dark pools take a role as a "buffer zone" for informed traders – a gambling place for those who are less well-informed to trade. This adds value to the trade-off of dark pools, and shall clearly not be neglected. When information becomes noisier, more informed traders will find dark pools more valuable places to trade. Also, if traders become risk-averse, the importance of this function for dark pools will increase to a great extent.

**Corollary 2.** Given any  $\sigma_e, \sigma_v > 0$ ,  $s_1 > \hat{s}$ , and in correspondent,  $d_1 > \hat{d}$ .

**Corollary 3. (Adverse selection)**  $\forall \sigma_e, \sigma_v > 0$ ,  $0 < \underline{\gamma}_e < \overline{\gamma}_e$ ,  $0 < \underline{\gamma}_d < \overline{\gamma}_d$ , and  $\bar{R} - \underline{R} > 0$ .

---

<sup>29</sup>[Sarkar et al. \(2009\)](#) provide a more detailed description of such process.

**Proof.** If  $\sigma \in (0, +\infty)$ , by Theorem 2,  $0 < s_0 < s_1$ . Therefore by definition of (1.4), and (1.1), (1.2), (1.3), (1.22), (1.23), (1.24), (1.25), (1.28), it must be that  $\frac{A}{\sigma_v}, \alpha_d, \alpha_e \in (0, 1)$  and  $0 < \underline{\gamma}_e < \overline{\gamma}_e < 1, 0 < \underline{\gamma}_d < \overline{\gamma}_d < 1$ . Therefore  $0 < \underline{R} < \overline{R} < 1$ .  $\square$

Corollary 2 states that dark pools strictly decrease traders' participation in the exchange. Corollary 3 states that there exists adverse selection in both the exchange and the dark pool. Market makers lose money to informed traders on average.

## 1.5 Dark Pool Trading and Information Structure

In this section, we restrict our attention to the following questions. These questions will be discussed in Section 1.5.1, 1.5.2, and 1.5.3, respectively.

- (i) How do each venue's market participation and information asymmetry level vary with the information structure, i.e., " $\sigma_e$ "?
- (ii) How does adding a dark pool impact market participation and information asymmetry?
- (iii) How does adding a dark pool impact price discovery, and what are the determinants?

### 1.5.1 Information Precision and Market Characteristics

To recall, dark pools are of important value for informed traders who are less well-informed because they mitigate their informational risks. When information becomes more precise, such need decreases, and a migration of traders from one venue to another shall be observed. In this section, we study how the traders' participation and information asymmetry level in each venue vary with the informational structure. The results are shown in Proposition 1 and Proposition 2. The numerical example is in Figure 1.9. We use  $\sigma_e$  to capture the information precision for informed traders. A lower  $\sigma_e$  corresponds with lower noises, hence a higher precision in their signals.

**Proposition 1. (Exchange spread, Dark pool adverse selection costs)** If  $\sigma_e$  is large,

then both the exchange spreads and the dark pool adverse selection costs increase in information precision. That is, as  $\sigma_e$  decreases,

(Without DP):  $\frac{A^S}{\sigma_v}$  strictly increases;

(With DP): Similarly,  $\frac{A}{\sigma_v}$  increases,  $\widehat{R} - \underline{R}$  increases,

**Proposition 2. (Participation rates)** Suppose  $\sigma_e$  is large. Then for informed traders, as information precision increases, both the exchange and the dark pool participation increase. In contrast, for uninformed traders, as information precision increases, the exchange participation decreases while the dark pool participation increase. And total uninformed participation decreases. That is, as  $\sigma_e$  decreases,

(Without DP):  $\overline{\gamma_e^S} - \underline{\gamma_e^S}$  strictly increases, and  $\alpha_e^S$  strictly decreases;

(With DP): Similarly,  $\overline{\gamma_e} - \underline{\gamma_e}$ ,  $\overline{\gamma_d} - \underline{\gamma_d}$  increases,  $\alpha_e$  decreases,  $\alpha_d$  increases, and  $\alpha_e + \alpha_d$  decreases.<sup>30</sup>

**Remark 1.** when  $\sigma_e$  is large, as in Proposition 1 and Proposition 2, dark pool participation for informed traders and dark pool adverse selection cost INCREASES with information precision. When  $\sigma_e$  is small, however, they may DECREASE with information precision. We have not been able to obtain comparative statics when  $\sigma_e$  is small, but we show this inverted U-shape in the numerical example in Figure 1.9.<sup>31</sup> While we provide an explanation in the context, the explicit proof is of future work.

In the exchange, when signals become more precise, both the informed exchange participation,  $\overline{\gamma_e} - \underline{\gamma_e}$ , and exchange spread,  $A$ , increase, whereas the uninformed exchange

---

<sup>30</sup> $\overline{\gamma_e} - \underline{\gamma_e}$  and  $\overline{\gamma_d} - \underline{\gamma_d}$  capture the “meaningful” participation of informed trades, in the sense that they are the fractions of informed trades that trade in the “right” direction net the fractions that trade in the “wrong” direction.

<sup>31</sup>In all our plots, we use a set of parameters in which  $\mu_z = 60$ ,  $\mu = 30$ ,  $Z^+$ ,  $Z^-$  has Gamma distributions with mean 30 and variance 30 and  $G(d) = \frac{d}{3}$  for  $\bar{d} \in [0, 3]$ .

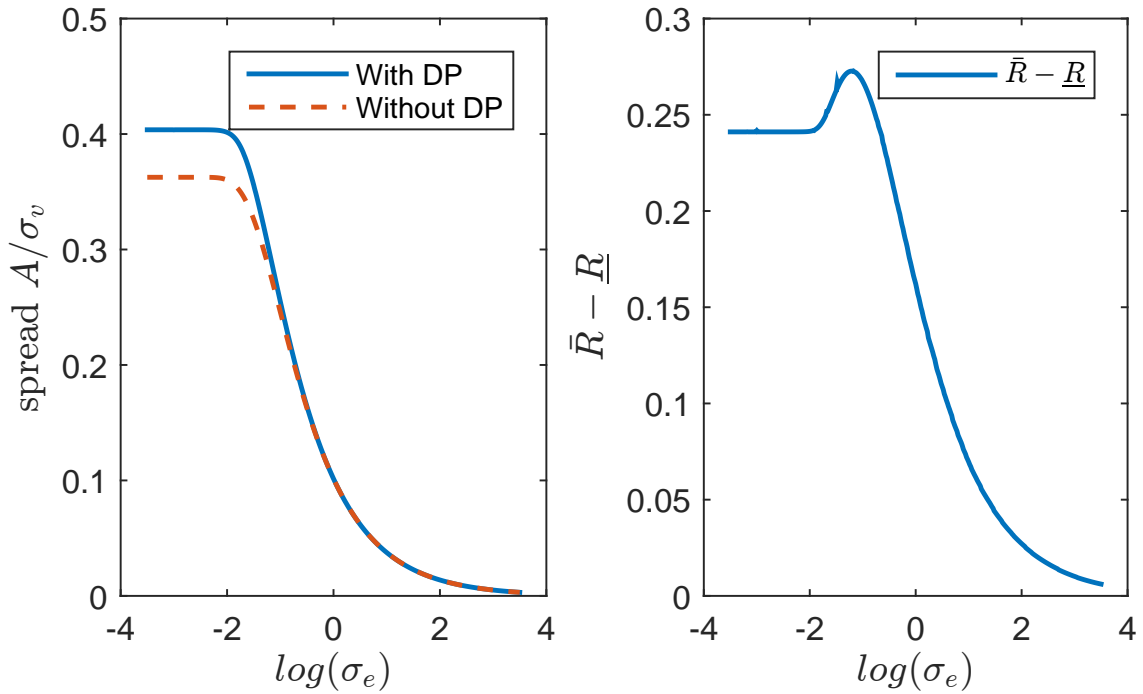


Figure 1.8: **Transaction Costs.** The left-hand figure shows the normalized spreads on the exchange and how they vary with  $\log(\sigma_e)$ ; the right-hand figure shows the adverse selection cost in the dark pool and how it vary with  $\log(\sigma_e)$ . In both figures,  $\log(\sigma_v) = 0$ .

participation,  $\alpha_e$ , decreases. The intuition is as follows. In equilibrium the informed traders are sorted by the strengths of their signals. when there is an increment in their information precision, the overall strengths of their signals are increased. Therefore, some informed traders migrate from “do not trade” to “trade in the dark pool” and from “trade in the dark pool” to “trade in the exchange.” This will cause a strict increase of information asymmetry level in the exchange, and hence an increase of the exchange spread. Consequently, some liquidity traders migrate from “trade in the exchange” to “trade in the dark pool,” which decreases the uninformed participation in the exchange.

In the dark pool, the dark pool informed participation,  $\bar{\gamma}_d - \underline{\gamma}_d$ , and the dark pool adverse selection,  $\widehat{R} - \underline{R}$ , exhibit an inverted U-shape with information precision. The intuition for the inverted U-shape is as follows. A change in the information precision changes the distribution of the signals’ strengths. When the information precision level is low (i.e.,  $\sigma_e$

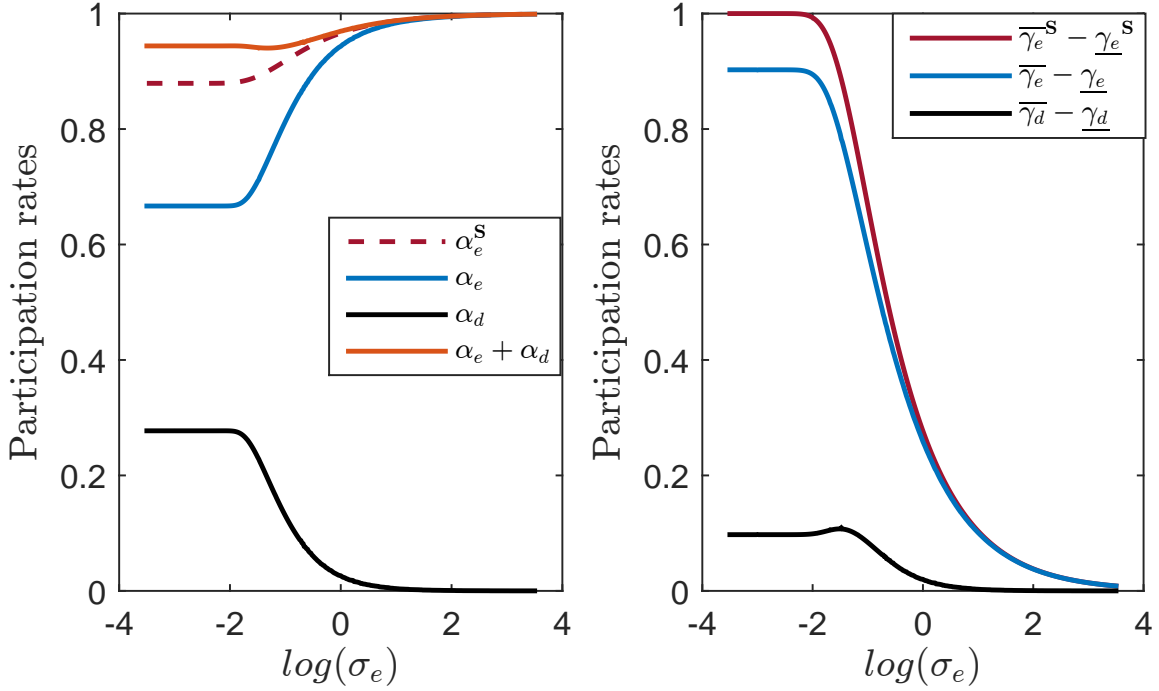


Figure 1.9: **Participation Rates.** The left figure plots the expected participation rates of the uninformed and how they vary with  $\log(\sigma_e)$ . The right one shows the participation rates for the informed traders how they vary with  $\log(\sigma_e)$ . In both plots,  $\log(\sigma_v) = 0$ ,  $\mu_z = 60$ ,  $\mu = 30$ .

is high), as the precision grows, signals become more concentrated in the relative “modest” group, and more informed traders migrate from “do not trade” to the dark pool. Overall, this induces a greater proportion of informed participation in the dark pool, and the dark pool adverse selection increases. In contrast, when the information precision level is high (i.e.,  $\sigma_e$  is low), as precision grows, signals become more concentrated in the relative “strong” group. Thus, more informed traders migrate from the dark pool to the exchange, leaving a lower proportion of informed trades in the dark pool, and the dark pool adverse selection decreases.

An interesting comparison with [Zhu \(2014\)](#) is that, although [Zhu \(2014\)](#) does not consider the information structure, he discusses the comparative statics of market behaviors as a function of  $\sigma_v$ .  $\sigma_v$  and  $\sigma_e$  are comparable in the sense that, all else equal, informed traders’ information advantage increases in both information precision (i.e., as  $\sigma_e$  decreases), and the

asset value uncertainty (i.e., as  $\sigma_v$  increases, see a more detailed discussion in Section 1.5.3).

We highlight two major differences between our predictions and those of [Zhu \(2014\)](#). First, our model predicts that traders’ participation exhibits a smooth variation cross-sectionally (i.e., when  $\sigma_v$  grows), whereas there is a discontinuity in that of [Zhu \(2014\)](#). In [Zhu \(2014\)](#), in equilibrium informed traders don’t trade in dark pools for some assets unless the asset’s value uncertainty is high (i.e.,  $\sigma_v$  is high). In contrast, we predict that both informed and liquidity traders trade in dark pools in a clustering fashion, regardless of  $\sigma_v$ . This is a more realistic prediction. If there are some assets for which dark pools only attract liquidity traders, one would expect a persistent gap between the average size of dark pools and the average size of lit markets. Yet, this is not true as we observe in [Figure 1.2](#). This, again, emphasizes that dark pools function as informational risk mitigators and that they are always lucrative for traders, informed or uninformed.

Second, [Zhu \(2014\)](#) predicts that informed traders’ participation in dark pool always squeezes out liquidity traders (i.e.,  $\alpha_d$  decreases as informed trades grow in the dark pool), whereas we predict that the two can grow simultaneously, especially when informed traders’ information is relatively imprecise. The explanation is that the informed trading intensity in the dark pool is always high in [Zhu \(2014\)](#) because traders have exact information. But in our model, the intensity is neutralized to some extent because some speculators trade in the “wrong” direction.

## 1.5.2 Dark Pool Impacts on Market Characteristics

In this section, we study how the market responds when a dark pool is added alongside an exchange. Precisely, we compare the equilibrium traders’ participation and exchange spread between the two models: the “Single-venue” model and the “Multi-venue” model. In the comparison, we fixed the information structure (i.e.,  $\sigma_e$ ). The result is shown in [Proposition 3](#). This result coincides with [Zhu \(2014\)](#), except that the effect on the exchange spread  $A$  is uncertain when information is imprecise (i.e.,  $\sigma_e$  high).

**Proposition 3.** Given any  $\sigma_v, \sigma_e > 0$ , then adding a dark pool alongside an exchange a) (**Participation**): decreases the participation in the exchange for both informed and uninformed traders, but increases the total market participation, and b) (**Exchange spread**): widens the spread on the exchange, if information precision is high ( $\sigma_e$  is small).

That is, suppose  $\frac{\mu_z}{\mu} \geq \frac{R}{1-R} \frac{1}{1-G(\hat{k})}$  where  $R = \mathbb{E} \left[ \min \left\{ 1, \frac{R^+}{R^-} \right\} \right]$ , and  $\hat{k}$  is uniquely determined by  $\hat{k} = \frac{1}{1+[1-G(\hat{k})\frac{\mu_z}{\mu}]}$  then

(i)  $(\overline{\gamma}_e^{\mathbb{S}} - \underline{\gamma}_e^{\mathbb{S}}) \geq (\overline{\gamma}_e - \underline{\gamma}_e)$ ,  $\alpha_e^{\mathbb{S}} \geq \alpha_e$ , and if  $\sigma_e$  is sufficiently small or large,  $\alpha_e^{\mathbb{S}} \leq \alpha_e + \alpha_d$ .

And,

(ii)  $\frac{A^{\mathbb{S}}}{\sigma_v} \leq \frac{A}{\sigma_v}$  if  $\sigma_e$  is small.

**Remark 2.** When information precision is high ( $\sigma_e$  is low), as in Proposition 3, we proved that  $\frac{A^{\mathbb{S}}}{\sigma_v} \leq \frac{A}{\sigma_v}$  (i.e., adding a dark pool WIDENS the exchange spread). When information precision is low ( $\sigma_e$  is high), however, it is possible that  $\frac{A^{\mathbb{S}}}{\sigma_v} > \frac{A}{\sigma_v}$  (i.e., adding a dark pool NARROWS the exchange spread).<sup>32</sup> While we discuss this briefly in Appendix 1.8.6, the explicit analysis is of future work.

Proposition 3 states that adding a dark pool will decrease informed and uninformed traders' exchange participation but increase the total participation. Thus, dark pools create additional liquidity. This, again, is explained by the migration of traders. Because adding a dark pool enlarges the opportunity sets for both informed and uninformed traders, there will be migrations of both types of traders from both "Not trade" and "trade in the exchange" to "trade in the dark pool." Therefore, the dark pool attracts not only additional liquidity but also part of the liquidity from the exchange. As a consequence, the exchange participation decreases, but the total participation of traders increases. This is captured in figure 1.9 in which  $\alpha_e \leq \alpha_e^{\mathbb{S}} \leq \alpha_e + \alpha_d$ .

---

<sup>32</sup>When  $\sigma_e$  is large, it is either  $\frac{A^{\mathbb{S}}}{\sigma_v} < \frac{A}{\sigma_v}$  when  $\sigma_e$  is large, or undetermined (in which, as  $\sigma_e \rightarrow +\infty$ ,  $\frac{A^{\mathbb{S}}}{\sigma_v}$  equals  $\frac{A}{\sigma_v}$ , and their first order derivatives with respect to  $\sigma_e$  are equal.)

The impact of a dark pool to the exchange spread, however, is not straightforward. The spread depends on the level of information asymmetry in the exchange, which in turn depends on the intensity of informed and uninformed trades. As we have pointed out, the addition of a dark pool induces an outflow of both informed and uninformed traders. The resulting proportion of the two in the exchange depends on which overwhelms the other. When the informed traders have high information precision (i.e., low  $\sigma_e$ ), a large fraction of them strictly prefers to stay in the exchange, and only a small fraction will migrate to the dark pool, compared with the migration of uninformed traders. As a result, the exchange information asymmetry strictly increases and exchange spread, “ $\frac{A}{\sigma_v}$ ,” is enlarged. When the informed traders have low precision in their information (i.e.,  $\sigma_e$  is high), however, there is a large fraction of the informed who prefer to migrate to the “buffer zone,” the dark pool, and the relative proportion of informed traders in the exchange decreases. As a result, the exchange spread may or may not decrease, depending on how intense the migration is.<sup>33</sup>

### 1.5.3 Dark Pool Impacts on Price Discovery

Price discovery is measured by the informativeness of  $P_1$ . At the end of period 1, the market maker observes the period 1 exchange order flows  $V_b, V_s$ , which respectively represents the “buy” volume and the “sell” volume and announces a closing price  $P_1 = \mathbb{E}[\tilde{v}|V_b, V_s]$ .  $P_1$  is perceived as a proxy for the fundamental value of the asset. This is so because  $\mathbb{E}[\tilde{v}|P_1, V_b, V_s] = \mathbb{E}[\mathbb{E}[\tilde{v}|P_1, V_b, V_s]|P_1] = P_1$ . We are interested in how informative  $P_1$  is, that is, how close  $P_1$  is to the true value of the asset.

We consider similar measures as suggested by [Zhu \(2014\)](#). Without loss of generality, we assume that the true value  $\tilde{v} = +\sigma_v$ . Let the likelihood ratio

$$r = \log \frac{\Pr(\tilde{v} = +\sigma_v|V_b, V_s)}{\Pr(\tilde{v} = -\sigma_v|V_b, V_s)} = \log \frac{\phi_z(Z^+ = \frac{1}{\alpha_e}[V_b - \bar{\gamma}_e\mu]) \cdot \phi_z(Z^- = \frac{1}{\alpha_e}[V_s - \underline{\gamma}_e\mu])}{\phi_z(Z^- = \frac{1}{\alpha_e}[V_b - \bar{\gamma}_e\mu]) \cdot \phi_z(Z^+ = \frac{1}{\alpha_e}[V_s - \underline{\gamma}_e\mu])}.$$

---

<sup>33</sup>Note that  $\frac{A}{\sigma_v}$  depends on both  $\frac{\bar{\gamma}_e - \underline{\gamma}_e}{\bar{\gamma}_e + \underline{\gamma}_e}$  and  $\frac{\bar{\gamma}_e - \underline{\gamma}_e}{\alpha_e}$ , when  $\sigma_e$  is large,  $\frac{\bar{\gamma}_e - \underline{\gamma}_e}{\alpha_e}$  decreases when adding a dark pool but not necessarily  $\frac{\bar{\gamma}_e - \underline{\gamma}_e}{\bar{\gamma}_e + \underline{\gamma}_e}$ . The overall effect on  $\frac{A}{\sigma_v}$  is uncertain.



And

$$\begin{aligned} P_1 &= \sigma_v \Pr(\tilde{v} = +\sigma_v | V_b, V_s) + (-\sigma_v) \Pr(\tilde{v} = -\sigma_v | V_b, V_s) \\ &= \frac{\Pr(\tilde{v} = +\sigma_v | V_b, V_s) - \Pr(\tilde{v} = -\sigma_v | V_b, V_s)}{\Pr(\tilde{v} = +\sigma_v | V_b, V_s) + \Pr(\tilde{v} = -\sigma_v | V_b, V_s)} \sigma_v \end{aligned}$$

Therefore

$$P_1 = \frac{e^r - 1}{e^r + 1} \sigma_v.$$

Clearly, if  $r$  is higher,  $P_1$  is closer to the true value  $\sigma_v$ . If  $r = +\infty$ , then  $P_1 = \sigma_v$ , in which case  $P_1$  is completely informative. Therefore,  $r$  can be considered as a measure of the informativeness.

Another measure of informativeness that we consider is the scaled root-mean-squared error (RMSE), in which

$$\text{RMSE} = \frac{[\mathbb{E}[(\tilde{v} - P_1)^2 | \tilde{v} = \sigma_v]]^{.5}}{\sigma_v} = \mathbb{E} \left[ \frac{4}{(e^r + 1)^2} | \tilde{v} = \sigma_v \right].$$

It is scaled by  $\sigma_v$ . Since  $r \in (0, 1)$ , the scaled pricing error (RMSE) is between 0 and 1. If RMSE is higher, there are more pricing errors, and there is less price discovery.

Since  $V_b, V_s$  are random variables,  $r$  is also a random variable. When  $\mu_z, \sigma_z^2$  are large enough, we can approximate the density of  $\phi_z(\cdot)$  by a normal distribution  $\mathcal{N}(.5\mu_z, .5\sigma_z^2)$ .<sup>34</sup> Substituting the density functions, we get an approximate  $r$  by

$$r^{Approx} = \frac{2(\overline{\gamma_e} - \underline{\gamma_e})\mu}{\alpha_e^2 \sigma_z^2} (V_b - V_s).$$

Given that  $\tilde{v} = \sigma_v$ , Since  $V_b - V_s$  has a distribution of  $\mathcal{N}((\overline{\gamma_e} - \underline{\gamma_e})\mu, \alpha_e^2 \sigma_z^2)$ , so  $r^{Approx}$  has a distribution of

$$\mathcal{N} \left( 2\mathcal{I}(\overline{\gamma_e}, \underline{\gamma_e}, \alpha_e)^2, 4\mathcal{I}(\overline{\gamma_e}, \underline{\gamma_e}, \alpha_e)^2 \right),$$

where

$$\mathcal{I}(\overline{\gamma_e}, \underline{\gamma_e}, \alpha_e) = \frac{(\overline{\gamma_e} - \underline{\gamma_e})\mu}{\alpha_e \sigma_z}.$$

---

<sup>34</sup>We use the same approximation as in [Zhu \(2014\)](#), in which it shows that when  $\mu_z$  and  $\sigma_z^2$  are large enough,  $Z^+$  is approximately normal.

Thus, the magnitude of  $\mathcal{I}(\underline{\gamma}_e, \underline{\gamma}_e, \alpha_e)$  can be taken as a measure of the price discovery in the exchange. To be consistent with definitions of [Zhu \(2014\)](#), we also refer to it as “signal-to-noise” ratio. We consider two measures of price discovery: the signal-to-noise ratio  $\mathcal{I}(\underline{\gamma}_e, \underline{\gamma}_e, \alpha_e)$  and the scaled RMSE under the normal approximation.

By the same argument as [Zhu \(2014\)](#), under the normal approximation, a higher signal-to-noise ratio  $\mathcal{I}(\underline{\gamma}_e, \underline{\gamma}_e, \alpha_e)$  always corresponds to a lower scaled RMSE. That is, they are in nature the same measure. Therefore, we only plot the “signal-to-noise” in our numerical example in [Figure 1.10](#).

We introduce a measure for the informed traders: that is, the measure of their “information advantage”:

$$\boldsymbol{\sigma} = \frac{\sigma_v}{\sigma_e}.$$

An informed speculator’s “information advantage” is defined as the asset’s fundamental uncertainty  $\sigma_v$  times the precision of the signals  $\frac{1}{\sigma_e}$ . Clearly, a higher  $\sigma_v$  reflects a high level of undisclosed information, therefore, a higher profitability of the informed speculators. Also, a lower  $\sigma_e$  means a higher precision of the private information, and hence a higher informational profit. [Proposition 4](#) summarizes the price discovery as a function of  $\boldsymbol{\sigma}$  and the impact of a dark pool to price discovery.

**Proposition 4.** Price discovery (i.e. the informativeness of  $P_1$ ) in the exchange is an increasing function of informed traders’ “information advantage” ( $\boldsymbol{\sigma}$ ). And, there exists a threshold,  $\bar{\boldsymbol{\sigma}} > 0$ , such that, a) when  $\boldsymbol{\sigma} < \bar{\boldsymbol{\sigma}}$ , adding a dark pool impairs price discovery, and b) when  $\boldsymbol{\sigma}$  is large, adding a dark pool enhances price discovery.

That is, suppose  $\hat{k} \leq \frac{\mu_z}{\mu} < +\infty$ , where  $\hat{k}$  is uniquely determined by  $\hat{k} = \frac{1}{1+[1-G(\hat{k})]\frac{\mu_z}{\mu}}$ , then  $\mathcal{I}(\underline{\gamma}_e, \underline{\gamma}_e, \alpha_e)$ ,  $\mathcal{I}(\underline{\gamma}_e^{\mathbb{S}}, \underline{\gamma}_e^{\mathbb{S}}, \alpha_e^{\mathbb{S}})$  increase in  $\boldsymbol{\sigma}$  and RMSE,  $\text{RMSE}^{\mathbb{S}}$  decrease in  $\boldsymbol{\sigma}$ , when  $\sigma_e > 0$  is large enough, and  $\exists \bar{\boldsymbol{\sigma}} > 0$  such that

(i) if  $\boldsymbol{\sigma} \in (0, \bar{\boldsymbol{\sigma}})$ , adding a dark pool will strictly decrease the informativeness of the price in exchange, that is,  $\mathcal{I}(\underline{\gamma}_e, \underline{\gamma}_e, \alpha_e) < \mathcal{I}(\underline{\gamma}_e^{\mathbb{S}}, \underline{\gamma}_e^{\mathbb{S}}, \alpha_e^{\mathbb{S}})$ , and  $\text{RMSE} > \text{RMSE}^{\mathbb{S}}$

(ii) if  $\sigma$  is sufficiently large, adding a dark pool will increase the informativeness of the price in exchange, that is,  $\mathcal{I}(\overline{\gamma}_e, \underline{\gamma}_e, \alpha_e) \geq \mathcal{I}(\overline{\gamma}_e^{\mathbb{S}}, \underline{\gamma}_e^{\mathbb{S}}, \alpha_e^{\mathbb{S}})$ , and  $\text{RMSE} \leq \text{RMSE}^{\mathbb{S}}$

When a dark pool is added alongside an exchange, the impact on price discovery is depending on the resulting ratio of informed traders and uninformed traders in the exchange. As we have discussed in Section 1.5.2, when a dark pool is introduced to the market, it induces migrations of both informed traders and liquidity traders from the exchange to the dark pool. When  $\sigma$  is high, on average, informed traders have high profitability, a high proportion of the informed would rather stay in the exchange, and only a small proportion migrate from the exchange to the dark pool, compared with the liquidity traders. Therefore adding a dark pool increases the “signal-to-noise” ratio and improves the informativeness of  $P_1$  in the exchange. When  $\sigma$  is low, however, on average the informed have low profitability so that a higher proportion would rather migrate from the exchange to trade in the “buffer zone,” the dark pool, compared with the liquidity traders. This leaves a lower proportion of informed traders in the exchange. The “signal-to-noise” ratio decreases and price discovery declines.

In Figure 1.10, the right plots “signal-to-noise” ratio as a function of  $\sigma = \frac{\sigma_v}{\sigma_e}$ . It increases with  $\sigma$ , indicating that informed traders’ trading intensity grows with higher “informational advantage,” and hence price discovery increases. Introducing a dark pool alongside an exchange decreases price discovery when  $\sigma$  is low (i.e.,  $\sigma_v$  is low or  $\sigma_e$  is high), and increases when  $\sigma$  is high (i.e.,  $\sigma_v$  is high or  $\sigma_e$  is low). The left further illustrates the dark pool impact on price discovery in a 2-dimensional context (i.e.,  $\sigma_v$  and  $\sigma_e$ ). .

The results highlight an important effect dark pools have on price discovery – an *amplification Effect* That is, dark pools enhances price discovery when it is high, whereas dark pools impairs price discovery when it is low. An economy needs to be prudent in introducing dark pools to its equity market, especially when the economy has a poor information environment (low quality in information disclosure, poor legal systems and enforcement, etc.) We provide a more detailed discussion in Section 1.6.2.

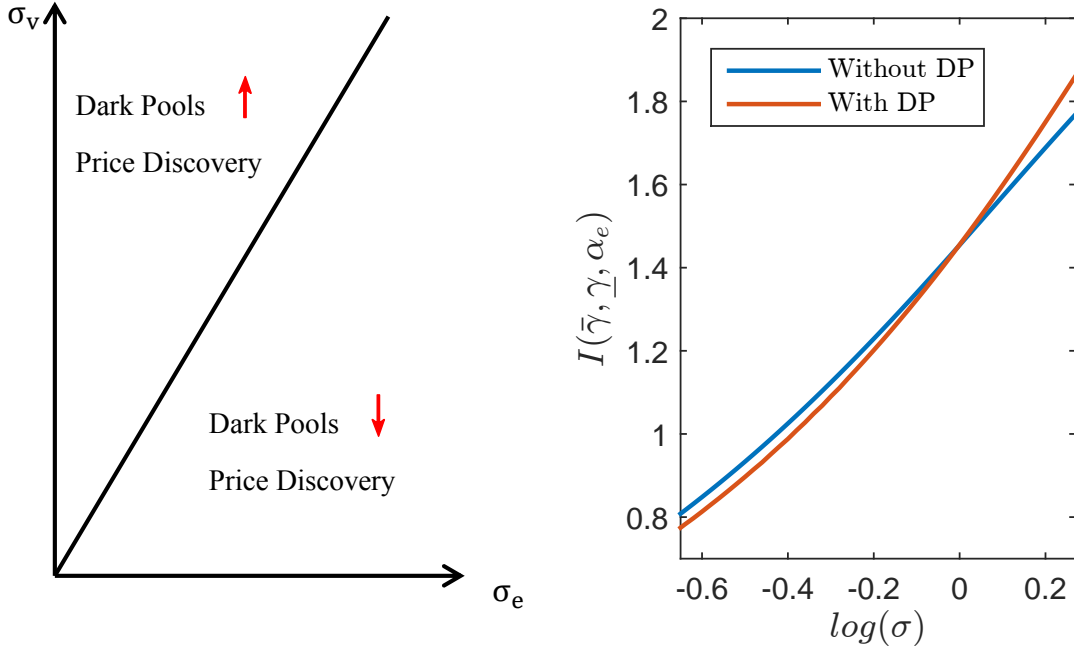


Figure 1.10: **Price Discovery.** The left plots the dark pool impact on price discovery with 2-dimension:  $\sigma_v$  and  $\sigma_e$ . The right plots the “Signal-to-noise” ratio  $I(\bar{\gamma}, \underline{\gamma}_e, \alpha_e)$  as a function of  $\sigma = \frac{\sigma_v}{\sigma_e}$ .

This result is in contrast with [Zhu \(2014\)](#), in which adding a dark pool strictly increases the price discovery. According to our analysis, the important reason [Zhu \(2014\)](#) predicts a strict increase is due to the fact that it assumes an extreme case where signals for informed traders are perfect (i.e.,  $\sigma_e \rightarrow 0$  in our model). As we have pointed out, when information is in high precision (i.e.,  $\sigma_e$  is low), the majority of the informed traders prefer the exchange, where dark pools will attract relatively less fraction informed traders from the exchange, compared with the liquidity traders, and leave a higher ratio of informed-to-uninformed traders in the exchange, hence improve price discovery. Thus, [Zhu \(2014\)](#) is consistent with our prediction. In reality, however, [Zhu \(2014\)](#)’s prediction may not hold because the information structure is much richer and exhibits significant cross-sectional difference (we will discuss this in [Section 1.6.2](#)). Policies and measures should be tailored to this issue in a different information environment.

[Zhu \(2014\)](#) also depicts a scenario when uninformed liquidity trader types are discrete. It

shows that in this case, to a large degree, price discovery will be harmed by the introduction of dark pools because uninformed traders of discrete types are more likely to get “stuck” in their original venues while some informed traders flow from the exchange to dark pools and decrease price discovery. Our prediction corresponds to this scenario. In our prediction, the discrete type and “stickiness” of uninformed traders will further increase the chance that price discovery be harmed.

**Determinants of the impact.** From the perspective of a regulator, when introducing dark pools, an important issue is what fraction of the assets will be harmed in their price discovery. In order to answer that question, one should examine the determinants and the overall impact dark pools have on price discovery.

We consider a proxy which we refer to as the “likelihood that dark pools harm price discovery.”

$$\bar{\sigma}_v = \sup_{x>0} \{x | \forall \sigma_v \in (0, x), \mathcal{I}(\bar{\gamma}_e^{\mathbb{S}}, \underline{\gamma}_e^{\mathbb{S}}, \alpha_e^{\mathbb{S}}) > \mathcal{I}(\bar{\gamma}_e, \underline{\gamma}_e, \alpha_e)\}.$$

By Proposition 4, such  $\bar{\sigma}_v$  must exist. A higher  $\bar{\sigma}_v$  reflects a higher fraction of assets whose price discovery will be harmed by adding a dark pool.

We consider two determinants. The first is the precision of traders’ private information, the inverse of  $\sigma_e$ . Proposition 4 indicates that the likelihood dark pools harm price decreases with precision level. Another determinant we consider is the relative measure of informed traders,  $\frac{\mu}{\mu_z}$ . The effects of the two on  $\bar{\sigma}_v$  is summarized in Proposition 5.

**Proposition 5.** the likelihood that price discovery will be harmed by dark pool trading ( $\bar{\sigma}_v$ ) decreases in information precision,  $\sigma_e$ , and increases in the relative measure of informed traders,  $\frac{\mu}{\mu_z}$ .

That is, Suppose  $\hat{k} \leq \frac{\mu_z}{\mu} < +\infty$ , where  $\hat{k}$  is uniquely determined by  $\hat{k} = \frac{1}{1+[1-G(\hat{k})\frac{\mu_z}{\mu}]}$ , then

- (i)  $\bar{\sigma}_v$  increases in  $\sigma_e$ . As  $\sigma_e \rightarrow 0^+$ ,  $\bar{\sigma}_v \rightarrow 0$ , and as  $\sigma_e \rightarrow +\infty$ ,  $\bar{\sigma}_v \rightarrow +\infty$ . And,
- (ii) for any sequence of  $\{(\frac{\mu}{\mu_z})\}$ , there exists a subsequence  $\{(\frac{\mu}{\mu_z})_n\}$  such that as  $(\frac{\mu}{\mu_z})_n$  increases,

$\bar{\sigma}_v$  increases, also, as  $(\frac{\mu}{\mu_z})_n \rightarrow 0^+$ ,  $\bar{\sigma}_v \rightarrow 0$ .<sup>35</sup>

The numerical example is given in Figure 1.11. Proposition 5 states that dark pools are beneficial for price discovery in an economy with a good information environment (i.e., high information precision and low size of informed traders), whereas they are bad for price discovery in an economy with a poor information environment (i.e., low information precision and high size of informed traders). Proposition 5 gives regulators insights into how to improve the economy and informativeness of prices. Policies and measures can be taken to enhance the market performance. Also, it points out important considerations for countries that are going to allow dark pools and provides them a benchmark to measure market quality. More details are in Section 1.6.2.

## 1.6 Discussion: Empirical and Regulatory

In this section, we provide a discussion about empirical implications and policy suggestions. The discussion is intended to provide insight into seemingly contradictory results in the empirical literature, as well as give exploration of channels for future research and regulatory concern. In these analyses, the economic force we consider is the variation of the information structure, more precisely, the informed traders' "information advantage,"  $\sigma = \frac{\sigma_v}{\sigma_e}$ , or the information imprecision,  $\sigma_e$ , if  $\sigma_v$  is fixed. We refer to "good information environment" by more precise information and less informed traders. Although we attempt to attribute the difference of the findings to the different information structures, we preserve a conservative interpretation in these predictions. In general, our model suggests that dark pool activity and its impacts display significant cross-sectional variation and thus should be evaluated differently in various economic environments.

---

<sup>35</sup>We cannot directly show that  $\bar{\sigma}_v$  increases in  $\frac{\mu}{\mu_z}$ , but we are able to show a upper bound of  $\bar{\sigma}_v$  that is increasing in  $\frac{\mu}{\mu_z}$ .

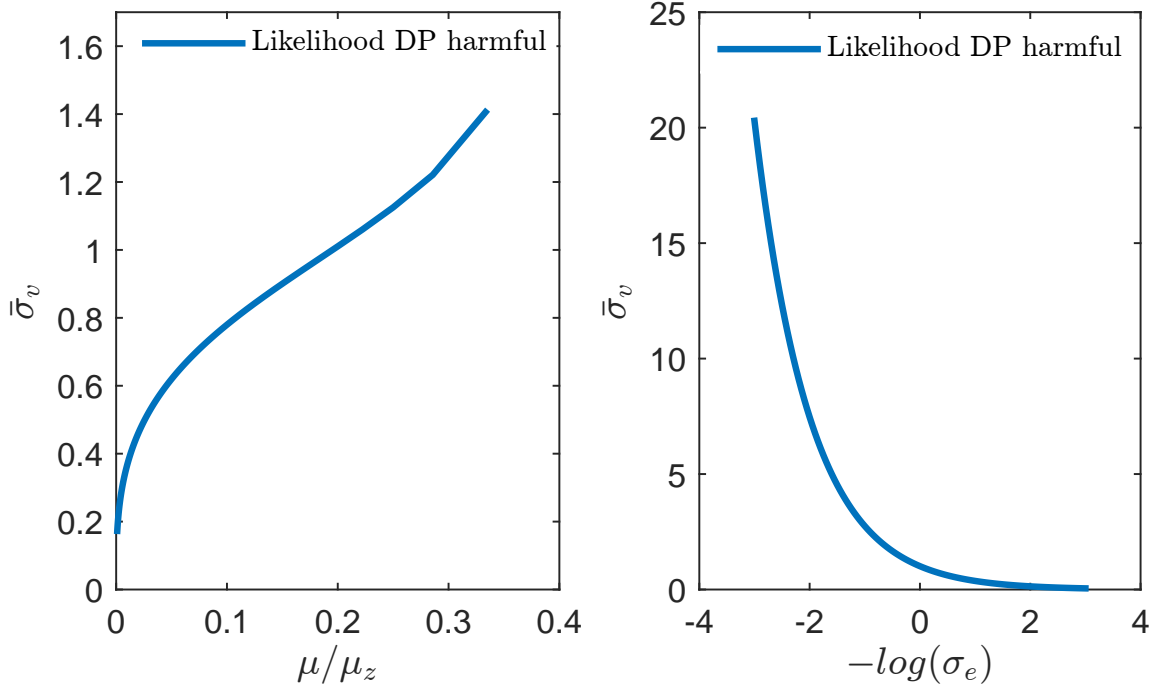


Figure 1.11: **The Likelihood DPs Harm Price Discovery**  $\bar{\sigma}_v$ . The left-hand figure plots the threshold  $\bar{\sigma}_v$  as a function of the relative size of informed traders,  $\frac{\mu}{\mu_z}$ . The right-hand shows the threshold  $\bar{\sigma}_v$  as a function of the information precision,  $-\log(\sigma_e)$ . On the left,  $\log(\sigma_e) = 0$ . On the right,  $\frac{\mu}{\mu_z} = .2$ .

### 1.6.1 A Summary of Testable Empirical Predictions

**1. Dark pool execution probability.** We predict that dark pool non-execution probability increases with information precision (i.e.  $1 - \frac{\bar{R}+R}{2}$  increases as  $\sigma_e$  decreases). Also, an asset's exchange spread increases with its dark pool non-execution probability (i.e.  $\frac{A}{\sigma_v}$  increases in  $1 - \frac{\bar{R}+R}{2}$ ).

This prediction suggest that the trade-off of dark pools is higher in an economy with a good information environment. The trade-off is documented in many empirical papers. For example, [Gresse \(2006\)](#), [Conrad et al. \(2003\)](#), [Næs and Ødegaard \(2006\)](#) , and [Ye \(2010\)](#) study crossing networks in the US and conclude that dark pools, in comparison with exchanges, have lower trading costs (within spread price) but higher non-execution probability. [He and Lepone \(2014\)](#) studied Australia's Centre Point dark pool and found that the dark pool execution probability increases with dark pool activity. In contrast, [Kwan](#)

et al. (2015) find that the dark pool execution probability increases in the trading friction in exchanges: the minimal price improvement.

The change of execution probability can be explained as follows: the execution depends on two factors: traders' total participation and dark pool information asymmetry level. The former irons the difference between the two sides in the pool and increases the execution rate, whereas the latter does the opposite. In the numerical example in Figure 1.12, we show that, without pricing frictions in the exchange, the expected dark pool execution rate decreases as the information becomes more precise ( $\sigma_e$  decreases).

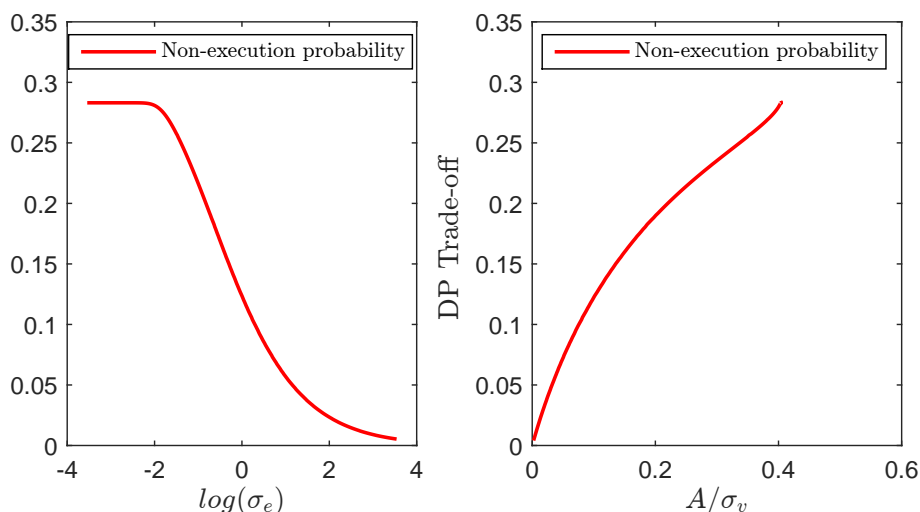


Figure 1.12: **Execution Probability and Trade-off of A Dark Pool.** The left figure plots the non-execution probability as a function of  $\log(\sigma_e)$ . The right-hand figure plots the non-execution probability as a function of the exchange spread  $A/\sigma_v$ . In both plots,  $\log(\sigma_v) = 0$ .

**2. Dark pool usage and market characteristics.** All else equal, in an economy/industry/asset that has a high information precision, dark pool market share decreases with information precision and with exchange spread, whereas in an economy/industry/asset that has low information precision, dark pool market share increases with information precision and with exchange spread. More precisely,

- (1) dark pool market share has an inverted U-shape relationship with the information precision,



(2) dark pool market share has an inverted U-shape relationship with the exchange spread.

The prediction follows from Proposition 1, Proposition 2, and Remark 1. To measure dark pool usage, we analyze the volumes in each venue. Since informed traders have no profit to trade in period 2 due to the disclosure of information, they cancel their unexecuted orders and leave the market in period 2. The remaining orders continue to execute in the exchange. The expected trading volume in the dark pool, in the exchange, and total consolidated volume are, respectively:

$$V_d = (\bar{R}\underline{\gamma}_d + \underline{R}\bar{\gamma}_d)\mu + \frac{\bar{R} + \underline{R}}{2}\alpha_d\mu_z, \quad (1.29)$$

$$V_e = (\underline{\gamma}_e + \bar{\gamma}_e)\mu + \alpha_e\mu_z + (1 - \alpha_e - \alpha_d)\mu_z + \left(1 - \frac{\bar{R} + \underline{R}}{2}\right)\alpha_d\mu_z, \quad (1.30)$$

$$V = V_d + V_e. \quad (1.31)$$

We distinguish the components of dark volumes by “Dark uninformed volumes” and “Dark informed volumes” respectively as:

$$V_d^U = \left(1 - \frac{\bar{R} + \underline{R}}{2}\right)\alpha_d\mu_z, \quad (1.32)$$

$$V_d^I = V_d - V_d^U. \quad (1.33)$$

Figure 1.13 illustrates equilibrium behavior of dark pool market share and dark pool “informed volume” share. Though this prediction coincides with Zhu (2014), our model emphasizes the role of the trader’s information structure.

This prediction is consistent with Ray (2010) and Preece (2012), which report a similar inverted U-shape between dark pool usage and exchange spread. Other empirical studies have reported contradictory results using different datasets. For studies using different US datasets, Hatheway et al. (2013) and Weaver (2014) find a positive association while O’Hara and Ye (2011) and Ready (2014) find a negative association between dark trading and exchange spread. ASIC (2013) and Comerton-Forde and Putniņš (2015) study Australian dark trading and find a positive relationship. Degryse et al. (2015) find a positive relationship for European dark fragmentation. Our model suggests that such a relationship varies cross-sectionally, depending on the specific information structure. The cross-sectional difference

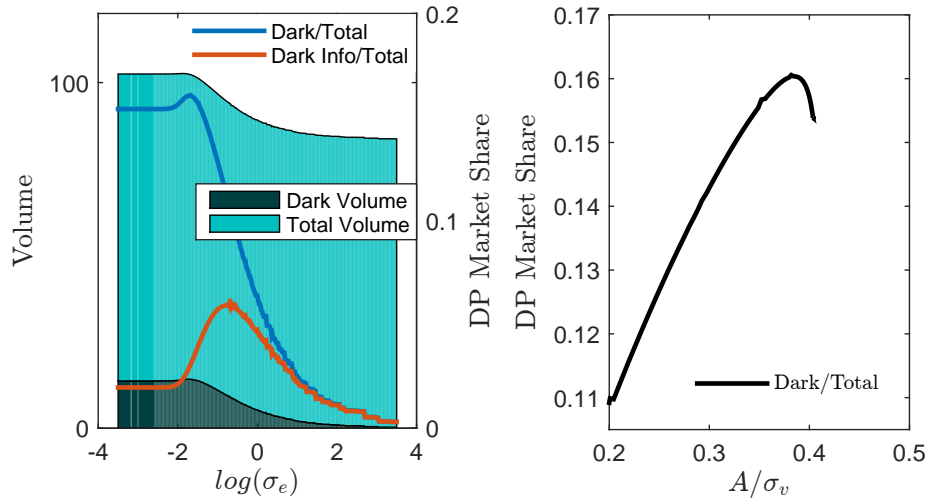


Figure 1.13: **Dark Volumes and Market Share.** The left figure plots the dark pool volume, total volume and dark pool market share as a function of  $\log(\sigma_e)$ . The right-hand figure shows the dark pool market share as a function of the exchange spread  $A/\sigma_v$ . In both plots  $\log(\sigma_v) = 0$ .

is reflected in [Nimalendran and Ray \(2014\)](#), [Buti et al. \(2011b\)](#), and [O’Hara and Ye \(2011\)](#). More cross-sectional studies that specify the characteristics of firms and countries are needed.

**3. Information content of dark pool trades.** In an economy with high information precision, the information content of dark pool trades decreases with information precision and with exchange spread. By contrast, in an economy low information precision, the information content of dark pool trades increases with information precision and with exchange spread. More precisely,

- (1) the information content of dark pool trades has an inverted U-shape relationship with the information precision,
- (2) the information content of dark pools trades has an inverted U-shape relationship with the exchange spread.

The prediction follows from [Proposition 1](#) and [Proposition 2](#). We use two measures for the dark pool information content. The first measure is the DP *Predictive Fraction* – the fraction of dark pool volumes that are traded in the “right direction” (i.e., fraction of volumes that

predict the movement of prices). The higher the fraction is, the higher is the information content of a dark pool. In this model, the Predictive Fraction is defined as

$$\text{DP Predictive Fraction} = \frac{\underline{R}(\bar{\gamma}_d\mu + .5\alpha_d\mu_z)}{V_d}.$$

Another measure we consider is the normalized adverse selection costs,  $\bar{R} - \underline{R}$ . The inverted U-shape of the two measures with the exchange spread is depicted in Figure 1.14. There

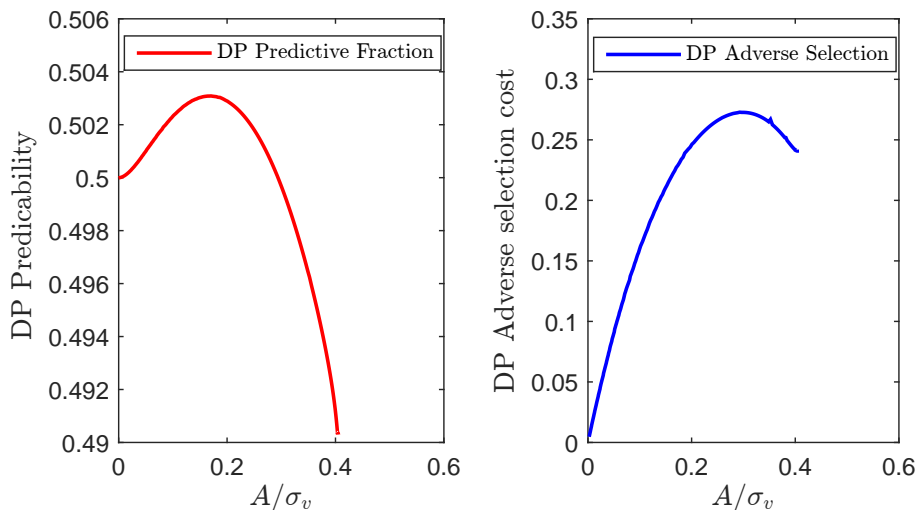


Figure 1.14: **Predictability of Dark Pool Trades.** The left figure plots the dark pool “Predictive Fraction” as a function of spread  $A/\sigma_v$ . The right-hand plots the dark pool adverse selection costs as a function of spread  $A/\sigma_v$ .

are relatively few studies that look at this issue. [Peretti and Tapiero \(2014\)](#) conclude that dark pool trades can significantly forecast price movements. [Nimalendran and Ray \(2014\)](#) study trades in a large crossing network and find that the information content in a dark pool is positively associated with the exchange spread.

But as we point out, under different information environments, the dark pool informational content may differ cross-sectionally. Further study in this area is needed.

#### 4. Impacts of adding a dark pool alongside an exchange. We predict that

(i) **Liquidity externality.** Adding a dark pool alongside an exchange decreases the exchange volume but increases the overall volume.

(ii) **Price discovery and exchange spread.** Dark pools have an *amplification effect* on price discovery. That is, the introduction of dark pools enhances price discovery when price discovery is high, and impairs price discovery when price discovery is low. Moreover, the improvement of price discovery is associated with a wider exchange spread, whereas the deterioration of price discovery can be associated with a wider or narrower spread.

Prediction 4 follows directly from Proposition 3, Proposition 4 and Remark 2.<sup>36</sup> Few studies focus on the direct impact of introducing dark pool trading. For example, [Hendershott and Mendelson \(2000\)](#) and [Hendershott and Jones \(2005\)](#) found that there was a reduction in price efficiency after Island ECN stopped displaying its limit order book. [Chlistalla and Lutat \(2011\)](#) finds that the entrance of Chi-X, a dark pool in the US, decreased spread.

Other research studies the relationship between price discovery and dark pool trading intensity within the fragmented framework. O’Hara and Ye (2011) and Jiang et al (2012) find a positive association between price discovery and dark pool trading, whereas [Hatheway et al. \(2013\)](#) and [Weaver \(2014\)](#) find the opposite. [Comerton-Forde and Putniņš \(2015\)](#) conduct a more comprehensive cross-sectional study and show that, when the fraction of non-block trades in dark pools is high (above 10%, suggesting that dark pools contain a high fraction of informational orders), then dark trading harms price discovery, whereas if dark pools contain less informational orders, dark trading improves price discovery. [Comerton-Forde and Putniņš \(2015\)](#)’s prediction is consistent with ours in the sense that we predict an inverted U-shape for the relation of dark pool information content and the information precision. More research is still needed on the important question of the effect of dark pool activity on price efficiency for different types of stocks in the cross-section.

---

<sup>36</sup>Remark 2 points out, when private information is imprecise, it is possible that price discovery is decreased while spread increases. If this is the case a dark pool can be strictly detrimental to the exchange.

## 1.6.2 Regulatory Considerations

Price discovery is the essential economic function of an exchange. As Alan and Schwartz (2013) point out, price discovery, as a public good, gives investors confidence and promotes the interests of listed entities and the broader community through an efficient secondary market for capital. More precisely, an exchange-produced price benefits a broad spectrum of market participants who use it for marking to market, derivatives valuation, mutual-fund cash flow estimation, estates, and dark pool pricing. Thus, the efficiency of how prices are discovered becomes a serious matter in measuring market quality. In the periods of time when markets are deeply fragmented by dark pool trading, it is of extreme importance for regulators to be wary of the impacts dark pools have on price discovery.

This paper shows that there is no certain claim in the issue whether dark pools harm price discovery. The information structure is essential to determining the impact of dark pools on price discovery. Dark pools enhance price discovery when the information environment is good and they impair price discovery when the information environment is poor. The information structure determines the level of price discovery and dark pools' impact on it. The use of dark pools should be case sensitive. In this section, we provide a brief discussion about information structure and regulatory suggestions.

**1. Information environment and its determinants.** There are two factors to consider for an information environment: the precision of (private) information and the number of informed traders. The notion of a better information environment includes a higher precision in traders' (private) information and fewer informed traders.<sup>37</sup> The information environment in equity markets largely depends on the following aspects: (1) The nature of the security. Researchers have found that firms with greater growth volatility (such as high R&D firms, young firms), smaller size, or fewer analyst followers have lower informational precision in traders' predictions (Li et al. 2012, Maffett 2012, Lang and Lundholm 1993, Ba-

---

<sup>37</sup>Private information is not necessarily insider information. A big fraction of it is information that is publicly available but hard to collect, transmit, and process by the majority of the public.

ginski and Hassell 1997). (2) Traders' ability to obtain and process information. Information precision is positively associated with the overall level of traders' experience. Generally, a more matured financial market, with more competition, more innovation in trading technologies, and years of trade has a higher level of trader ability for information acquisition (Louis et al. 2014, Clement 1999, Chen et al. 2005). (3) The opacity of the economy's information environment in the macro-setting. This includes the strength of legal institutions and law enforcement against insider trading, the functionality of the public disclosure system, and the availability and efficiency of media transmission. Generally, public disclosure and media channels can enhance the precision of informed traders' forecasts<sup>38</sup> and stronger legal systems can significantly reduce the number of insiders.

**2. What should regulators do?** Regulators should be cautious in controlling dark pool trading in order to not harm price discovery. To do that, regulators should examine the following aspects. First, dark pool trading should be regulated to a level that distinguishes firm characteristics. As we have pointed out, traders generally possess low precision for high R&D firms, young firms, small firms, and less-analyzed firms. Introducing dark pools to these firms might cause a decrease in price discovery. Second, a monitoring system measuring the public's ability to process information should be built, and dark pool trading should be under dynamic revision. Third, countries should continue to improve their judicial system to prevent insider trading, and, at the same time, take measures to improve the efficiency of public disclosure, including accounting information enhancement and financial reporting regulations. Countries should also ensure there are more effective financial media channels. In general, regulators should improve countries' information environment.

**3, Dark pools in emerging markets?** Based on current evaluations of the information environment in several emerging markets, a great proportion of emerging markets are governed by poor legal systems and have limited implemental power against insider trading

---

<sup>38</sup>Although there is a debate regarding the association between public and private information, researchers generally find that public disclosures may be processed into private information by informed investors, and there is a positive correlation between the precisions of public and private information. See Botosan et al. (2004) and Kim and Verrecchia (1994).

and poor quality of information disclosure. These countries should be extremely cautious in dark pool trading. For example, [Bhattacharya and Daouk \(2002\)](#) found that the enforcement of insider trading laws in 81 emerging markets is significantly low compared with developed countries. [Wang and Wu \(2011\)](#) and [Yu and Lu \(2009\)](#) document poor quality of financial information in mainland China, and they show that up to a quarter of listed firms in mainland China explicitly admitted to the poor quality of their financial information by restating their previous financial reports. [Tang et al. \(2013\)](#) finds that a poor corporate governance system interacts with abnormal insider trading to aggravate the information environment in Taiwan. [Budsaratragoon et al. \(2012\)](#) tests insider trading regulations in Thailand and find that severe informational asymmetry, lax enforcement and poor pricing efficiency are endemic. As we point out, dark pools have an amplification effect on price discovery, so introducing dark pool trading in those countries may aggravate the situation.

## 1.7 Conclusion

This paper studies the impact of dark pools on price discovery in a noisy information framework. We find that the addition of a dark pool to the traditional exchange has an *amplification effect* on price discovery, i.e., it enhances price discovery when the information has high precision and impairs price discovery when the information has low precision. The results reconcile the conflicting empirical findings in current literature and suggest new channels of research to disentangle the relationship between dark pool trading and market quality.

We highlight the dark pool's function as an informational risk mitigator. In equilibrium, information is sorted by market fragmentation. That is, traders with strong signals trade in the exchange, traders with modest signals trade in the dark pool, and traders with weak signals do not trade. When information precision is low, a large proportion of informed traders with modest signals crowd in the dark pool to reduce their information risk. Adding a dark pool, thus, shifts a higher fraction of informed traders from the exchange, compared with liquidity traders, leaving a lower informed-to-uninformed ratio in the exchange and thus

decreasing price discovery. In contrast, when information precision is high, a large proportion of informed traders with strong signals crowd in the exchange. Adding a dark pool shifts only a small fraction of informed traders from the exchange, compared with liquidity traders, increasing the informed-to-uninformed ratio in the exchange and increasing price discovery.

There are several observations that complement the overall effects on market quality. First, when information precision is low, the market can experience a deterioration of price discovery along with a widened exchange spread. In this case, dark pools are strictly detrimental to the exchange. Second, dark pools always attract informed traders and liquidity traders in a clustered fashion. We should observe both informed and uninformed traders in all trading venues. Third, the ability of dark trades to predict price movement has an inverted U-shape with exchange spread. Therefore, assets with modest exchange liquidity have a high information content in their dark pool trades.

There are aspects regulators should be aware of. First, dark pools and their impacts have significant variance cross-sectionally. The information structure of different assets, industries, and countries differs in nature. The use of dark pools is thus case sensitive. Second, in a deeply fragmented market, policies that help improve the information environment are needed to enhance price discovery. These measures include, among others, enhancing public disclosure by improving accounting and reporting regulations, strengthening legal systems, and implementing laws against insider trading.

## 1.8 Appendix

### 1.8.1 Proof of Lemma 1

Since each trader is infinitesimal and orders are limited by the amount, his or her action has no impact on the market parameters (i.e., the exchange spread  $A$  and the dark pool execution probabilities  $(\bar{R}, \underline{R})$ ). Therefore, splitting the order cannot affect the (per unit) profit in each venue. Without loss of generality, we focus on the case of a positive signal (the case for



a negative signal is similar). Suppose that the informed traders have signal  $s > 0$ . Then he has a belief  $B(s) > \frac{1}{2}$ . Because the profit of a “Buy” order in each venue is strictly higher than the profit of a “Sell” order, thus it is optimal to choose the “Buy” direction. From his or her perspective, given the exchange spread  $A$  and the dark pool execution probabilities  $(\bar{R}, \underline{R})$ , the expected (per unit) profit for trading in the lit market, dark pool, and not trade depends on his or her confidence level  $B(|s|)$  and is determined by (1.5), (1.6), and (1.7), respectively.

Because these payoffs are linear in  $B(|s|)$ , given any belief  $B(|s|)$ , there is always one venue that is no worse than any of other venues. This relationship is shown in Figure 1.5a. When  $s \neq \pm s_0$  or  $\pm s_1$ , the payoff of trading in one venue is strictly better than others, and it is optimal to send the entire order to that venue. When  $s = \pm s_0$  or  $\pm s_1$ , there are two venues that yield the same payoff, and the trader can choose to split the order or not between these two venues. However, since the realization of the signal among the informed traders are continuously distributed, the measure of informed traders who receive a particular signal is zero. That is, such traders who are indifferent to these two venues has a mass of zero in the market. Therefore, in probability one, all informed traders send entire order to either the exchange or the dark pool, or not trader at all.  $\square$

### 1.8.2 Proof of Lemma 2

From a type  $d$  liquidity trader’s perspective, the expected per unit payoff from trading in the lit market, dark pool, and completely deference are determined by (1.8), (1.9), and (1.10), respectively. Since each individual has no impact to the market, given  $A$ ,  $\bar{R}$ ,  $\underline{R}$ , the per unit payoff in each venue is fixed. There is always one venue that is no worse than others. In addition, the payoff is linear in the number of units transacted. Hence there is no need to split among different venues or among different periods.  $\square$

### 1.8.3 Proof of Theorem 1

Hereafter we normalize some variables via dividing by  $\sigma_e$ , i.e., let  $\mathbf{s} = \frac{s}{\sigma_e}$ ,  $\mathbf{s}_0 = \frac{s_0}{\sigma_e}$ ,  $\mathbf{s}_1 = \frac{s_1}{\sigma_e}$ ,  $\widehat{\mathbf{s}} = \frac{\widehat{s}}{\sigma_e}$ ,  $\boldsymbol{\sigma} = \frac{\sigma_v}{\sigma_e}$ . Then it is equivalent to prove that, given  $\boldsymbol{\sigma} \geq 0$ , there is a unique cut-off  $\widehat{\mathbf{s}}$  such that  $h_I^{\mathbb{S}}(\mathbf{s})$ ,  $h_{U,\iota}^{\mathbb{S}}(d)$ ,  $A^{\mathbb{S}}$ ,  $\overline{\gamma}_e^{\mathbb{S}}$ ,  $\underline{\gamma}_e^{\mathbb{S}}$ ,  $\alpha_e^{\mathbb{S}}$  consist a equilibrium, in which

$$h_I^{\mathbb{S}}(\mathbf{s}) = \begin{cases} (\text{“Buy”}, \text{Exchange(Lit)}) & \text{if } \mathbf{s} \geq \widehat{\mathbf{s}}, \\ (\text{“Sell”}, \text{Exchange(Lit)}) & \text{if } \mathbf{s} < -\widehat{\mathbf{s}}, \\ \text{Not trade} & \text{otherwise,} \end{cases} \quad (1.34)$$

$$h_{U,\iota}^{\mathbb{S}}(d) = \begin{cases} (\text{“Buy” if } \iota=\text{Buyer, or “Sell” if } \iota=\text{Seller, Exchange(Lit)}) & \text{if } d \geq 2B(\widehat{\mathbf{s}}) - 1, \\ \text{Delay trade} & \text{otherwise,} \end{cases} \quad (1.35)$$

$$\overline{\gamma}_e^{\mathbb{S}} = 1 - \Phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma}), \quad (1.36)$$

$$\underline{\gamma}_e^{\mathbb{S}} = 1 - \Phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}), \quad (1.37)$$

$$\alpha_e^{\mathbb{S}} = 1 - G(2B(\widehat{\mathbf{s}}) - 1), \quad (1.38)$$

$$\frac{A^{\mathbb{S}}}{\sigma_v} = \frac{\overline{\gamma}_e^{\mathbb{S}} - \underline{\gamma}_e^{\mathbb{S}}}{\overline{\gamma}_e^{\mathbb{S}} + \underline{\gamma}_e^{\mathbb{S}} + \alpha_e^{\mathbb{S}} \frac{\mu z}{\mu}}, \quad (1.39)$$

where  $\widehat{\mathbf{s}}$  is determined by

$$2B(\widehat{\mathbf{s}}) - 1 = \frac{A^{\mathbb{S}}}{\sigma_v}. \quad (1.40)$$

We prove the theorem in two steps. First, we show that if  $\widehat{\mathbf{s}}$  is given, the other variables  $h_I^{\mathbb{S}}(\mathbf{s})$ ,  $h_{U,\iota}^{\mathbb{S}}(d)$ ,  $A^{\mathbb{S}}$ ,  $\overline{\gamma}_e^{\mathbb{S}}$ ,  $\underline{\gamma}_e^{\mathbb{S}}$ ,  $\alpha_e^{\mathbb{S}}$  solved from (1.34)-(1.39) form an equilibrium. Then we show that such  $\widehat{\mathbf{s}}$  exists and is unique.

Suppose that  $\widehat{\mathbf{s}}$  exists. By (1.40), an informed trader with signal  $\widehat{\mathbf{s}}$  is indifferent between trading in the exchange and not trade. Since  $B(\mathbf{s})$  is increasing in  $\mathbf{s}$ ,  $h_I^{\mathbb{S}}(\mathbf{s})$  is an optimal strategy for informed traders. Similarly, since a type  $\widehat{d} = 2B(\widehat{\mathbf{s}}) - 1$  uninformed liquidity trader is indifferent between trading on the exchange and deferring trade,  $h_{U,\iota}^{\mathbb{S}}(d)$  is an optimal strategy for uninformed traders. By the law of large numbers, given  $h_I^{\mathbb{S}}(\mathbf{s})$  and  $h_{U,\iota}^{\mathbb{S}}(d)$ , the fraction of uninformed traders who trade in the exchange would be  $\alpha_e^{\mathbb{S}} = \Pr(d \geq$

$\widehat{d}) = 1 - G(2B(\widehat{\mathbf{s}}) - 1)$ . Thus, the fraction of informed traders who trade in the “right direction” would be  $\overline{\gamma}_e^{\mathbf{s}} = \Pr(\mathbf{s} \geq \widehat{\mathbf{s}}) = 1 - \Phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma})$ , and the fraction of informed traders who trade in the “wrong direction” would be  $\underline{\gamma}_e^{\mathbf{s}} = \Pr(\mathbf{s} < \widehat{\mathbf{s}}) = 1 - \Phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma})$ . In addition, for given  $\overline{\gamma}_e^{\mathbf{s}}$ ,  $\underline{\gamma}_e^{\mathbf{s}}$ ,  $\alpha_e^{\mathbf{s}}$ , we can find  $A^{\mathbf{s}}$  from (1.39) and it would make the market maker on the exchange breaks even on average. Thus,  $h_I^{\mathbf{s}}(\widehat{\mathbf{s}})$ ,  $h_{U,\iota}^{\mathbf{s}}(d)$ ,  $A^{\mathbf{s}}$ ,  $\overline{\gamma}_e^{\mathbf{s}}$ ,  $\underline{\gamma}_e^{\mathbf{s}}$ ,  $\alpha_e^{\mathbf{s}}$  indeed form an equilibrium.

Then we will prove that such  $\widehat{\mathbf{s}}$  exists and is unique. After substituting the expressions of  $A^{\mathbf{s}}$ ,  $\overline{\gamma}_e^{\mathbf{s}}$ ,  $\underline{\gamma}_e^{\mathbf{s}}$ ,  $\alpha_e^{\mathbf{s}}$  into (1.40), we obtain the following equation for  $\widehat{\mathbf{s}}$ :

$$\frac{\Phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) - \Phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma})}{2 - \Phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) - \Phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma}) + (1 - G(2B(\widehat{\mathbf{s}}) - 1))\frac{\mu_z}{\mu}} = 2B(\widehat{\mathbf{s}}) - 1. \quad (1.41)$$

Define

$$f(s) = (2B(s) - 1) \left[ 2 - \Phi(s + \boldsymbol{\sigma}) - \Phi(s - \boldsymbol{\sigma}) + (1 - G(2B(s) - 1))\frac{\mu_z}{\mu} \right] - [\Phi(s + \boldsymbol{\sigma}) - \Phi(s - \boldsymbol{\sigma})],$$

and the derivative of  $f(s)$  is

$$f'(s) = 2B'(s) \left[ 2 - \Phi(s + \boldsymbol{\sigma}) - \Phi(s - \boldsymbol{\sigma}) + (1 - G(2B(s) - 1))\frac{\mu_z}{\mu} \right] - 2(2B(s) - 1)G'(2B(s) - 1)B'(s).$$

We can easily find that  $f(\frac{1}{2}) < 0$ ,  $f(+\infty) > 0$ ,  $f'(0) > 0$ ,  $f'(+\infty) = 0$ . Because  $G'(x) + xG''(x) \geq 0$ ,  $\forall x \in [0, 1]$ , we have  $f''(s) < 0$ . Thus there exists a unique  $\widehat{\mathbf{s}}$  such that  $f(\widehat{\mathbf{s}}) = 0$ .  $\square$

#### 1.8.4 Proof of Theorem 2

Hereafter we normalize some variables via dividing by  $\sigma_e$ , i.e.,  $\mathbf{s} = \frac{s}{\sigma_e}$ ,  $\mathbf{s}_0 = \frac{s_0}{\sigma_e}$ ,  $\mathbf{s}_1 = \frac{s_1}{\sigma_e}$ ,  $\widehat{\mathbf{s}} = \frac{\widehat{s}}{\sigma_e}$ ,  $\boldsymbol{\sigma} = \frac{\sigma_v}{\sigma_e}$ . Then finding the equilibrium is equivalent to solving the following system

of equations:

$$B(\mathbf{s}_0)(\bar{R} + \underline{R}) = \bar{R}, \quad (1.42)$$

$$B(\mathbf{s}_1) [(1 - \bar{R}) + (1 - \underline{R})] = \frac{A}{\sigma_v} + (1 - \bar{R}), \quad (1.43)$$

$$\bar{R} = \mathbb{E} \left[ \min \left\{ 1, \frac{\bar{\gamma}_d \mu + \alpha_d Z^+}{\underline{\gamma}_d \mu + \alpha_d Z^-} \right\} \right], \quad (1.44)$$

$$\underline{R} = \mathbb{E} \left[ \min \left\{ 1, \frac{\underline{\gamma}_d \mu + \alpha_d Z^-}{\bar{\gamma}_d \mu + \alpha_d Z^+} \right\} \right], \quad (1.45)$$

$$\frac{A}{\sigma_v} = \frac{\bar{\gamma}_e - \underline{\gamma}_e}{(\bar{\gamma}_e + \underline{\gamma}_e) + \alpha_e \frac{\mu_z}{\mu}}, \quad (1.46)$$

$$\bar{\gamma}_e = 1 - \Phi(\mathbf{s}_1 - \boldsymbol{\sigma}), \quad (1.47)$$

$$\underline{\gamma}_e = 1 - \Phi(\mathbf{s}_1 + \boldsymbol{\sigma}), \quad (1.48)$$

$$\bar{\gamma}_d = \Phi(\mathbf{s}_1 - \boldsymbol{\sigma}) - \Phi(\mathbf{s}_0 - \boldsymbol{\sigma}), \quad (1.49)$$

$$\underline{\gamma}_d = \Phi(\mathbf{s}_1 + \boldsymbol{\sigma}) - \Phi(\mathbf{s}_0 + \boldsymbol{\sigma}), \quad (1.50)$$

$$\alpha_e = 1 - G(2B(\mathbf{s}_1) - 1), \quad (1.51)$$

$$\alpha_d = G(2B(\mathbf{s}_1) - 1) - G(2B(\mathbf{s}_0) - 1), \quad (1.52)$$

where

$$B(\mathbf{s}) = \frac{\phi(\mathbf{s} - \boldsymbol{\sigma})}{\phi(\mathbf{s} - \boldsymbol{\sigma}) + \phi(\mathbf{s} + \boldsymbol{\sigma})}. \quad (1.53)$$

Before proving the existence of solutions to the system of equations, we introduce the following lemma.

**Lemma 3.** Let  $\mathbf{s}_0 \geq 0$  and  $\mathbf{s}_1 = \mathbf{s}_0 + \epsilon$ , we have

$$\begin{aligned} \lim_{\epsilon \rightarrow 0^+} \underline{R} &= \mathbb{E} \left[ \min \left\{ 1, \frac{\phi(\mathbf{s}_0 + \boldsymbol{\sigma})\mu + 2G'(2B(\mathbf{s}_0) - 1)B'(\mathbf{s}_0)Z^-}{\phi(\mathbf{s}_0 - \boldsymbol{\sigma})\mu + 2G'(2B(\mathbf{s}_0) - 1)B'(\mathbf{s}_0)Z^+} \right\} \right], \\ \lim_{\epsilon \rightarrow 0^+} \bar{R} &= \mathbb{E} \left[ \min \left\{ 1, \frac{\phi(\mathbf{s}_0 - \boldsymbol{\sigma})\mu + 2G'(2B(\mathbf{s}_0) - 1)B'(\mathbf{s}_0)Z^+}{\phi(\mathbf{s}_0 + \boldsymbol{\sigma})\mu + 2G'(2B(\mathbf{s}_0) - 1)B'(\mathbf{s}_0)Z^-} \right\} \right], \\ \lim_{\epsilon \rightarrow 0^+} \frac{A}{\sigma_v} &= \frac{\Phi(\mathbf{s}_0 + \boldsymbol{\sigma}) - \Phi(\mathbf{s}_0 - \boldsymbol{\sigma})}{2 - \Phi(\mathbf{s}_0 + \boldsymbol{\sigma}) - \Phi(\mathbf{s}_0 - \boldsymbol{\sigma}) + [1 - G(2B(\mathbf{s}_0) - 1)] \frac{\mu_z}{\mu}}. \end{aligned}$$

Moreover, if  $\mathbf{s}_0 = 0$  or  $\boldsymbol{\sigma} = 0$ , then  $\lim_{\epsilon \rightarrow 0^+} \underline{R} = \lim_{\epsilon \rightarrow 0^+} \bar{R} = 1$ . Therefore, we define  $\underline{R}$ ,  $\bar{R}$ , and  $\frac{A}{\sigma_v}$  use these limits when  $\mathbf{s}_0 = \mathbf{s}_1$ .

**Proof.** We can prove this by the Taylor expansion. Suppose that  $\epsilon$  is sufficiently small. Because  $\mathbf{s}_0 \geq 0$  and  $\mathbf{s}_1 = \mathbf{s}_0 + \epsilon$ , we have, by the Taylor expansion, that  $\underline{\gamma}_d = \phi(\mathbf{s}_0 + \boldsymbol{\sigma})\epsilon + o(\epsilon)$ ,  $\overline{\gamma}_d = \phi(\mathbf{s}_0 - \boldsymbol{\sigma})\epsilon + o(\epsilon)$ , and  $\alpha_d = 2G'(2B(\mathbf{s}_0) - 1)B'(\mathbf{s}_0)\epsilon + o(\epsilon)$ . Therefore we have

$$\begin{aligned}\underline{R} &= \mathbb{E} \left[ \min \left\{ 1, \frac{\phi(\mathbf{s}_0 + \boldsymbol{\sigma})\mu\epsilon + 2G'(2B(\mathbf{s}_0) - 1)B'(\mathbf{s}_0)Z^{-\epsilon}}{\phi(\mathbf{s}_0 - \boldsymbol{\sigma})\mu\epsilon + 2G'(2B(\mathbf{s}_0) - 1)B'(\mathbf{s}_0)Z^{+\epsilon}} + o(\epsilon) \right\} \right], \\ \overline{R} &= \mathbb{E} \left[ \min \left\{ 1, \frac{\phi(\mathbf{s}_0 - \boldsymbol{\sigma})\mu\epsilon + 2G'(2B(\mathbf{s}_0) - 1)B'(\mathbf{s}_0)Z^{+\epsilon}}{\phi(\mathbf{s}_0 + \boldsymbol{\sigma})\mu\epsilon + 2G'(2B(\mathbf{s}_0) - 1)B'(\mathbf{s}_0)Z^{-\epsilon}} + o(\epsilon) \right\} \right].\end{aligned}$$

Similarly, by the Taylor expansion, we have  $\overline{\gamma}_e = 1 - \Phi(\mathbf{s}_0 - \boldsymbol{\sigma}) - \phi(\mathbf{s}_0 - \boldsymbol{\sigma})\epsilon + o(\epsilon)$ ,  $\underline{\gamma}_e = 1 - \Phi(\mathbf{s}_0 + \boldsymbol{\sigma}) - \phi(\mathbf{s}_0 + \boldsymbol{\sigma})\epsilon + o(\epsilon)$ , and  $\alpha_e = [1 - G(2B(\mathbf{s}_0) - 1)] - 2G'(\cdot)B'(\mathbf{s}_0)\epsilon + o(\epsilon)$ .

Therefore

$$\begin{aligned}\frac{A}{\sigma_v} &= \\ &= \frac{\Phi(\mathbf{s}_0 + \boldsymbol{\sigma}) - \Phi(\mathbf{s}_0 - \boldsymbol{\sigma}) - [\phi(\mathbf{s}_0 + \boldsymbol{\sigma}) - \phi(\mathbf{s}_0 - \boldsymbol{\sigma})]\epsilon}{2 - \Phi(\mathbf{s}_0 + \boldsymbol{\sigma}) - \Phi(\mathbf{s}_0 - \boldsymbol{\sigma}) + [1 - G(2B(\mathbf{s}_0) - 1)]\frac{\mu z}{\mu} - [\phi(\mathbf{s}_0 + \boldsymbol{\sigma}) + \phi(\mathbf{s}_0 - \boldsymbol{\sigma}) + 2G'(\cdot)B'(\mathbf{s}_0)]\epsilon} \\ &+ o(\epsilon).\end{aligned}$$

Let  $\epsilon \rightarrow 0^+$ , and we prove the lemma.  $\square$

We prove the theorem in a similar way as in the proof of Theorem 1. First, we show that if  $\mathbf{s}_0$  and  $\mathbf{s}_1$  are given, the other variables  $h_I(\cdot)$ ,  $h_{U,\iota}(\cdot)$ ,  $A$ ,  $\overline{R}$ ,  $\underline{R}$ ,  $\overline{\gamma}_e$ ,  $\underline{\gamma}_e$ ,  $\overline{\gamma}_d$ ,  $\underline{\gamma}_d$ ,  $\alpha_d$ ,  $\alpha_e$  solved from (1.44)-(1.52) form an equilibrium. Then we show that  $(\mathbf{s}_0, \mathbf{s}_1)$  exists and is unique.

Given  $A$ ,  $\overline{R}$ ,  $\underline{R}$ ,  $\overline{\gamma}_e$ ,  $\underline{\gamma}_e$ ,  $\overline{\gamma}_d$ ,  $\underline{\gamma}_d$ ,  $\alpha_d$ ,  $\alpha_e$  and that  $\mathbf{s}_0, \mathbf{s}_1$  determined by (1.42), (1.43),  $0 < \mathbf{s}_0 < \mathbf{s}_1$ , we show that it is optimal for informed speculators and uninformed liquidity buyers (and sellers) to following the strategy described respectively by  $h_I(\cdot)$  and  $h_{U,\iota}(\cdot)$ ,  $\iota \in \{Buyer, Seller\}$ .

Consider an informed speculator who receives a signal  $\mathbf{s} \geq 0$  (the case when  $\mathbf{s} \leq 0$  is symmetric with respect to the vertical axis, and hence the analysis is similar and skipped here). Suppose that  $0 < \mathbf{s}_0 < \mathbf{s}_1$ . From his or her perspective, the expected payoffs in the lit market, the dark pool, and no-trade are, respectively,  $[B(\mathbf{s})\sigma_v - (1 - B(\mathbf{s}))\sigma_v] - A$ ,

$B(\mathbf{s})\underline{R}\sigma_v - (1 - B(\mathbf{s}))\bar{R}\sigma_v$ , and 0. Figure 1.5a captures the payoff as a function of  $B(\mathbf{s})$ . As one can see in the graph, since the payoffs are linear with respect to  $B(\mathbf{s})$ , and  $B(\mathbf{s})$  is strictly increasing with respect to  $\mathbf{s}$ , the optimal strategy for an informed speculator with signal  $\mathbf{s}$  should use the exchange (the lit market) to trade when his or her signal  $\mathbf{s} \geq \mathbf{s}_1$ , and the dark pool when  $\mathbf{s}_0 \leq \mathbf{s} < \mathbf{s}_1$ , and stay outside when  $\mathbf{s} < \mathbf{s}_0$ . This is marked as the red line in Figure 1.5a.

The fractions of each type of traders in each venue  $\bar{\gamma}_e, \underline{\gamma}_e, \bar{\gamma}_d, \underline{\gamma}_d, \alpha_e, \alpha_d$  are determined by (1.47), (1.48), (1.49), (1.50), (1.51), (1.52), respectively, and  $A, \bar{R}, \underline{R}$  are given by (1.46), (1.44), (1.45). Thus properties (ii), (iii) and (iv) in Definition 2 are satisfied.

Then we need to show that such pair of cut-off  $(\mathbf{s}_0, \mathbf{s}_1)$  exists and satisfies  $0 < \mathbf{s}_0 < \mathbf{s}_1$ . In order to show this, we consider equations (1.42) and (1.43) and show that there is a intersection for the two lines represented by these two equations.

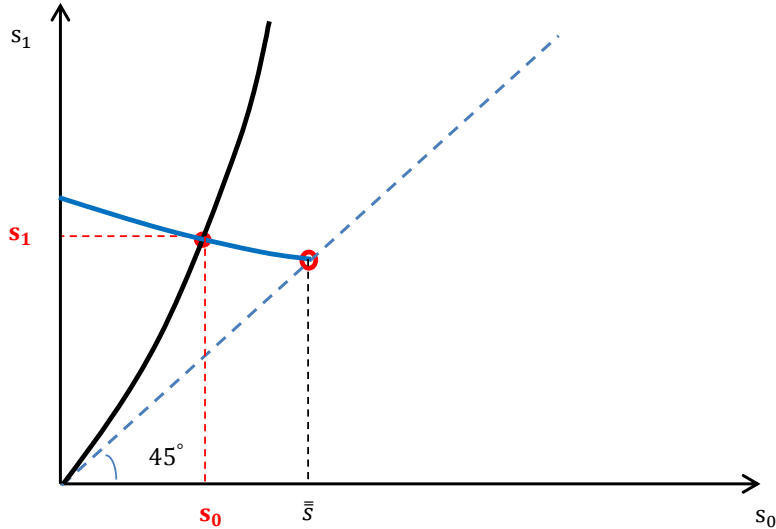


Figure 1.15: Equilibrium Existence

For equation (1.42), we show that  $(\mathbf{s}_0, \mathbf{s}_1) = (0, 0)$  satisfies equation (1.42) and behaves as the black line in Figure 1.15.

(i) Suppose  $\mathbf{s}_0 = 0, \mathbf{s}_1 = 0$ , then  $B(\mathbf{s}_0) = \frac{1}{2}$ , and by Lemma 3,  $\bar{R} = \underline{R} = 1$ . Therefore

equation (1.42) is satisfied.

(ii) Now suppose that  $\mathbf{s}_0 > 0$ , then  $\frac{1}{2} < B(\mathbf{s}_0) < 1$ . To satisfy (1.42), we need that  $\underline{R} < \bar{R} \leq 1$ , thus  $|\underline{\gamma}_d| < |\bar{\gamma}_d|$ . To obtain this, it must be true that  $\mathbf{s}_1 > \mathbf{s}_0$  if such  $\mathbf{s}_1$  exists. By continuity such  $\mathbf{s}_1$  must exist for a small enough  $\mathbf{s}_0$ . (Note that if  $\mathbf{s}_0$  is too large, such  $\mathbf{s}_1$  may not exist.)

(iii) We also show that there exist some  $\bar{s}$  such that  $\mathbf{s}_1 \rightarrow +\infty$  when  $\mathbf{s}_0 \rightarrow \bar{s}$ . We rewrite equation (1.42) as  $B(\mathbf{s}_0) = \frac{\bar{R}}{\bar{R} + \underline{R}}$ . As  $\mathbf{s}_1 \rightarrow +\infty$ ,  $\bar{\gamma}_d \rightarrow 1 - \Phi(\mathbf{s}_0 - \sigma)$ ,  $\underline{\gamma}_d \rightarrow 1 - \Phi(\mathbf{s}_0 + \sigma)$ ,  $\alpha_d \rightarrow 1 - B(\mathbf{s}_0)$ . Hence  $\bar{R} \rightarrow \mathbb{E} \left[ \min \left\{ 1, \frac{1 - \Phi(\mathbf{s}_0 - \sigma) + [1 - B(\mathbf{s}_0)]Z^+}{1 - \Phi(\mathbf{s}_0 + \sigma) + [1 - B(\mathbf{s}_0)]Z^-} \right\} \right]$ , and  $\underline{R} \rightarrow \mathbb{E} \left[ \min \left\{ 1, \frac{1 - \Phi(\mathbf{s}_0 + \sigma) + [1 - B(\mathbf{s}_0)]Z^-}{1 - \Phi(\mathbf{s}_0 - \sigma) + [1 - B(\mathbf{s}_0)]Z^+} \right\} \right]$ . Therefore, for any  $\mathbf{s}_0 \in [0, \infty)$ , there must exist  $\bar{\gamma}_d > \underline{\gamma}_d$ , thus  $\bar{R} > \underline{R}$ . Then let  $\mathbf{s}_1 \rightarrow +\infty$ , the left hand side of the equation,  $B(\mathbf{s}_0)$ , is equal to  $\frac{1}{2}$  if  $\mathbf{s}_0 = 0$ , and is equal to 1 if  $\mathbf{s}_0 \rightarrow +\infty$ . However, the right hand side of the equation,  $\frac{\bar{R}}{\bar{R} + \underline{R}}$ , is greater than  $\frac{1}{2}$  if  $\mathbf{s}_0 = 0$ , and equal to  $\frac{1}{2}$  if  $\mathbf{s}_0 \rightarrow +\infty$ . This is because  $\lim_{s \rightarrow +\infty} \frac{1 - \Phi(\mathbf{s}_0 - \sigma)}{1 - B(\mathbf{s}_0)} = \lim_{s \rightarrow +\infty} \frac{1 - \Phi(\mathbf{s}_0 + \sigma)}{1 - B(\mathbf{s}_0)} = 0$ , so  $\lim_{s \rightarrow +\infty} \bar{R} = \lim_{s \rightarrow +\infty} \underline{R} = \mathbb{E} \left[ \min \left\{ 1, \frac{Z^+}{Z^-} \right\} \right]$ . By continuity, there must exist an  $\bar{s} \in (0, +\infty)$  such that, as  $\mathbf{s}_0 \rightarrow \bar{s}$ ,  $\mathbf{s}_1 \rightarrow +\infty$ , LHS = RHS. That is, equation (1.42) is satisfied.

For equation (1.43). We rewrite it as

$$B(\mathbf{s}_1) = \frac{\frac{A}{\sigma_v}}{(1 - \bar{R}) + (1 - \underline{R})} + \frac{(1 - \bar{R})}{(1 - \bar{R}) + (1 - \underline{R})}. \quad (1.54)$$

(i) Suppose that  $\mathbf{s}_0 = 0$ , we prove that there must exist a  $\mathbf{s}_1 > 0$  satisfy (1.43). Note that for any given  $\sigma \in (0, +\infty)$ ,  $A > 0$  is satisfied. If  $\mathbf{s}_1 = 0$ , we have  $B(\mathbf{s}_1) = \frac{1}{2}$  and  $\bar{R} = \underline{R} = 1$ . Plugging into (1.43) gives us  $A = 0$ , which contradicts the fact that  $A > 0$ . If  $\mathbf{s}_1 < 0$ , then  $B(\mathbf{s}_1) < \frac{1}{2}$ ,  $\bar{\gamma}_d < \underline{\gamma}_d$ , and  $0 < \bar{R} < \underline{R} < 1$  (we don't consider any  $\bar{R}, \underline{R} < 0$ ). Hence  $\frac{(1 - \bar{R})}{(1 - \bar{R}) + (1 - \underline{R})} > \frac{1}{2} > B(\mathbf{s}_1)$ . In order for (1.43) to be satisfied, we have  $A < 0$ , which contradicts with that fact that  $A > 0$ . Then we show the existence of  $\mathbf{s}_1$  using the continuity of equation (1.54). Its left hand side  $B(\mathbf{s}_1)$  is increasing in  $\mathbf{s}_1$  and  $B(0) = \frac{1}{2}$ ,  $\lim_{\mathbf{s}_1 \rightarrow \infty} B(\mathbf{s}_1) = 1$ . If  $\mathbf{s}_1 = 0$ , the right hand side equals  $\frac{\frac{A}{\sigma_v}}{(1 - \bar{R}) + (1 - \underline{R})} + \frac{1}{2} > \frac{1}{2}$ . However, when  $\mathbf{s}_1 \rightarrow \infty$ , we have  $A \rightarrow 0$  and  $1 > \bar{R} > \underline{R}$ , hence the right hand side equals  $0 + \frac{1 - \bar{R}}{(1 - \bar{R}) + (1 - \underline{R})} < \frac{1}{2}$ . By

the continuity of equation (1.54), there must exist a  $\mathbf{s}_1 \in (0, +\infty)$  such that the equation is satisfied.

(ii) Next we prove that there exist an  $\bar{\mathbf{s}} > 0$  and small enough  $\epsilon > 0$  such that for  $\mathbf{s}_0 = \bar{\mathbf{s}}$ ,  $\mathbf{s}_1 = \bar{\mathbf{s}} + \epsilon$ , equation (1.43) is satisfied as  $\epsilon \rightarrow 0^+$ . Consider any  $\mathbf{s}_0 = \mathbf{s}$ ,  $\mathbf{s}_1 = \mathbf{s} + \epsilon$ , when  $\epsilon > 0$  is sufficiently small. By Lemma 3, equation (1.43) is equivalent to

$$B(\mathbf{s}) = \frac{\frac{A}{\sigma_v}}{(1 - \bar{R}) + (1 - \underline{R})} + \frac{(1 - \bar{R})}{(1 - \bar{R}) + (1 - \underline{R})}, \quad (1.55)$$

where  $\frac{A}{\sigma_v} = \frac{\Phi(\mathbf{s}+\sigma) - \Phi(\mathbf{s}-\sigma)}{2 - \Phi(\mathbf{s}+\sigma) - \Phi(\mathbf{s}-\sigma) + [1 - G(2B(\mathbf{s}) - 1)] \frac{\mu z}{\mu}}$ ,  $\underline{R} = \mathbb{E} \left[ \min \left\{ 1, \frac{\phi(\mathbf{s}-\sigma)\mu + 2G'(2B(\mathbf{s})-1)B'(\mathbf{s})Z^+}{\phi(\mathbf{s}+\sigma)\mu + 2G'(2B(\mathbf{s})-1)B'(\mathbf{s})Z^-} \right\} \right]$ , and  $\bar{R} = \mathbb{E} \left[ \min \left\{ 1, \frac{\phi(\mathbf{s}+\sigma)\mu + 2G'(2B(\mathbf{s})-1)B'(\mathbf{s})Z^-}{\phi(\mathbf{s}-\sigma)\mu + 2G'(2B(\mathbf{s})-1)B'(\mathbf{s})Z^+} \right\} \right]$ . Consider  $\mathbf{s}$  on  $[0, \infty)$ . The left hand side of equation (1.55) increases with respect to  $\mathbf{s}$ . We have  $B(0) = \frac{1}{2}$ , and  $\lim_{\mathbf{s} \rightarrow \infty} B(\mathbf{s}) = 1$ . Now consider the right hand side of equation (1.55). By Lemma 3, we know that if  $\mathbf{s} \rightarrow 0^+$ , the limit of the right hand side is  $\frac{\frac{A}{\sigma_v}}{(1-\bar{R})+(1-\underline{R})} + \frac{1}{2} > \frac{1}{2}$ . If  $\mathbf{s} \rightarrow \infty$ , we have  $A \rightarrow 0$  and  $1 > \bar{R} > \underline{R}$ , hence the limit of the right hand side is  $0 + \frac{1-\bar{R}}{(1-\bar{R})+(1-\underline{R})} < \frac{1}{2}$ . By continuity there must exist a  $\bar{\mathbf{s}} \in (0, \infty)$  such that equation (1.55) is satisfied at  $(\bar{\mathbf{s}}, \bar{\mathbf{s}})$  (i.e.,  $\mathbf{s}_0 = \mathbf{s}_1 = \bar{\mathbf{s}}$ ).

The above argument can be summarized by Figure 1.15. Given  $\sigma > 0$  fixed, the black curve represents the  $(\mathbf{s}_0, \mathbf{s}_1)$  pairs that satisfy equation (1.42). It goes through the point  $(0, 0)$ , is always above the line  $\mathbf{s}_1 = \mathbf{s}_0$ , and  $\mathbf{s}_1 \rightarrow +\infty$  when  $\mathbf{s}_0 \rightarrow \bar{\mathbf{s}}$ . The red curve represents the  $(\mathbf{s}_0, \mathbf{s}_1)$  pairs that satisfy equation (1.43). When  $\mathbf{s}_0 = 0$ ,  $\mathbf{s}_1 \in (0, \infty)$ . And there exists some  $\bar{\mathbf{s}} \in (0, +\infty)$  such that  $\mathbf{s}_0 = \mathbf{s}_1 = \bar{\mathbf{s}}$ , satisfies equation (1.43). Then because all functions are continuous, there must exist a pair  $(\mathbf{s}_0, \mathbf{s}_1)$ ,  $0 < \mathbf{s}_0 < \mathbf{s}_1 < +\infty$ , such that both equations (1.42) and (1.43) are satisfied. It is the intersection of the black curve and the red curve in Figure 1.15. The existence is then established.  $\square$

### 1.8.5 Proof of Proposition 1 and Proposition 2

To prove Propositions 1 and 2, we need the following two lemmas.

**Lemma 4.** Suppose  $s(\sigma)$  is continuously differentiable over  $(0, +\infty)$ , and  $\lim_{\sigma \rightarrow 0^+} s(\sigma)\sigma = 0$ ,



then

$$\begin{aligned}\lim_{\sigma \rightarrow 0^+} (\phi(s(\sigma) + \sigma) - \phi(s(\sigma) - \sigma)) s'(\sigma) &= 0 \\ \lim_{\sigma \rightarrow 0^+} (\Phi(s(\sigma) + \sigma) - \Phi(s(\sigma) - \sigma)) s'(\sigma) &= 0\end{aligned}$$

In addition,

(i) If  $\lim_{\sigma \rightarrow 0^+} s(\sigma) = \pm\infty$ ,  $|\sigma s'(\sigma)| \leq s(\sigma)$  for sufficiently small  $\sigma$ .

(ii) If  $-\infty < \lim_{\sigma \rightarrow 0^+} s(\sigma) < +\infty$ ,  $\lim_{\sigma \rightarrow 0^+} \sigma s'(\sigma) = 0$ .

**Proof.** (i) Suppose that  $\lim_{\sigma \rightarrow 0^+} s(\sigma) = +\infty$ . There exists  $\epsilon > 0$  such that  $\forall \sigma \in (0, \epsilon)$ ,  $s(\sigma)\sigma > 0$  and  $\frac{d(s(\sigma)\sigma)}{d\sigma} > 0$ . Thus  $\frac{d(s(\sigma)\sigma)}{d\sigma} = \sigma s'(\sigma) + s(\sigma) \geq 0$ , and

$$|\sigma s'(\sigma)| \leq |s(\sigma)|,$$

for  $\sigma \in (0, \epsilon)$ . Similarly, if  $\lim_{\sigma \rightarrow 0^+} s(\sigma) = -\infty$ , we have that

$$|\sigma s'(\sigma)| \leq |s(\sigma)|,$$

for sufficiently small  $\sigma$ .

Therefore, by mean value theorem, we have

$$\begin{aligned}\lim_{\sigma \rightarrow 0^+} (\phi(s(\sigma) + \sigma) - \phi(s(\sigma) - \sigma)) s'(\sigma) &= \lim_{\sigma \rightarrow 0^+} \int_{s(\sigma) - \sigma}^{s(\sigma) + \sigma} -xe^{-\frac{x^2}{2}} dx s'(\sigma) \\ &= \lim_{\sigma \rightarrow 0^+} -2\sigma s(\sigma) e^{-\frac{s(\sigma)^2}{2}} s'(\sigma).\end{aligned}$$

Because  $|\sigma s'(\sigma)| \leq |s(\sigma)|$  and  $\lim_{\sigma \rightarrow 0^+} \left| -2s(\sigma)^2 e^{-\frac{s(\sigma)^2}{2}} \right| = 0$ , we obtain

$$\lim_{\sigma \rightarrow 0^+} (\phi(s(\sigma) + \sigma) - \phi(s(\sigma) - \sigma)) s'(\sigma) = 0.$$

Similarly, we have

$$\begin{aligned}\lim_{\sigma \rightarrow 0^+} (\Phi(s(\sigma) + \sigma) - \Phi(s(\sigma) - \sigma)) s'(\sigma) &= \lim_{\sigma \rightarrow 0^+} \int_{s(\sigma) - \sigma}^{s(\sigma) + \sigma} e^{-\frac{x^2}{2}} dx s'(\sigma) \\ &= \lim_{\sigma \rightarrow 0^+} 2\sigma e^{-\frac{s(\sigma)^2}{2}} s'(\sigma).\end{aligned}$$

Additionally,  $\lim_{\sigma \rightarrow 0^+} \left| -2s(\sigma)e^{-\frac{s(\sigma)^2}{2}} \right| = 0$  gives us that

$$\lim_{\sigma \rightarrow 0^+} (\Phi(s(\sigma) + \sigma) - \Phi(s(\sigma) - \sigma)) s'(\sigma) = 0.$$

(ii) Suppose that  $\lim_{\sigma \rightarrow 0^+} s < +\infty$ . On one hand, we have that  $\lim_{\sigma \rightarrow 0^+} \frac{d(s(\sigma)\sigma)}{d\sigma} = \lim_{\sigma \rightarrow 0^+} \sigma s'(\sigma) + \lim_{\sigma \rightarrow 0^+} s(\sigma) = \lim_{\sigma \rightarrow 0^+} \sigma s'(\sigma) + s(0)$ . On the other hand, we have

$$\frac{d(s(\sigma)\sigma)}{d\sigma} \Big|_{\sigma=0} = \lim_{\sigma \rightarrow 0^+} \frac{s(\sigma)\sigma - 0}{\sigma - 0} = s(0). \quad (1.56)$$

Thus we have

$$\lim_{\sigma \rightarrow 0^+} \sigma s'(\sigma) = 0,$$

and

$$\begin{aligned} \lim_{\sigma \rightarrow 0^+} (\phi(s(\sigma) + \sigma) - \phi(s(\sigma) - \sigma)) s'(\sigma) &= \lim_{\sigma \rightarrow 0^+} \int_{s(\sigma) - \sigma}^{s(\sigma) + \sigma} -xe^{-\frac{x^2}{2}} dx s'(\sigma) \\ &= \lim_{\sigma \rightarrow 0^+} (-2\sigma s'(\sigma)) \cdot \lim_{\sigma \rightarrow 0^+} s(\sigma) e^{-\frac{s(\sigma)^2}{2}} \\ &= 0, \end{aligned}$$

$$\begin{aligned} \lim_{\sigma \rightarrow 0^+} (\Phi(s(\sigma) + \sigma) - \Phi(s(\sigma) - \sigma)) s'(\sigma) &= \lim_{\sigma \rightarrow 0^+} \int_{s(\sigma) - \sigma}^{s(\sigma) + \sigma} -xe^{-\frac{x^2}{2}} dx s'(\sigma) \\ &= \lim_{\sigma \rightarrow 0^+} (\sigma s'(\sigma)) \cdot \lim_{\sigma \rightarrow 0^+} e^{-\frac{s(\sigma)^2}{2}} \\ &= 0. \end{aligned}$$

□

**Lemma 5.**  $\lim_{\sigma \rightarrow 0^+} \widehat{s} = s^*$ , where  $s^* \in (0, +\infty)$  is determined by the following equation

$$s = \frac{2\phi(s)}{2 - 2\Phi(s) + \frac{\mu z}{\mu}}.$$

**Proof.** Because  $G(\cdot)$ ,  $\Phi(\cdot) \in C^2$ . The implicit function theorem and the uniqueness of  $\widehat{s}$  show that  $\widehat{s}(\sigma)$  is a continuously differentiable function over  $(0, +\infty)$ .

When  $\boldsymbol{\sigma} = 0$ , we have  $\overline{\gamma_e^S} - \underline{\gamma_e^S} = 0$  and  $\frac{A^S}{\sigma_v} = 0$ . Equation (1.40) gives us that  $B(\widehat{\mathbf{s}}) = \frac{1}{2}$  and  $\widehat{\mathbf{s}}(\boldsymbol{\sigma})\boldsymbol{\sigma} = 0$ .

Recall that

$$\frac{A^S}{\sigma_v} = \frac{\Phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) - \Phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma})}{2 - \Phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) - \Phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma}) + (1 - G(2B(\widehat{\mathbf{s}}) - 1))^{\frac{\mu_z}{\mu}}}, \quad (1.57)$$

$G(\cdot)$ ,  $\Phi(\cdot) \in C^2$ , and  $\frac{A^S}{\sigma_v}$  is differentiable of  $\boldsymbol{\sigma}$  over  $(0, +\infty)$ .

Taking the derivative, we get

$$\begin{aligned} \frac{d\left(\frac{A^S}{\sigma_v}\right)}{d\boldsymbol{\sigma}} &= \frac{(\phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) - \phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma})) \frac{d\widehat{\mathbf{s}}}{d\boldsymbol{\sigma}} + (\phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) + \phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma}))}{\overline{\gamma_e} + \underline{\gamma_e} + \alpha_e \frac{\mu_z}{\mu}} \\ &+ \frac{[\Phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) - \Phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma})] \left[ (\phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) + \phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma})) \frac{d\widehat{\mathbf{s}}}{d\boldsymbol{\sigma}} + (\phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) - \phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma})) \right]}{\left[ \overline{\gamma_e} + \underline{\gamma_e} + \alpha_e \frac{\mu_z}{\mu} \right]^2} \\ &+ \frac{2G'(2B(\widehat{\mathbf{s}}) - 1)^{\frac{\mu_z}{\mu}} [\Phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) - \Phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma})] \left( \frac{\partial B(\widehat{\mathbf{s}})}{\partial \widehat{\mathbf{s}}} \frac{d\widehat{\mathbf{s}}}{d\boldsymbol{\sigma}} + \frac{\partial B(\widehat{\mathbf{s}})}{\partial \boldsymbol{\sigma}} \right)}{\left[ \overline{\gamma_e} + \underline{\gamma_e} + \alpha_e \frac{\mu_z}{\mu} \right]^2}. \end{aligned}$$

Lemma 4 gives us

$$\lim_{\boldsymbol{\sigma} \rightarrow 0^+} \frac{d\left(\frac{A^S}{\sigma_v}\right)}{d\boldsymbol{\sigma}} = \lim_{\boldsymbol{\sigma} \rightarrow 0^+} \frac{2\phi(\widehat{\mathbf{s}})}{2 - 2\Phi(\widehat{\mathbf{s}}) + \frac{\mu_z}{\mu}}. \quad (1.58)$$

On the other hand, from equation (1.40), we have

$$\frac{A^S}{\sigma_v} = 2B(\widehat{\mathbf{s}}) - 1.$$

Taking derivative with respect to  $\boldsymbol{\sigma}$ , we get

$$\frac{d\left(\frac{A^S}{\sigma_v}\right)}{d\boldsymbol{\sigma}} = 2B(\widehat{\mathbf{s}}) [1 - B(\widehat{\mathbf{s}})] \left( 2\boldsymbol{\sigma} \frac{d\widehat{\mathbf{s}}}{d\boldsymbol{\sigma}} + 2\widehat{\mathbf{s}} \right).$$

Using Lemma 4 and  $\lim_{\boldsymbol{\sigma} \rightarrow 0^+} \widehat{\mathbf{s}}(\boldsymbol{\sigma})\boldsymbol{\sigma} = 0$ , we obtain

$$\lim_{\boldsymbol{\sigma} \rightarrow 0^+} \frac{d\left(\frac{A^S}{\sigma_v}\right)}{d\boldsymbol{\sigma}} = \lim_{\boldsymbol{\sigma} \rightarrow 0^+} \left( \boldsymbol{\sigma} \frac{d\widehat{\mathbf{s}}}{d\boldsymbol{\sigma}} + \widehat{\mathbf{s}} \right). \quad (1.59)$$

Combing equations (1.59) and (1.58), we have that

$$\lim_{\sigma \rightarrow 0^+} \left( \sigma \frac{d\widehat{\mathbf{s}}}{d\sigma} + \widehat{\mathbf{s}} \right) = \lim_{\sigma \rightarrow 0^+} \frac{2\phi(\widehat{\mathbf{s}})}{2 - 2\Phi(\widehat{\mathbf{s}}) + \frac{\mu z}{\mu}}.$$

Suppose that  $\lim_{\sigma \rightarrow 0^+} \widehat{\mathbf{s}} = +\infty$ , then we have, as we do in the proof of Lemma 4,  $\lim_{\sigma \rightarrow 0^+} \sigma \frac{d\widehat{\mathbf{s}}}{d\sigma} + \widehat{\mathbf{s}} > 0$ , which contradicts with  $\lim_{\sigma \rightarrow 0^+} \frac{2\phi(\widehat{\mathbf{s}})}{2 - 2\Phi(\widehat{\mathbf{s}}) + \frac{\mu z}{\mu}} = 0$ .

Then we have to show that the limit can not be zero. Because the limit can not be infinity, we have  $\lim_{\sigma \rightarrow 0^+} \sigma s'(\sigma) = 0$  from Lemma 4. Let  $f(s) = \frac{2\phi(s)}{2 - 2\Phi(s) + \frac{\mu z}{\mu}} - s$ . We can check that there is a unique  $\mathbf{s}^* \in (0, +\infty)$  such that  $f(\mathbf{s}^*) = 0$ . Therefore,

$$\lim_{\sigma \rightarrow 0^+} \widehat{\mathbf{s}} = \mathbf{s}^* \in (0, +\infty).$$

□

We then proceed to prove the propositions.

### Case I: Without a dark pool

By Lemma 4 and Lemma 5,  $\frac{A^{\mathbb{S}}}{\sigma_v}$ ,  $\widehat{\alpha}_e$ ,  $\overline{\gamma}_e^{\mathbb{S}}$ ,  $\underline{\gamma}_e^{\mathbb{S}}$  are differentiable functions of  $\sigma$ , and

$$\lim_{\sigma \rightarrow 0^+} \frac{d\left(\frac{A^{\mathbb{S}}}{\sigma_v}\right)}{d\sigma} = \mathbf{s}^* \in (0, +\infty).$$

Also, taking derivative of  $B(\widehat{\mathbf{s}})$  with respect to  $\sigma$ , we get

$$\frac{dB(\widehat{\mathbf{s}})}{d\sigma} = \frac{\partial B(\widehat{\mathbf{s}})}{\partial \widehat{\mathbf{s}}} \frac{d\widehat{\mathbf{s}}}{d\sigma} + \frac{\partial B(\widehat{\mathbf{s}})}{\partial \sigma} = B(\widehat{\mathbf{s}}) (1 - B(\widehat{\mathbf{s}})) \left( 2\sigma \frac{d\widehat{\mathbf{s}}}{d\sigma} + 2\widehat{\mathbf{s}} \right). \quad (1.60)$$

and the derivative of  $\widehat{\alpha}_e$  is

$$\frac{d\widehat{\alpha}_e}{d\sigma} = -G'(2B(\widehat{\mathbf{s}}) - 1)B(\widehat{\mathbf{s}}) (1 - B(\widehat{\mathbf{s}})) \left( 2\sigma \frac{d\widehat{\mathbf{s}}}{d\sigma} + 2\widehat{\mathbf{s}} \right).$$

When  $\sigma$  is sufficiently small, we get

$$\lim_{\sigma \rightarrow 0^+} \frac{d\widehat{\alpha}_e}{d\sigma} = -\frac{G'(0)\mathbf{s}^*}{2} \in (-\infty, 0).$$

Similarly, we take derivative of  $\overline{\gamma}_e^{\mathbb{S}} - \underline{\gamma}_e^{\mathbb{S}}$  with respect to  $\sigma$  and get

$$\frac{d(\overline{\gamma}_e^{\mathbb{S}} - \underline{\gamma}_e^{\mathbb{S}})}{d\sigma} = [\phi(\widehat{\mathbf{s}} + \sigma) - \phi(\widehat{\mathbf{s}} - \sigma)] \frac{d\widehat{\mathbf{s}}}{d\sigma} + [\phi(\widehat{\mathbf{s}} + \sigma) + \phi(\widehat{\mathbf{s}} - \sigma)],$$

and letting  $\sigma \rightarrow 0^+$ , we have

$$\lim_{\sigma \rightarrow 0^+} \frac{d(\overline{\gamma}_e^{\mathbb{S}} - \underline{\gamma}_e^{\mathbb{S}})}{d\sigma} = 2\phi(\widehat{\mathbf{s}}) \in (0, +\infty).$$

Note that  $\sigma = \frac{\sigma_v}{\sigma_e}$ , we conclude the following:

Given  $\sigma$  sufficiently small, as  $\sigma_v$  increases (or  $\sigma_e$  decreases),

- (i)  $\frac{A^{\mathbb{S}}}{\sigma_v}$  strictly increases.
- (ii)  $\overline{\gamma}_e^{\mathbb{S}} - \underline{\gamma}_e^{\mathbb{S}}$  strictly increases, and  $\alpha_e^{\mathbb{S}}$  strictly decreases.

### Case II, With a dark pool

Note that when  $\sigma = 0$ , we have  $\overline{\gamma}_e = \underline{\gamma}_e$  and  $\overline{\gamma}_d = \underline{\gamma}_d$ . Therefore  $\frac{A}{\sigma_v} = 0$  and  $\bar{R} = \underline{R}$ . Equations (1.42) and (1.43) show that  $B(\mathbf{s}_0) = \frac{1}{2}$  and  $B(\mathbf{s}_1) = \frac{1}{2}$ . If  $0 < \sigma < +\infty$ , we have, by Theorem 2, that  $0 < \mathbf{s}_0 < \mathbf{s}_1 < \infty$ . Therefore, we have  $\overline{\gamma}_e > \underline{\gamma}_e$ ,  $\overline{\gamma}_d > \underline{\gamma}_d$ ,  $\frac{A}{\sigma_v} > 0$ ,  $\bar{R} > \underline{R}$ , and  $\frac{1}{2} < B(\mathbf{s}_0) < B(\mathbf{s}_1) < 1$ . Then we are ready to conclude the following:

Given  $\sigma$  sufficiently small, as  $\sigma_v$  increases (or as  $\sigma_e$  decreases),

- (i)  $\frac{A}{\sigma_v}$  increases, and  $\bar{R} - \underline{R}$  increases.
- (ii)  $\overline{\gamma}_e - \underline{\gamma}_e$ ,  $\overline{\gamma}_d - \underline{\gamma}_d$  increases,  $\alpha_e$  decreases, and  $\alpha_d$  increases.

Let  $(\mathbf{s}_0, \mathbf{s}_1)$  be any equilibrium. Since  $G(\cdot)$ , and  $\Phi(\cdot)$  are twice differentiable, by the implicit function theorem, there exist continuously differentiable functions  $\mathbf{s}_0(\sigma)$ ,  $\mathbf{s}_1(\sigma)$  defined on  $(0, +\infty)$ .

When  $\sigma \in (0, +\infty)$ . By equation (1.42), we have  $B(\mathbf{s}_0) = \frac{\bar{R}}{\bar{R} + \underline{R}} \in (0, 1)$ . Thus rewrite it as

$$\frac{\bar{R}}{\underline{R}} = \frac{1}{\frac{1}{B(\mathbf{s}_0)} - 1},$$

and the derivative can be found as following:

$$\begin{aligned}
\frac{d\left(\frac{\bar{R}}{R}\right)}{d\sigma} &= \frac{1}{\underline{R}^2} \left[ \frac{d\bar{R}}{d\sigma} \underline{R} - \frac{d\underline{R}}{d\sigma} \bar{R} \right] \\
&= \frac{1}{[1 - B(\mathbf{s}_0)]^2} \left( \frac{\partial B(\mathbf{s}_0)}{\partial \mathbf{s}_0} \frac{d\mathbf{s}_0}{d\sigma} + \frac{\partial B(\mathbf{s}_0)}{\partial \sigma} \right) \\
&= \frac{B(\mathbf{s}_0)}{1 - B(\mathbf{s}_0)} \left( 2\sigma \frac{d\mathbf{s}_0}{d\sigma} + 2\mathbf{s}_0 \right).
\end{aligned}$$

Also, we know  $\lim_{\sigma \rightarrow 0^+} B(\mathbf{s}_0) = \frac{1}{2}$  and  $\lim_{\sigma \rightarrow 0^+} \bar{R} = \lim_{\sigma \rightarrow 0^+} \underline{R} = 1$ , thus

$$\lim_{\sigma \rightarrow 0^+} \frac{d\bar{R}}{d\sigma} - \lim_{\sigma \rightarrow 0^+} \frac{d\underline{R}}{d\sigma} = \lim_{\sigma \rightarrow 0^+} \left( 2\sigma \frac{d\mathbf{s}_0}{d\sigma} + 2\mathbf{s}_0 \right).$$

Equation (1.43) shows that

$$2B(\mathbf{s}_1) - 1 - [B(\mathbf{s}_1)\underline{R} - (1 - B(\mathbf{s}_1))\bar{R}] = \frac{A}{\sigma_v}.$$

Taking derivative on both sides, we get

$$\frac{d\left(\frac{A}{\sigma_v}\right)}{d\sigma} = (2 - \underline{R} - \bar{R})B(\mathbf{s}_0) [1 - B(\mathbf{s}_0)] \left( \sigma \frac{d\mathbf{s}_1}{d\sigma} + \mathbf{s}_1 \right) + [1 - B(\mathbf{s}_0)] \frac{d\bar{R}}{d\sigma} - B(\mathbf{s}_0) \frac{d\underline{R}}{d\sigma},$$

and because  $\frac{A}{\sigma_v} = \frac{\Phi(\mathbf{s}_1 + \sigma) - \Phi(\mathbf{s}_1 - \sigma)}{2 - \Phi(\mathbf{s}_1 + \sigma) - \Phi(\mathbf{s}_1 - \sigma) + (1 - G(2B(\mathbf{s}_1) - 1)) \frac{\mu_z}{\mu}}$ , we have

$$\begin{aligned}
&\frac{d\left(\frac{A}{\sigma_v}\right)}{d\sigma} \\
&= \frac{(\phi(\mathbf{s}_1 + \sigma) - \phi(\mathbf{s}_1 - \sigma)) \frac{d\mathbf{s}_1}{d\sigma} + (\phi(\mathbf{s}_1 + \sigma) + \phi(\mathbf{s}_1 - \sigma))}{\bar{\gamma}_e + \underline{\gamma}_e + \alpha_e \frac{\mu_z}{\mu}} \\
&\quad + \frac{[\Phi(\mathbf{s}_1 + \sigma) - \Phi(\mathbf{s}_1 - \sigma)] \left[ (\phi(\mathbf{s}_1 + \sigma) + \phi(\mathbf{s}_1 - \sigma)) \frac{d\mathbf{s}_1}{d\sigma} + (\phi(\mathbf{s}_1 + \sigma) - \phi(\mathbf{s}_1 - \sigma)) \right]}{\left[ \bar{\gamma}_e + \underline{\gamma}_e + \alpha_e \frac{\mu_z}{\mu} \right]^2} \\
&\quad + \frac{2G'(2B(\mathbf{s}_1) - 1) \frac{\mu_z}{\mu} [\Phi(\mathbf{s}_1 + \sigma) - \Phi(\mathbf{s}_1 - \sigma)] \left( \frac{\partial B(\mathbf{s}_1)}{\partial \mathbf{s}_1} \frac{d\mathbf{s}_1}{d\sigma} + \frac{\partial B(\mathbf{s}_1)}{\partial \sigma} \right)}{\left[ \bar{\gamma}_e + \underline{\gamma}_e + \alpha_e \frac{\mu_z}{\mu} \right]^2}.
\end{aligned}$$

Similarly to what we shown in the proof of Lemma 5, we obtain

$$\lim_{\sigma \rightarrow 0^+} \frac{d\left(\frac{A}{\sigma_v}\right)}{d\sigma} = \frac{1}{2} \left( \lim_{\sigma \rightarrow 0^+} \frac{d\bar{R}}{d\sigma} - \lim_{\sigma \rightarrow 0^+} \frac{d\underline{R}}{d\sigma} \right) = \lim_{\sigma \rightarrow 0^+} \left( \sigma \frac{d\mathbf{s}_0}{d\sigma} + \mathbf{s}_0 \right), \quad (1.61)$$

and

$$\lim_{\sigma \rightarrow 0^+} \frac{d\left(\frac{A}{\sigma_v}\right)}{d\sigma} = \lim_{\sigma \rightarrow 0^+} \frac{2\phi(\mathbf{s}_1)}{2 - 2\Phi(\mathbf{s}_1) + \frac{\mu_z}{\mu}}. \quad (1.62)$$

Combing equations (1.61) and (1.62) gives us

$$\lim_{\sigma \rightarrow 0^+} \left( \sigma \frac{d\mathbf{s}_0}{d\sigma} + \mathbf{s}_0 \right) = \lim_{\sigma \rightarrow 0^+} \frac{2\phi(\mathbf{s}_1)}{2 - 2\Phi(\mathbf{s}_1) + \frac{\mu_z}{\mu}}$$

Suppose  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_0 = +\infty$ . Using the similar argument as in the proof of Lemma 5, we obtain  $\lim_{\sigma \rightarrow 0^+} \left( \sigma \frac{d\mathbf{s}_0}{d\sigma} + \mathbf{s}_0 \right) > 0$ . However, as  $\mathbf{s}_0 \rightarrow +\infty$ , we have  $\mathbf{s}_1 \rightarrow +\infty$  and  $\frac{2\phi(\mathbf{s}_1)}{2 - 2\Phi(\mathbf{s}_1) + \frac{\mu_z}{\mu}} \rightarrow 0$ . This is a contradiction. Therefore, it must be that  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_0 < +\infty$ .

By Lemma 4,  $\lim_{\sigma \rightarrow 0^+} \sigma \frac{d\mathbf{s}_0}{d\sigma} = 0$ . So we have

$$\lim_{\sigma \rightarrow 0^+} \mathbf{s}_0 = \lim_{\sigma \rightarrow 0^+} \frac{2\phi(\mathbf{s}_1)}{2 - 2\Phi(\mathbf{s}_1) + \frac{\mu_z}{\mu}}$$

Define  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_0 \triangleq \mathbf{s}_0(0^+)$ ,  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_1 \triangleq \mathbf{s}_1(0^+)$ , and we have

$$\begin{aligned} \lim_{\sigma \rightarrow 0^+} \frac{d\left(\frac{A}{\sigma_v}\right)}{d\sigma} &= \lim_{\sigma \rightarrow 0^+} \mathbf{s}_0 = \mathbf{s}_0(0^+) \geq 0, \\ \lim_{\sigma \rightarrow 0^+} \frac{d(\overline{\gamma_e} - \underline{\gamma_e})}{d\sigma} &= 2\phi(\mathbf{s}_1(0^+)) \geq 0, \\ \lim_{\sigma \rightarrow 0^+} \frac{d\alpha_e}{d\sigma} &= -\frac{G'(0)\mathbf{s}_1(0^+)}{2} \leq 0, \\ \lim_{\sigma \rightarrow 0^+} \frac{d\alpha_d}{d\sigma} &= \frac{G'(0)(\mathbf{s}_1(0^+) - \mathbf{s}_0(0^+))}{2} \geq 0, \\ \lim_{\sigma \rightarrow 0^+} \frac{d(\alpha_e + \alpha_d)}{d\sigma} &= -\frac{G'(0)\mathbf{s}_0(0^+)}{2} \leq 0, \end{aligned}$$

which conclude the proof. □

### 1.8.6 Proof of Proposition 3

To prove Proposition 3, we need the following lemmas.

**Lemma 6.** For any given  $\sigma \in (0, +\infty)$ ,  $\widehat{\mathbf{s}}(\sigma) < \mathbf{s}_1(\sigma)$ .

**Proof.** Substitute the expressions of  $\frac{A}{\sigma_v}$  into equation (1.40) and (1.43), then  $\widehat{\mathbf{s}}$ ,  $\mathbf{s}_1$  are respectively determined by the following two equations

$$\begin{aligned} \frac{\Phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) - \Phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma})}{2 - \Phi(\widehat{\mathbf{s}} + \boldsymbol{\sigma}) - \Phi(\widehat{\mathbf{s}} - \boldsymbol{\sigma}) + (1 - G(2B(\widehat{\mathbf{s}}) - 1))\frac{\mu_z}{\mu}} &= 2B(\widehat{\mathbf{s}}) - 1, \\ \frac{\Phi(\mathbf{s}_1 + \boldsymbol{\sigma}) - \Phi(\mathbf{s}_1 - \boldsymbol{\sigma})}{2 - \Phi(\mathbf{s}_1 + \boldsymbol{\sigma}) - \Phi(\mathbf{s}_1 - \boldsymbol{\sigma}) + (1 - G(2B(\mathbf{s}_1) - 1))\frac{\mu_z}{\mu}} &= 2B(\mathbf{s}_1) - 1 \\ &\quad - [B(\mathbf{s}_1)\underline{R} - (1 - B(\mathbf{s}_1))\bar{R}]. \end{aligned}$$

Let  $f(s) = \frac{\Phi(s+\boldsymbol{\sigma}) - \Phi(s-\boldsymbol{\sigma})}{2 - \Phi(s+\boldsymbol{\sigma}) - \Phi(s-\boldsymbol{\sigma}) + (1 - G(2B(s) - 1))\frac{\mu_z}{\mu}}$ , and its derivative is

$$f'(s) = \frac{D_1(s) + D_2(s)}{\left[2 - \Phi(s + \boldsymbol{\sigma}) - \Phi(s - \boldsymbol{\sigma}) + (1 - G(2B(s) - 1))\frac{\mu_z}{\mu}\right]^2},$$

where

$$\begin{aligned} D_1(s) &= (\phi(s + \boldsymbol{\sigma}) - \phi(s - \boldsymbol{\sigma})) \left(2 - \Phi(s + \boldsymbol{\sigma}) - \Phi(s - \boldsymbol{\sigma}) + (1 - G(2B(s) - 1))\frac{\mu_z}{\mu}\right) < 0, \\ D_2(s) &= -(\Phi(s + \boldsymbol{\sigma}) - \Phi(s - \boldsymbol{\sigma})) \left(-\phi(s + \boldsymbol{\sigma}) - \phi(s - \boldsymbol{\sigma}) - 2G'(2B(s) - 1)B'(s)\frac{\mu_z}{\mu}\right) > 0. \end{aligned}$$

Since  $G'(s) + sG''(s) \geq 0$ , one can represent  $f(s)$  as the blue curve in Figure 1.16.

Let  $\widehat{h}(s) = 2B(s) - 1$  and  $h(s) = 2B(s) - 1 - [B(s)\underline{R} - (1 - B(s))\bar{R}]$ . By equation (1.42), for any  $s > \mathbf{s}_0$ , we have  $B(s) > B(\mathbf{s}_0) = \frac{\bar{R}}{\underline{R} + \bar{R}}$ . That is,  $[B(s)\underline{R} - (1 - B(s))\bar{R}] > 0$ . Therefore  $\widehat{h}(s) > h(s)$ . In Figure 1.16,  $\widehat{h}(s)$  is represented by the red curve, while  $h(s)$  is represented by the green curve which is below  $\widehat{h}(s)$ . Obviously, the intersection point  $\mathbf{s}_1$  is larger than  $\widehat{\mathbf{s}}$ . The Lemma is proved.  $\square$

**Lemma 7.** If  $\boldsymbol{\sigma} \rightarrow +\infty$ , there exists a unique  $\widehat{k} \in (\frac{1}{2}, 1)$  such that  $\lim_{\boldsymbol{\sigma} \rightarrow +\infty} \overline{\gamma}_e^{\mathbb{S}} = 1$ ,  $\lim_{\boldsymbol{\sigma} \rightarrow +\infty} \underline{\gamma}_e^{\mathbb{S}} = 0$ ,  $\lim_{\boldsymbol{\sigma} \rightarrow +\infty} \alpha_e^{\mathbb{S}} = 1 - G(\widehat{k})$ , and  $\lim_{\boldsymbol{\sigma} \rightarrow +\infty} \frac{A}{\sigma_v} = \widehat{k}$ , where  $\widehat{k}$  is determined by

$$\widehat{k} = \frac{1}{1 + \left[1 - G(\widehat{k})\right]\frac{\mu_z}{\mu}}. \quad (1.63)$$

In addition, such  $\widehat{k}$  is smaller if  $\frac{\mu_z}{\mu}$  is larger.



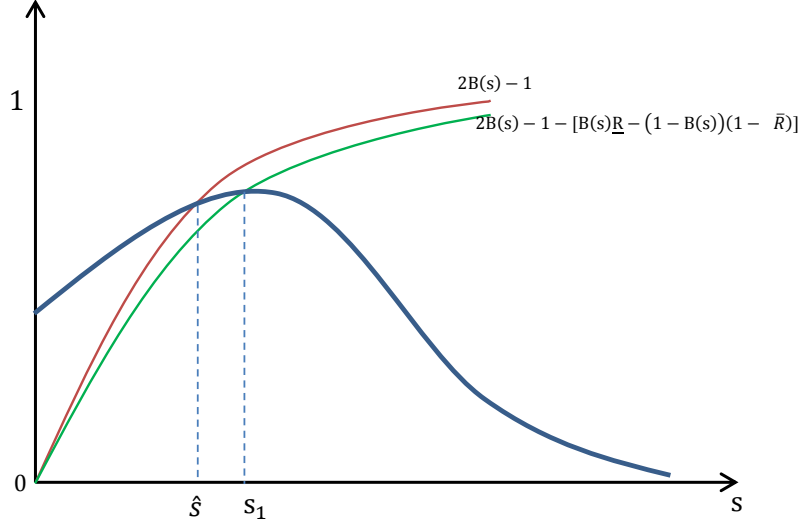


Figure 1.16:  $h(s)$  and  $\hat{h}(s)$

**Proof.** Suppose  $\lim_{\sigma \rightarrow +\infty} \hat{\mathbf{s}}\sigma = +\infty$ . Then, when  $\sigma \rightarrow +\infty$ , we have  $2B(\hat{\mathbf{s}}) - 1 \rightarrow 1$ . Thus equation (1.40) gives us that  $\frac{A^S}{\sigma_v} = 1$ . However  $\hat{\alpha}_e = 1 - G(1) > 0$ , which implies  $\frac{A^S}{\sigma_v} < 1$ . Therefore, we have

$$\lim_{\sigma \rightarrow +\infty} \hat{\mathbf{s}}\sigma < +\infty.$$

Let  $\lim_{\sigma \rightarrow +\infty} \hat{\mathbf{s}}\sigma = \hat{C} \in [0, +\infty)$ , where  $\hat{C}$  will be determined later. Then we have  $\lim_{\sigma \rightarrow +\infty} \hat{\mathbf{s}} = 0$ , thus  $\lim_{\sigma \rightarrow +\infty} (\hat{\mathbf{s}} - \sigma) = -\infty$ . Therefore,  $\lim_{\sigma \rightarrow +\infty} \bar{\gamma}_e^S = 1$ ,  $\lim_{\sigma \rightarrow +\infty} \gamma_e^S = 0$ . Let  $\hat{k} = \lim_{\sigma \rightarrow +\infty} 2B(\hat{\mathbf{s}}) - 1 = \frac{1 - e^{-2\hat{C}}}{1 + e^{-2\hat{C}}}$ , and we have  $\lim_{\sigma \rightarrow +\infty} \alpha_e^S = 1 - G(\hat{k})$  and  $\lim_{\sigma \rightarrow +\infty} \frac{A^S}{\sigma_v} = \frac{1}{1 + [1 - G(\hat{k})]^{\frac{\mu_z}{\mu}}}$ . However,  $\hat{k}$  has to satisfy equation (1.63) such that equation (1.40) is satisfied.

Let  $f(k) = k - \frac{1}{1 + [1 - G(k)]^{\frac{\mu_z}{\mu}}}$ , and we can easily verify that  $f(0) < 0$ , and  $f(1) > 0$ . Therefore, there exists a  $\hat{k} \in (0, 1)$  such that  $f(\hat{k}) = 0$ , and  $\hat{C} = \frac{1}{2} \ln \frac{1 + \hat{k}}{1 - \hat{k}}$ .  $\square$

**Lemma 8.** Let  $R = \mathbb{E} \left[ \min \left\{ 1, \frac{Z^+}{Z^-} \right\} \right]$ . Consider any equilibrium  $\mathbf{s}_0(\sigma), \mathbf{s}_1(\sigma)$  for  $\sigma \rightarrow +\infty$ . We have  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_0\sigma < +\infty$ . In addition, the limits of variables can be determined in the following two statements.

(i) If  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_1\sigma < +\infty$ , we have  $\lim_{\sigma \rightarrow +\infty} \bar{\gamma}_e = 1$ ,  $\lim_{\sigma \rightarrow +\infty} \gamma_e = 0$ ,  $\lim_{\sigma \rightarrow +\infty} \bar{\gamma}_d = 0$ ,  $\lim_{\sigma \rightarrow +\infty} \gamma_d = 0$ ,  $\lim_{\sigma \rightarrow +\infty} \alpha_e = 1 - G(k_1)$ ,  $\lim_{\sigma \rightarrow +\infty} \alpha_d = G(k_1)$ ,  $\lim_{\sigma \rightarrow +\infty} \frac{A}{\sigma_v} = \frac{1}{1 + [1 - G(k_1)]^{\frac{\mu_z}{\mu}}}$ ,  $\lim_{\sigma \rightarrow +\infty} \bar{R} = R$ , and

$\lim_{\sigma \rightarrow +\infty} \underline{R} = R$ , where  $k_1 \in (\frac{1}{2}, 1)$  is determined by

$$(1 - R)k_1 = \frac{1}{1 + [1 - G(k_1)] \frac{\mu_z}{\mu}}. \quad (1.64)$$

(ii) If  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_0 \sigma = +\infty$ , we have  $\lim_{\sigma \rightarrow +\infty} \overline{\gamma}_e = 1 - k_3$ ,  $\lim_{\sigma \rightarrow +\infty} \underline{\gamma}_e = 0$ ,  $\lim_{\sigma \rightarrow +\infty} \overline{\gamma}_d = k_3$ ,  $\lim_{\sigma \rightarrow +\infty} \underline{\gamma}_d = 0$ ,  $\lim_{\sigma \rightarrow +\infty} \alpha_e = 1 - G(1)$ ,  $\lim_{\sigma \rightarrow +\infty} \alpha_d = G(1) - G(2k_2 - 1)$ ,  $\lim_{\sigma \rightarrow +\infty} \frac{A}{\sigma_v} = \frac{1 - k_3}{1 - k_3 + [1 - G(1)] \frac{\mu_z}{\mu}}$ ,  $\lim_{\sigma \rightarrow +\infty} \bar{R} = \frac{k_2}{1 - k_2} \frac{[1 - G(1)] \frac{\mu_z}{\mu}}{1 - k_3 + [1 - G(1)] \frac{\mu_z}{\mu}}$ , and  $\lim_{\sigma \rightarrow +\infty} \underline{R} = \frac{[1 - G(1)] \frac{\mu_z}{\mu}}{1 - k_3 + [1 - G(1)] \frac{\mu_z}{\mu}}$ , where  $k_2 \in [\frac{1}{2}, 1)$  and  $k_3 \in [0, 1)$  are determined by

$$\frac{[1 - G(1)] \frac{\mu_z}{\mu}}{1 - k_3 + [1 - G(1)] \frac{\mu_z}{\mu}} = \mathbb{E} \left[ \min \left\{ 1, \frac{Z^-}{\frac{k_3}{G(1) - G(2k_2 - 1)} + Z^+} \right\} \right], \quad (1.65)$$

$$k_2 = \frac{\mathbb{E} \left[ \min \left\{ 1, \frac{k_3}{\frac{G(1) - G(2k_2 - 1)}{Z^-} + Z^+} \right\} \right]}{\mathbb{E} \left[ \min \left\{ 1, \frac{k_3}{\frac{G(1) - G(2k_2 - 1)}{Z^-} + Z^+} \right\} \right] + \mathbb{E} \left[ \min \left\{ 1, \frac{Z^-}{\frac{k_3}{G(1) - G(2k_2 - 1)} + Z^+} \right\} \right]}. \quad (1.66)$$

**Proof.** Consider any continuously differentiable functions  $\mathbf{s}_0(\sigma), \mathbf{s}_1(\sigma)$ .

First we show  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_0 \sigma < +\infty$  by contradiction. Suppose that  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_0 \sigma = +\infty$ , we have  $B(\mathbf{s}_0) = \frac{1}{1 + e^{-2\mathbf{s}_0 \sigma}} \rightarrow 1$ . Since  $\mathbf{s}_1 > \mathbf{s}_0$ , we have  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_1 \sigma = +\infty$ ,  $B(\mathbf{s}_1) = \frac{1}{1 + e^{-2\mathbf{s}_1 \sigma}} \rightarrow 1$ . In addition, Equation (1.42) gives us that  $\frac{\bar{R}}{R} = 1$ , i.e.,  $\bar{R} = R$ .

If  $\lim_{\sigma \rightarrow +\infty} (\mathbf{s}_0 - \sigma) < +\infty$ , then  $\overline{\gamma}_d > 0 = \underline{\gamma}_d$ , which is a contradiction to  $\bar{R} = R$ . If  $\lim_{\sigma \rightarrow +\infty} (\mathbf{s}_0 - \sigma) = +\infty$ , then  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_1 - \sigma = +\infty$ . Therefore we have  $\lim_{\sigma \rightarrow +\infty} \overline{\gamma}_e = \lim_{\sigma \rightarrow +\infty} \underline{\gamma}_e = 0$ ,  $\lim_{\sigma \rightarrow +\infty} \frac{A}{\sigma_v} = 0$ , and by equation (1.43), we have  $\lim_{\sigma \rightarrow +\infty} B(\mathbf{s}_1) = \lim_{\sigma \rightarrow +\infty} \frac{1 - \bar{R}}{1 - \bar{R} + 1 - R} = \frac{1}{2}$ , whche is a contradiction to  $\lim_{\sigma \rightarrow +\infty} B(\mathbf{s}_1) = 1$ . Therefore, we have

$$\lim_{\sigma \rightarrow +\infty} \mathbf{s}_0 \sigma = C \in [0, +\infty).$$

Then we show the two statements.

(i) Suppose that  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_0 \sigma = C_0 \in [0, +\infty)$  and  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_1 \sigma = C_1 \in [0, +\infty)$ , then we have  $\lim_{\sigma \rightarrow +\infty} (\mathbf{s}_0 - \sigma) \rightarrow -\infty$  and  $\lim_{\sigma \rightarrow +\infty} (\mathbf{s}_1 - \sigma) \rightarrow -\infty$ . Therefore,  $\lim_{\sigma \rightarrow +\infty} \overline{\gamma}_e = 1$ ,  $\lim_{\sigma \rightarrow +\infty} \underline{\gamma}_e = \lim_{\sigma \rightarrow +\infty} \overline{\gamma}_d = \lim_{\sigma \rightarrow +\infty} \underline{\gamma}_d = 0$ .

We show that  $\lim_{\sigma \rightarrow +\infty} \bar{R} = \lim_{\sigma \rightarrow +\infty} \underline{R} = R$ . If  $C_0 = C_1$ , Lemma 3 and  $\lim_{\sigma \rightarrow +\infty} \frac{\phi(\mathbf{s}_0 - \sigma)}{B'(\mathbf{s}_0)} = \lim_{\sigma \rightarrow +\infty} \frac{\phi(\mathbf{s}_0 + \sigma)}{B'(\mathbf{s}_0)} = 0$  give us that  $\lim_{\sigma \rightarrow +\infty} \bar{R} = \lim_{\sigma \rightarrow +\infty} \underline{R} = \mathbb{E} \left[ \min \left\{ 1, \frac{Z^+}{Z^-} \right\} \right] = R$ . If  $C_0 < C_1$ , because  $\lim_{\sigma \rightarrow +\infty} \bar{\gamma}_d = \lim_{\sigma \rightarrow +\infty} \underline{\gamma}_d = 0$  and  $\lim_{\sigma \rightarrow +\infty} \alpha_d > 0$ , we have  $\lim_{\sigma \rightarrow +\infty} \bar{R} = \lim_{\sigma \rightarrow +\infty} \underline{R} = \mathbb{E} \left[ \min \left\{ 1, \frac{Z^+}{Z^-} \right\} \right] = R$ .

Then equation (1.42) gives us that  $\lim_{\sigma \rightarrow +\infty} B(\mathbf{s}_0) = 1/2$  and  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_0 \sigma = 0$ . Let  $k_1 = \lim_{\sigma \rightarrow +\infty} 2B(\mathbf{s}_1) - 1$ , and we have  $\lim_{\sigma \rightarrow +\infty} \alpha_e = 1 - G(k_1)$ ,  $\lim_{\sigma \rightarrow +\infty} \alpha_d = G(k_1)$  and  $\lim_{\sigma \rightarrow +\infty} \frac{A}{\sigma^v} = \frac{1}{1 + [1 - G(k_1)] \frac{\mu_z}{\mu}}$ . Rewrite equation (1.43) in the following form

$$(2B(\mathbf{s}_1) - 1)(1 - R) = \frac{1}{1 + [1 - G(2B(\mathbf{s}_1) - 1)] \frac{\mu_z}{\mu}},$$

and  $k_1$  has to satisfy equation (1.64).

Let  $f(k) = (1 - R)k - \frac{1}{1 + [1 - G(k)] \frac{\mu_z}{\mu}}$ . We can verify that  $f(0) < 0$  and  $f(1) > 0$  if  $1 + [1 - G(1)] \frac{\mu_z}{\mu} > \frac{1}{1 - R}$ . There is a  $k_1 \in (0, 1)$  such that  $f(k_1) = 0$ , and  $C_1 = \frac{1}{2} \ln \frac{1 + k_1}{1 - k_1}$ .

(ii) Suppose that  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_0 \sigma = C_2 \in [0, +\infty)$  and  $\lim_{\sigma \rightarrow +\infty} \mathbf{s}_1 \sigma = +\infty$ . We have  $\lim_{\sigma \rightarrow +\infty} \bar{\gamma}_e = 0$ ,  $\lim_{\sigma \rightarrow +\infty} \underline{\gamma}_d = 0$ , and  $\lim_{\sigma \rightarrow +\infty} \alpha_e = 1 - G(1)$ .

Suppose that  $\lim_{\sigma \rightarrow +\infty} (\mathbf{s}_1 - \sigma) = C_3 \in [-\infty, +\infty]$ . Let  $k_2 = \lim_{\sigma \rightarrow +\infty} B(\mathbf{s}_0) = \frac{1}{1 + e^{-2C_2}} \in [\frac{1}{2}, 1)$  and  $k_3 = \lim_{\sigma \rightarrow +\infty} \bar{\gamma}_d = \Phi(C_3) \in [0, 1]$ . Then we have  $\lim_{\sigma \rightarrow +\infty} \bar{\gamma}_e = 1 - k_3$ ,  $\lim_{\sigma \rightarrow +\infty} \underline{\gamma}_d = k_3$ ,  $\lim_{\sigma \rightarrow +\infty} \alpha_d = G(1) - G(2k_2 - 1)$ ,  $\lim_{\sigma \rightarrow +\infty} \frac{A}{\sigma^v} = \frac{1 - k_3}{1 - k_3 + [1 - G(1)] \frac{\mu_z}{\mu}}$ . Combining equations (1.42) and (1.43), we have  $\lim_{\sigma \rightarrow +\infty} \bar{R} = \frac{k_2}{1 - k_2} \frac{[1 - G(1)] \frac{\mu_z}{\mu}}{1 - k_3 + [1 - G(1)] \frac{\mu_z}{\mu}}$ ,  $\lim_{\sigma \rightarrow +\infty} \underline{R} = \frac{[1 - G(1)] \frac{\mu_z}{\mu}}{1 - k_3 + [1 - G(1)] \frac{\mu_z}{\mu}}$ . In addition, by equations (1.44) and (1.45),  $k_2$  and  $k_3$  have to satisfy equations (1.65) and (1.66).

Suppose that  $1 + [1 - G(1)] \frac{\mu_z}{\mu} \leq \frac{1}{1 - R}$ . For equation (1.65), the left hand side is increasing with respect to  $k_3$ , while the right hand side is decreasing with respect to  $k_3$ . In addition, when  $k_3 = 0$ ,  $LHS - RHS = \frac{[1 - G(1)] \frac{\mu_z}{\mu}}{1 + [1 - G(1)] \frac{\mu_z}{\mu}} - \mathbb{E} \left[ \min \left\{ 1, \frac{Z^-}{Z^+} \right\} \right] = 1 - R - \frac{1}{1 + [1 - G(1)] \frac{\mu_z}{\mu}} \leq 0$ , and when  $k_3 = 1$ ,  $LHS - RHS = 1 - \mathbb{E} \left[ \min \left\{ 1, \frac{Z^-}{G(1) - G(2k_2 - 1) + Z^+} \right\} \right] > 0$ . Thus, given any  $k_2 \in [\frac{1}{2}, 1)$ , there exists a unique  $k_3(k_2) \in (0, 1)$  that solves equation (1.65). Furthermore, as  $k_2$  increases, the right hand side of equation (1.65) decreases, thus  $k_3(k_2)$  is decreasing with respect to  $k_2$ . When  $k_2 \rightarrow 1$ , we have  $k_3(k_2) \rightarrow 0$ . Thus  $\underline{R} \rightarrow \frac{[1 - G(1)] \frac{\mu_z}{\mu}}{1 + [1 - G(1)] \frac{\mu_z}{\mu}} \geq 0$ .

For equation (1.66), we substitute  $k_3$  with the expression solved from (1.65), and it becomes a function of  $k_2$  only. When  $k_2 = 1/2$ , we have  $k_3 \in [0, 1)$ . Then  $LHS - RHS \leq 0$ . While when  $k_2 \rightarrow 1$ , we have  $LHS - RHS \geq 0$ . Therefore, there exist  $k_2 \in [1/2, 1)$  and  $k_3 \in [0, 1)$  such that equations (1.65) and (1.66) are satisfied. Additionally, we have  $C_2 = \frac{1}{2} \ln \frac{k_2}{1-k_2}$ ,  $C_3 = \Phi^{-1}(k_3)$ .

□

We now proceed to prove the proposition. From Lemma 6, we have  $\widehat{\mathbf{s}} < \mathbf{s}_1$  for all  $\sigma \in (0, +\infty)$ . Thus,  $\overline{\gamma}_e^{\mathbb{S}} - \underline{\gamma}_e^{\mathbb{S}} = \Phi(\widehat{\mathbf{s}} + \sigma) - \Phi(\widehat{\mathbf{s}} - \sigma) > \Phi(\mathbf{s}_1 + \sigma) - \Phi(\mathbf{s}_1 - \sigma) = \overline{\gamma}_e - \underline{\gamma}_e$  and  $\alpha_e^{\mathbb{S}} = 1 - G(2B(\widehat{\mathbf{s}}) - 1) > 1 - G(2B(\mathbf{s}_1) - 1) = \alpha_e$ .

Let  $\widehat{k}$ ,  $k_1$ ,  $k_2$ ,  $k_3$  as in (1.63), (1.64), (1.65), (1.66). Suppose  $1 - R > \frac{1}{1+[1-G(\widehat{k})] \frac{\mu_z}{\mu}}$ , then as  $\sigma \rightarrow +\infty$ , by Lemma 7, we have

$$\lim_{\sigma \rightarrow +\infty} \frac{A^{\mathbb{S}}}{\sigma_v} = \frac{1}{1 + \left[1 - G(\widehat{k})\right] \frac{\mu_z}{\mu}},$$

$$\lim_{\sigma \rightarrow +\infty} \alpha_e^{\mathbb{S}} = 1 - G(\widehat{k}).$$

By Lemma 8(i),  $\lim_{\sigma \rightarrow +\infty} \frac{A}{\sigma_v} = \frac{1}{1+[1-G(k_1)] \frac{\mu_z}{\mu}}$  and  $\lim_{\sigma \rightarrow +\infty} \alpha_e = 1 - G(k_1)$ . We can verify that  $\widehat{k} < k_1$  from equations (1.63) and (1.64). Therefore  $\frac{1}{1+[1-G(\widehat{k})] \frac{\mu_z}{\mu}} < \frac{1}{1+[1-G(k_1)] \frac{\mu_z}{\mu}}$  and  $1 - G(\widehat{k}) > 1 - G(k_1)$ . That is,  $\lim_{\sigma \rightarrow +\infty} \frac{A^{\mathbb{S}}}{\sigma_v} < \lim_{\sigma \rightarrow +\infty} \frac{A}{\sigma_v}$ . We can easily verify that  $\lim_{\sigma \rightarrow +\infty} \alpha_e^{\mathbb{S}} < \lim_{\sigma \rightarrow +\infty} \alpha_e + \alpha_d$ .

By Lemma 8(ii),  $\lim_{\sigma \rightarrow +\infty} \frac{A}{\sigma_v} = \frac{1-k_3}{1-k_3+[1-G(1)] \frac{\mu_z}{\mu}}$  and  $\lim_{\sigma \rightarrow +\infty} \alpha_e = 1 - G(1)$ . Then by equation (1.65),  $\frac{1-k_3}{1-k_3+[1-G(1)] \frac{\mu_z}{\mu}} = 1 - \mathbb{E} \left[ \min \left\{ 1, \frac{Z^-}{\frac{k_3}{G(1)-G(2k_2-1)} + Z^+} \right\} \right] > 1 - \mathbb{E} \left[ \min \left\{ 1, \frac{Z^-}{Z^+} \right\} \right] = 1 - R$ . Since we suppose that  $1 - R > \frac{1}{1+[1-G(\widehat{k})] \frac{\mu_z}{\mu}}$ , we have that  $\frac{1}{1+[1-G(\widehat{k})] \frac{\mu_z}{\mu}} < \frac{1-k_3}{1-k_3+[1-G(1)] \frac{\mu_z}{\mu}}$ , that is,  $\lim_{\sigma \rightarrow +\infty} \frac{A^{\mathbb{S}}}{\sigma_v} < \lim_{\sigma \rightarrow +\infty} \frac{A}{\sigma_v}$ .

Since  $\widehat{k} < 1$ ,  $k_3 > 0$ , we proved that  $\lim_{\sigma \rightarrow +\infty} \overline{\gamma}_e^{\mathbb{S}} - \underline{\gamma}_e^{\mathbb{S}} \leq \lim_{\sigma \rightarrow +\infty} \overline{\gamma}_e - \underline{\gamma}_e$ ,  $\lim_{\sigma \rightarrow +\infty} \alpha_e^{\mathbb{S}} \geq \lim_{\sigma \rightarrow +\infty} \alpha_e$ .

Next we consider the case when  $\sigma \rightarrow 0^+$ . Recall that when  $\sigma = 0$ , we have  $\frac{A^{\mathbb{S}}}{\sigma_v} = \frac{A}{\sigma_v} = 0$ . So we have to compare their derivatives at 0. From the proof of Lemma 5, we have that

$\lim_{\sigma \rightarrow 0^+} \widehat{\mathbf{s}} = \lim_{\sigma \rightarrow 0^+} \frac{2\phi(\widehat{\mathbf{s}})}{2-2\Phi(\widehat{\mathbf{s}}) + \frac{\mu_z}{\mu}}$ , and  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_0 = \lim_{\sigma \rightarrow 0^+} \frac{2\phi(\mathbf{s}_1)}{2-2\Phi(\mathbf{s}_1) + \frac{\mu_z}{\mu}}$ . Since  $\frac{2\phi(s)}{2-2\Phi(s) + \frac{\mu_z}{\mu}}$  decreases in  $s$ , we

show that either  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_0 < \lim_{\sigma \rightarrow 0^+} \widehat{\mathbf{s}} < \lim_{\sigma \rightarrow 0^+} \mathbf{s}_1$ , or  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_0 = \lim_{\sigma \rightarrow 0^+} \widehat{\mathbf{s}} = \lim_{\sigma \rightarrow 0^+} \mathbf{s}_1$ . Therefore we have two cases to consider. (i)  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_0 < \lim_{\sigma \rightarrow 0^+} \widehat{\mathbf{s}} < \lim_{\sigma \rightarrow 0^+} \mathbf{s}_1$ . Since  $\frac{dA}{d\sigma}$  increases in  $s$  when  $\sigma \rightarrow 0^+$ , we have that  $\frac{A^{\mathbb{S}}}{\sigma_v} < \frac{A}{\sigma_v}$ , as  $\sigma \rightarrow 0^+$ . (ii)  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_0 = \lim_{\sigma \rightarrow 0^+} \widehat{\mathbf{s}} = \lim_{\sigma \rightarrow 0^+} \mathbf{s}_1$ . In this case  $\frac{dA^{\mathbb{S}}}{d\sigma} = \frac{dA}{d\sigma}$ , it is undetermined whether  $\frac{A^{\mathbb{S}}}{\sigma_v} < \frac{A}{\sigma_v}$  or  $\frac{A^{\mathbb{S}}}{\sigma_v} > \frac{A}{\sigma_v}$ , as  $\sigma \rightarrow 0^+$ . However, we cannot distinguish between case (i) and case (ii).

### 1.8.7 Proof of Proposition 4

As  $\sigma \rightarrow +\infty$ , we have  $\lim_{\sigma \rightarrow +\infty} \overline{\gamma_e^{\mathbb{S}}} - \underline{\gamma_e^{\mathbb{S}}} = 1$ , and  $\lim_{\sigma \rightarrow +\infty} \alpha_e^{\mathbb{S}} = 1 - G(\widehat{k})$ . We consider the two case in Lemma 8: (i) We have  $\lim_{\sigma \rightarrow +\infty} \overline{\gamma_e} - \underline{\gamma_e} = 1$  and  $\lim_{\sigma \rightarrow +\infty} \alpha_e = 1 - G(k_1)$ . Thus  $\lim_{\sigma \rightarrow +\infty} \frac{\overline{\gamma_e^{\mathbb{S}}} - \underline{\gamma_e^{\mathbb{S}}}}{\alpha_e^{\mathbb{S}}} \leq \lim_{\sigma \rightarrow +\infty} \frac{\overline{\gamma_e} - \underline{\gamma_e}}{\alpha_e}$  because  $\widehat{k} < k_1$ . (ii) We have  $\lim_{\sigma \rightarrow +\infty} \overline{\gamma_e} - \underline{\gamma_e} = 1 - k_3$  and  $\lim_{\sigma \rightarrow +\infty} \alpha_e = 1 - G(1)$ . From  $\frac{1}{1+[1-G(\widehat{k})]\frac{\mu_z}{\mu}} < \frac{1-k_3}{1-k_3+[1-G(1)]\frac{\mu_z}{\mu}}$ , we have  $\frac{1}{1-G(\widehat{k})} < \frac{1-k_3}{1-G(1)}$ , i.e., Thus  $\lim_{\sigma \rightarrow +\infty} \frac{\overline{\gamma_e^{\mathbb{S}}} - \underline{\gamma_e^{\mathbb{S}}}}{\alpha_e^{\mathbb{S}}} \leq \lim_{\sigma \rightarrow +\infty} \frac{\overline{\gamma_e} - \underline{\gamma_e}}{\alpha_e}$ .

As  $\sigma \rightarrow 0^+$ , by Lemma 6, we have  $\widehat{\mathbf{s}} < \mathbf{s}_1$ ,  $\forall \sigma > 0$ . Since  $\frac{\overline{\gamma_e^{\mathbb{S}}} - \underline{\gamma_e^{\mathbb{S}}}}{\alpha_e^{\mathbb{S}}} = \frac{\Phi(\widehat{\mathbf{s}} + \sigma) - \Phi(\widehat{\mathbf{s}} - \sigma)}{1 - G(2B(\widehat{\mathbf{s}}) - 1)}$  and  $\frac{\overline{\gamma_e} - \underline{\gamma_e}}{\alpha_e} = \frac{\Phi(\mathbf{s}_1 + \sigma) - \Phi(\mathbf{s}_1 - \sigma)}{1 - G(2B(\mathbf{s}_1) - 1)}$ , to show that  $\frac{\overline{\gamma_e^{\mathbb{S}}} - \underline{\gamma_e^{\mathbb{S}}}}{\alpha_e^{\mathbb{S}}} > \frac{\overline{\gamma_e} - \underline{\gamma_e}}{\alpha_e}$  for small  $\sigma$ , it is sufficient to show that there exists  $\bar{\sigma} > 0$ , s.t.  $\forall \sigma \in (0, \bar{\sigma})$ ,  $\frac{\Phi(s + \sigma) - \Phi(s - \sigma)}{1 - G(2B(s) - 1)}$  decreases in  $s$ . If  $\frac{\mu_z}{\mu} < +\infty$ , Lemma 5 gives us  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_1 \geq \lim_{\sigma \rightarrow 0^+} \widehat{\mathbf{s}} > 0$ . Also, recall that  $\lim_{\sigma \rightarrow 0^+} \mathbf{s}_1 \sigma = \lim_{\sigma \rightarrow 0^+} \widehat{\mathbf{s}} \sigma = 0$ .

We now consider the derivative of  $\frac{\Phi(s + \sigma) - \Phi(s - \sigma)}{1 - G(2B(s) - 1)}$  with respect to  $s$ .

$$\begin{aligned} \frac{d \left( \frac{\Phi(s + \sigma) - \Phi(s - \sigma)}{1 - G(2B(s) - 1)} \right)}{ds} &= \frac{(\Phi(s + \sigma) - \Phi(s - \sigma)) G'(2B(s) - 1) \frac{e^{-2s\sigma}}{(1 + e^{-2s\sigma})^2} 4\sigma}{(1 - G(2B(s) - 1))^2} \\ &\quad + \frac{(\phi(s + \sigma) - \phi(s - \sigma)) (1 - G(2B(s) - 1))}{(1 - G(2B(s) - 1))^2}. \end{aligned} \quad (1.67)$$

Let  $M = \max_{s \in [0, +\infty]} \left[ \frac{4G'(2B(s) - 1)}{1 - G(2B(s) - 1)} + 1 \right] < +\infty$ . If  $\lim_{\sigma \rightarrow 0^+} s > 0$  and  $\lim_{\sigma \rightarrow 0^+} s\sigma = 0$ , there exists

$\bar{\sigma} > 0$ , such that  $\forall \sigma \in (0, \bar{\sigma})$ ,  $s > Me^2\sigma$  and  $s\sigma < 1$ . Therefore by the mean value theorem,

$$\begin{aligned} \frac{d\left(\frac{\Phi(s+\sigma)-\Phi(s-\sigma)}{1-G(2B(s)-1)}\right)}{ds} &< \frac{2\sigma [\phi(s-\sigma)4G'(2B(s)-1)\sigma + \phi(s+\sigma)(1-G(2B(s)-1))(-s-\sigma)]}{(1-G(2B(s)-1))^2} \\ &< \frac{2\sigma \left\{ \phi(s-\sigma) \left[ \frac{4G'(2B(s)-1)}{1-G(2B(s)-1)} + 1 \right] \sigma + \phi(s+\sigma)(-s) \right\}}{1-G(2B(s)-1)} \\ &= \frac{2\sigma \{ \phi(s+\sigma)Me^{2s\sigma}\sigma - \phi(s+\sigma)s \}}{1-G(2B(s)-1)} \\ &< 0. \end{aligned}$$

Thus  $\exists \bar{\sigma} > 0$ , such that  $\forall \sigma \in (0, \bar{\sigma})$ ,  $s \in [\hat{\mathbf{s}}, \mathbf{s}_1]$ , we have  $\frac{d\left(\frac{\Phi(s+\sigma)-\Phi(s-\sigma)}{1-G(2B(s)-1)}\right)}{ds} < 0$ . Since  $\hat{\mathbf{s}} < \mathbf{s}_1$ , we have  $\frac{\Phi(\hat{\mathbf{s}}+\sigma)-\Phi(\hat{\mathbf{s}}-\sigma)}{1-G(2B(\hat{\mathbf{s}})-1)} > \frac{\Phi(\mathbf{s}_1+\sigma)-\Phi(\mathbf{s}_1-\sigma)}{1-G(2B(\mathbf{s}_1)-1)}$ . We proved the proposition.  $\square$

### 1.8.8 Proof of Proposition 5

We need the follow Lemma to proceed the proof.

**Lemma 9.**  $\forall \sigma > 0, \exists C(\sigma) > 0$ , such that  $\forall 0 \leq s \leq C(\sigma)$ ,  $\frac{d\left(\frac{\Phi(s+\sigma)-\Phi(s-\sigma)}{1-G(2B(s)-1)}\right)}{ds} > 0$ .

**Proof.** Consider (1.67),  $\forall \sigma > 0$ ,  $\left. \frac{d\left(\frac{\Phi(s+\sigma)-\Phi(s-\sigma)}{1-G(2B(s)-1)}\right)}{ds} \right|_{s=0} > 0$ . Therefore,  $\exists C(\sigma) > 0$  such that  $\forall 0 \leq s \leq C(\sigma)$ , we have

$$\frac{d\left(\frac{\Phi(s+\sigma)-\Phi(s-\sigma)}{1-G(2B(s)-1)}\right)}{ds} > 0.$$

$\square$

First we show that  $\bar{\sigma}_v = \sup_{x>0} \left\{ x | \forall \sigma_v \in (0, x), \frac{\bar{\gamma}e^{\bar{\sigma}} - \gamma e^{\bar{\sigma}}}{\alpha_e^{\bar{\sigma}}} > \frac{\bar{\gamma}e - \gamma e}{\alpha_e} \right\}$  is increasing in  $\sigma_e$ . Because  $\frac{\mu_z}{\mu} < +\infty$  and  $\frac{\mu_z}{\mu}$  sufficiently large, according to Proposition 4, there must exist a  $\hat{\sigma}$  such that  $\bar{\sigma} = \sup_{x>0} \left\{ x | \forall \sigma \in (0, x), \frac{\bar{\gamma}e^{\bar{\sigma}} - \gamma e^{\bar{\sigma}}}{\alpha_e^{\bar{\sigma}}} > \frac{\bar{\gamma}e - \gamma e}{\alpha_e} \right\}$ . By definition  $\bar{\sigma} = \frac{\bar{\sigma}_v}{\sigma_e}$ , i.e.,  $\bar{\sigma}_v = \bar{\sigma}\sigma_e$ , where  $\bar{\sigma}$  is a constant.  $\bar{\sigma}_v$  is increasing in  $\sigma_e$ . As  $\sigma_e \rightarrow 0^+$ ,  $\bar{\sigma}_v \rightarrow 0$  and as  $\sigma_e \rightarrow +\infty$ ,  $\bar{\sigma}_v \rightarrow +\infty$ .

Next we prove that if  $\frac{\mu_z}{\mu}$  is large enough, there exists a subsequence  $\{(\frac{\mu_z}{\mu})_i\}$  such that  $\bar{\sigma}_v$  decreases as  $(\frac{\mu_z}{\mu})_i$  increases.

Let  $C(\sigma)$  defined as  $\sup_x \left\{ x | \forall s \in (0, x), \frac{d\left(\frac{\Phi(s+\sigma)-\Phi(s-\sigma)}{1-G(2B(s)-1)}\right)}{ds} > 0 \right\}$ . By Lemma 9, such  $C(\sigma)$  exists for all  $\sigma > 0$ . Note that if  $\frac{\mu_z}{\mu} \rightarrow +\infty$ , we have  $\hat{\mathbf{s}}, \mathbf{s}_1 \rightarrow 0$ . Therefore, as  $\frac{\mu_z}{\mu}$  becomes

sufficiently large, there exists  $\sigma(\frac{\mu_z}{\mu})$ , such that

$$\widehat{\mathbf{s}}, \mathbf{s}_1 < C(\sigma(\frac{\mu_z}{\mu})).$$

Thus,  $\left. \frac{d(\frac{\Phi(s+\sigma)-\Phi(s-\sigma)}{1-G(2B(s)-1)})}{ds} \right|_{s=\widehat{\mathbf{s}}} > 0$ ,  $\left. \frac{d(\frac{\Phi(s+\sigma)-\Phi(s-\sigma)}{1-G(2B(s)-1)})}{ds} \right|_{s=\mathbf{s}_1} > 0$ . And since  $\widehat{\mathbf{s}} < \mathbf{s}_1$ ,  $\frac{\overline{\gamma}_e^{\widehat{\mathbf{s}}}-\underline{\gamma}_e^{\widehat{\mathbf{s}}}}{\alpha_e^{\widehat{\mathbf{s}}}} < \frac{\overline{\gamma}_e-\underline{\gamma}_e}{\alpha_e}$

This is to say, when  $\frac{\mu_z}{\mu}$  is sufficiently large, we find an upper bound of  $\bar{\sigma}$ , i.e.,  $\bar{\sigma} < \sigma(\frac{\mu_z}{\mu})$ .

Therefore, there exists a subsequence  $\{(\frac{\mu_z}{\mu})_i\}$  such that, as  $(\frac{\mu_z}{\mu})_i$  increases,  $\sigma(\frac{\mu_z}{\mu})$  decreases, and  $\bar{\sigma}$  decreases, and as  $(\frac{\mu_z}{\mu})_i \rightarrow +\infty$ ,  $\sigma(\frac{\mu_z}{\mu}) \rightarrow 0$ ,  $\bar{\sigma} \rightarrow 0$ .

## CHAPTER 2

# Information Diffusion With Centralized and Bilateral Trading

### 2.1 Introduction

In recent years, there has been an ongoing shift in focus from public firms to private firms. More and more investors are seeking to invest in the pre-IPO market, a secondary market where buyers and sellers privately negotiate transactions of private stocks, pursuant to certain guidelines and conditions of the Securities Act of 1933. For some firms such as Facebook, Twitter, Uber, Snapchat, and Airbnb, the pre-IPO trading volume surges prior to the IPO offering.<sup>1</sup> In the meantime, online platforms, brokerage firms and derivatives has sprung up to facilitate the pre-IPO transactions.<sup>2</sup> For example, in the US, robust pre-IPO markets have been created such as SecondMarket, founded in 2004, and Nasdaq Private Market, a joint venture formed in 2013 by Nasdaq OMX and SharesPost. In Germany, broker-dealers offer OTC trading for investors who want to buy or sell IPO shares during the offer process. Similarly, in emerging markets such as Mainland China and Taiwan, significant transaction volume and attention have been observed in their pre-IPO markets in recent years.<sup>3</sup>

---

<sup>1</sup>“Tech Workers Cash In Early”, by Susan Pulliam and Telis Demos, *The Wall Street Journal*, Mar 28, 2015.

<sup>2</sup>“How Wall Street Middlemen Help Silicon Valley Employees Cash In Early”, Susan Pulliam and Telis Demos, *The Wall Street Journal*, Mar 27, 2015.

<sup>3</sup>In Mainland China, the pre-IPO market for private companies is the New Over the Counter Market, launched in SZSE. In Taiwan, pre-IPO transaction is mandatory since 2005, which is conducted in its Emerging Stock Market (ESM).



Information asymmetry and information diffusion is a substantial concern of pre-IPO markets. Investors who trade in pre-IPO markets are heterogeneous in the information held, and some have superior information than others. Pre-IPO markets can serve as a channel to disperse such information and facilitate price discovery. For example, [Chang et al. \(2016\)](#) study Taiwans pre-IPO market (ESM) and find that pre-market prices are very informative about post-market prices. [Lffler et al. \(2005\)](#) who also study the German pre-IPO markets find a similar conclusion.

The current studies about pre-IPO markets, however, do not reveal the mechanism between the pre-IPO markets and IPO markets and the mechanism of information diffusion in a pre-IPO market. This paper attempts to answer these questions. Specifically, we examine 1) How information diffuse in pre-IPO markets, and 2) how pre-IPO markets interact with the IPO offering. In this paper, the pre-IPO is regarded as a bilateral market where traders meet and trade bilaterally, keeping private their transaction prices. The IPO market is modeled as a centralized market where traders trade with a monopolistic market maker using a publically observable price. There are both informed and uninformed traders. The bilateral trading is conducted before the centralized trading. Under functional bilateral trading, traders can infer the private information held by their trading partners and, thus, information diffuses.

There are three major testable implications of the results. First, bilateral trading does not always function as a channel for information diffusion. There is an interdependence between information diffusion, bilateral trading, and centralized trading. That is, if the liquidity in the centralized market is overly restrictive or overly non-restrictive, bilateral trading is not active and hence there is no corresponding information diffusion. Only if the centralized liquidity level is restricted within a certain range, there exists an active bilateral market and hence information diffuses. Second, information diffuses in a “biased” way. Traders design their bilateral contract to infer a certain type of information (good or bad, for example) and quickly learn if their conjecture is true. However, if the actual information is against their conjecture, they only learn the information partially. Third, if the centralized prices

are informative enough, they squeeze out information diffusion in the bilateral market.

The results of the model shed light on the information content and interaction between the pre-IPO and IPO market. It also explains some phenomena in other similar markets or trading procedures that involve both bilateral and centralized meetings and information asymmetry. Examples include procedures for the promotion of agricultural innovation and market dynamics for interbank foreign exchange transactions. Empirical studies can also be launched in these areas.

This paper contributes to the theoretical market microstructure literature that studies information diffusion and information asymmetry. The paper analyzes these issues in the particular framework of interaction between the centralized market and the decentralized market. A significant amount of prior studies have focused on either centralized markets (for example, stock exchanges) or decentralized markets (for example, bilateral OTC markets). In regards to the centralized market analysis, [Kyle \(1985\)](#) considers the strategy of informed traders. [Glosten and Milgrom \(1985\)](#) consider the behavior of competitive risk neutral market makers. Other papers include [Manela \(2014\)](#), [Duffie and Manso \(2007\)](#), [Admati and Pfleiderer \(1988\)](#), and [Colla and Mele \(2010\)](#). In regards to the decentralized market analysis, [Golosov et al. \(2014\)](#) and [Wolinsky \(1990\)](#) consider the learning and information revelation process in a decentralized bilateral market. Other papers include [Babus and Kondor \(2013\)](#), [Duffie et al. \(2009\)](#), and [Ozsoylev et al. \(2014\)](#). There is little literature, however, analyzing information diffusion in the framework of both centralized and decentralized markets. Some studies focus on the trade-off between the two markets. For example, [Miao \(2006\)](#) to explain the coexistence of both types of markets as a trade-off of searching costs (in decentralized markets) and transaction costs (in centralized markets). In addition, [Rust and Hall \(2002\)](#) explains traders' entering decisions by heterogeneity of transaction cost. This paper differs with the previous literature by considering the dynamics of information diffusion among the two trading systems.

This paper also speaks to important empirical issues that involve information and pricing in both bilateral and centralized markets, for example, the pre-IPO and IPO trading markets.

Our paper is related with [Cornelli et al. \(2006\)](#) who document European pre-IPO market (the grey market) and find a close relationship between information and the underwriter’s pricing decisions. The thesis of this paper is also closed related with that of [Chen and Zhang \(2016\)](#) who conclude that a more functional pre-IPO market helps increase the pricing efficiency in the IPO market. There are also other papers that study investors’ behavior before and after IPO. For example, [Kang et al. \(2015\)](#) analyze Korean markets and prove that insiders use pre-IPO markets to take advantage of their superior information. [Chang et al. \(2016\)](#) document the Taiwan Emerging Stock Market (ESM, Taiwan’s pre-IPO market) and find a positive correlation between the informativeness of pre-market prices and the pricing efficiency in the IPO market. Similarly, [Lffler et al. \(2005\)](#) use German pre-market data and finds similar results. Other papers with similar arguments include [Dambra et al. \(2015\)](#) and [Chua \(2015\)](#).

The paper is conducted in 5 sections. Section [2.2](#) analyzes a benchmark model where agents choose to participate in both a bilateral and a centralized market. In the benchmark model, the centralized market is absolutely liquid and the market maker has no superior information. Section [2.3](#) considers an illiquid centralized market in which the market maker has no superior information. It reveals the relationship between information diffusion, bilateral trading, and liquidity level in the centralized market. Section [2.4](#) considers the model in which the centralized market is illiquid and the market maker has some superior information that the uninformed traders do not have. We show that public information can squeeze out trade in the bilateral market. Section [2.5](#) concludes the paper.

## **2.2 Benchmark: A Liquid Centralized Market with An Uninformed Market Maker**

### **2.2.1 Setup**

**Assets.** Suppose there are two states in the world,  $S \in \{H, L\}$  and two assets  $j \in \{1, 2\}$ . Asset 1 is a risky asset whose payoff is dependent on the state of the world. It pays 2 units

of consumption if  $H$  is realized and 0 units if  $L$  is realized. Asset 2 is a risk-free asset that pays 1 unit of consumption in either state.

**Market Participants.** There is a measure 1 of agents with von Neumann-Morgenstern expected utility  $\mathbb{E}[u(c)]$ , where  $\mathbb{E}$  is the expectation operator and  $u(c)$  is a CRRA utility function,  $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$ ,  $0 < \gamma \leq 1$ . The agents may trade to hedge and maximize their utility. There are two types of agents: the *informed agents* and the *uninformed agents*. The informed types observe a private noisy signal about the state of the world while the uninformed types have no such information. There is a monopolistic market maker (MM) who operates the centralized market. The market maker is risk-neutral and sets prices to maximize her profit.

**Trading Timeline.** There are 4 periods in the model: morning, noon, afternoon, and evening. In the morning, Nature draws a realization on the state of the world,  $S \in \{H, L\}$ , which is unknown to everybody. Then Nature draws a binary signal  $S_I \in \{H, L\}$ . We define the accuracy of the signal  $\delta_I$  as the conditional probability  $\delta_I = Pr(S = s | S_I = s)$ ,  $s \in \{H, L\}$ . We assume that  $1/2 \leq \delta_I \leq 1$ . The signal is private: only a fraction  $\alpha \in (0, 1)$  of agents observe  $S_I$ . The agents who observe  $S_I$  are the informed agents. Those who do not observe the signal are the uninformed agents. Each agent is assigned the same endowment with equal amounts of the two assets, denoted as  $e = (1, 1)$ . Since the measure of these agents is 1, the aggregate endowment of both assets is 1. After the realization of the signal  $S_I$ , but before noon, the market maker announces a price  $p$  to the public, indicating that the market maker is willing and committed to trade 1 unit of the risky asset with  $p$  units of the risk-free asset. We assume  $0 < p < 2$ .<sup>4</sup>

At noon, the *bilateral market* is open to all agents, where agents are randomly matched in pairs to negotiate a trade. The negotiation takes the following process. With probability .5, one of the two agents is selected as the *proposer*. The proposer then proposes a menu of offers to the other agent. A bilateral offer is composed of a quantity and a price and denoted

---

<sup>4</sup>If  $p > 2$ , agents would never trade with the market maker because the maximum consumption agents can obtain from the risky asset is 2.

as  $(\bar{z}, q)$ , with  $\bar{z}$  representing the quantity and  $q$  the price. For example, a trading option  $(\bar{z}, q)$  means that the proposer is willing to deliver  $\bar{z}$  of asset 1 in exchange for  $q\bar{z}$  of asset 2. We assume that the proposer can only offer to trade a maximum of  $\bar{z}$  units. However, the proposer can make any number of bilateral offers. The other partner who receives those offers is the *responder*. He can either accept one of the offers or reject all. If an offer  $(\bar{z}, q)$  is accepted, the proposer's portfolio becomes  $(1 - \bar{z}, 1 + q\bar{z})$ , and the responder's portfolio becomes  $(1 + \bar{z}, 1 - q\bar{z})$ . If all the selections are rejected, both agents keep their portfolio  $(1, 1)$  and move to the afternoon. An agent does not know whether his opponent is informed or not. Moreover, the bilateral trading is conducted in a private manner: an agent only observes the trading he is involved in but not the trading activities of other agents. After this trading period, everyone moves to the afternoon.

In the afternoon, the *centralized market* is open to all agents to trade with the market maker. In this period, an agent can offer to trade any amount  $y$  with the market maker at her committed price  $p$ . For purpose of this research, we do not consider the case when the market maker breaks her commitment. In our model, the proxy for the liquidity level in the centralized market is the restrictions on the trading quantity  $y$ . That is, if there are no restrictions on the trading quantity in the centralized market  $y$ , the market is completely *liquid*, whereas if there is any restriction on  $y$ , the market is *illiquid*. The more restricted  $y$  is, the more illiquid the centralized market is. Suppose an agent with portfolio  $(x_1, x_2)$  at the beginning of this period decides to trade an amount  $y$  (note that  $y \geq 0$  represents selling and  $y < 0$  purchasing), then after this trade the agent's portfolio becomes  $(x_1 - y, x_2 + py)$ . If the agent does not trade with the market maker, he has to keep the portfolio  $(x_1, x_2)$  and move to the evening.

In the evening, the true state is revealed and everyone consumes according to the portfolio held. Figure 2.1 describes the trading timeline of the model.

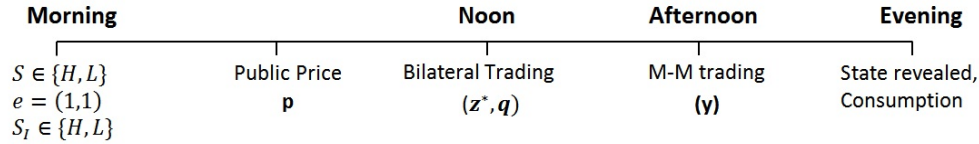


Figure 2.1: Timeline (Benchmark Model)

**Examples.** There are some practical trading examples which have similar procedures as our model. For example, modern technologies in biology and engineering have created many innovations in traditional agricultural regions. From time to time farmers receive new products which are alleged to contain certain technological advantages. The payoffs of using the new products is uncertain. Sometimes the uncertainty comes from the reputation of the company who made the products, and sometimes it comes from the cognitive barriers that the technology has created. For instance, these barriers could prevent the general public from identifying sources of information, from searching for useful information, and from dealing with information overload. Nevertheless, some farmers may have better information than others: they may have been trained as scientists, or simply have participated in earlier experiments on the products. Furthermore, in order to make better decisions, farmers may trade bilaterally with each other as a channel to access the insider opinions about the product before trading with the distributor. The bilateral trade between the farmers can be regarded as a process of information acquisition, and the distributor of the product can be viewed as the market maker in the model.

The model can also be associated with the pre-IPO and IPO market, where the pre-IPO market can be regarded as the bilateral searching market and the IPO market the centralized trading market. The pre-IPO market is the secondary market for shares of private companies held by employees and other investors. Recently, activities in the pre-IPO market have been booming in the US because of the runaway popularity of up-and-comers such as Uber Technologies, Snapchat Inc., Airbnb Inc. According to a Wall Street Journal Report, participants in pre-IPO trading estimate that 10 billion to 30 billion in stock changed hands

in 2014.<sup>5</sup> Ad hoc marketplaces for trading pre-IPO were created to facilitate the search for and transaction of pre-IPO stocks. For example, in the US, brokerage firms and funds and online trading platforms such as SecondMarket and SharesPost, and even exchange-sponsored pre-IPO markets such as Nasdaq Private Market, sprang up to help search for pre-IPO trading. Pre-IPO markets also exist in other emerging markets such as in Mainland China and Taiwan. In Mainland China, the National Equities Exchange and Quotations (or New Over-The-Counter Market) in the Shenzhen Stock Exchange serves the role of a bilateral trading system for pre-IPO transactions. In Taiwan, since 2005, private firms are required to trade on Taiwans Emerging Stock Market (ESM) for at least six months before they can apply for an IPO. Since there are both insiders and outsiders in the pre-IPO markets, information diffusion and interaction between the pre-IPO market and the post-IPO market is crucial.

**Market Maker Information.** In the benchmark model, the market maker receives no signals about the state of the world. This can be related to some markets where there is a strict control over the market maker's information acquisition. For example, for the NYSE Designated Market Maker, a significant amount of rules and principals have been set to prevent the DMM from obtaining or using inside information. Therefore, they face some of the same risks as other uninformed traders. Nevertheless, since acquiring information is costly, so the market maker cannot access the same information without paying a corresponding cost.

In Section 2.4, however, we consider the model where the market maker has certain information about the asset value. This can be related to the markets where the market maker, or product distributor, obtains private information because of their advantages in resources, search skills, and processing technologies. For example, in the IPO case, the underwriting firm certainly has better information than the general public. In this section, therefore, we want to study whether or not such information held by the market maker would

---

<sup>5</sup>See <http://blogs.wsj.com/briefly/2015/03/27/5-things-to-know-about-pre-ipo-stocks/>.

be sent out to the public when the market maker decides to announce his initial price. More details are provided in Section 2.4.

### 2.2.2 Equilibrium Definition

We now define a perfect Bayesian equilibrium of the trading game. Let  $\Theta = \{H, L, U\}$  denote the type of agents, representing an agent receiving an  $H$  signal, an  $L$  signal, or no signal respectively. Then we define an indicator variable  $\iota \in \{0, 1\}$ , which is equal to 1 if the agent is selected as a proposer and 0 otherwise. Let  $Q^3$  be the space of the private offers with  $Q = (-2, 2) \times \{no\ proposal\}$  representing the price space and 3 representing the number of proposals. As there are at most 3 types, the number of proposals is at most 3. We can then define an indicator variable  $r_i \in \{0, 1\}$ , which equals to 1 if proposal  $i$  is accepted and 0 otherwise.

#### (i): Agents' Actions

An agent's action in the first bilateral trading round is given below:

If he is selected as a proposer, his action is given by the map:

$$q : \Theta \times R_+ \rightarrow Q^3.$$

If he is selected as a responder, his action is given by the map:

$$r : \Theta \times R_+ \times Q^3 \rightarrow \{0, 1\}^N.$$

Let  $\Omega = \{0, 1\} \times Q^3 \times \{0, 1\}^3$  denote the bilateral trading history. For example,  $\omega = (0, (-1, 1.5), (0, 1))$  represents that the agent is selected as a responder and observes two proposals, with one proposal offering to **buy**  $\bar{z}$  units of the risky asset at price 1, and the other offering to **sell**  $\bar{z}$  units of risky asset at price 1.5, and he chooses to reject the first offer and accept the second one.

An agent's action in the market maker trading round is given by:

$$y : \Theta \times R_+ \times \Omega \rightarrow R.$$



The Market Maker's action is to choose price  $p \in R_+$ .

**(ii): Agents' Beliefs**

At the beginning of the bilateral trading round, agents' beliefs post-receiving the private signal  $H$  are given below:

If he is selected as a proposer, his belief is represented by:

$$b^p : \Theta \times R_+ \rightarrow [0, 1].$$

If he is selected as a responder, by observing the offers  $(q_1, \dots, q_n)$ , his belief becomes:

$$b^r : \Theta \times R_+ \times Q^3 \rightarrow [0, 1].$$

After the bilateral trading round, agent beliefs are updated as:

$$b_a : \Theta \times R_+ \times \Omega \rightarrow [0, 1].$$

Note that in this game, the market maker's belief is  $b_b^M = \frac{1}{2}$  before the centralized trading round, and is updated to  $b_a^M : (0, 1) \times Y \rightarrow [0, 1]$  after observing agents' trading volumes. Informed agent beliefs do not change over time, that is:  $b^p(H) = b^r(H) = b_a(H)$ ,  $b^p(L) = b^r(L) = b_a(L)$ .

**Definition 3.** A Perfect Bayesian Equilibrium is given by a) agents' strategy profiles  $\sigma = \{q, r, y\}$ , and agents' beliefs  $b = \{b^p, b^r, b_a\}$ , and b) market maker's strategy  $p$ , beliefs  $b^M = \{b_b^M, b_a^M\}$  such that:

- (i)  $\sigma$  is optimal for an individual agent given his belief  $b$ ;
- (ii)  $p$  is optimal for the market maker given his belief  $b^M$ ;
- (iii) beliefs are correct and consistent with Bayes Rule whenever possible.

### 2.2.3 Equilibrium Characterization

The equilibrium can be derived via backward induction. First, we shall characterize traders' strategies in the afternoon, the centralized trading round. Lemma 10 gives us the agent's optimal trading volume and the corresponding utility achieved.

**Lemma 10.** In the centralized trading round, the optimal trading volume for an agent with portfolio  $(x_1, x_2)$  and  $\delta = Pr(S = H|b)$  is given by:

$$y^*(p, x_1, x_2; \delta) = \frac{2x_1 + x_2(1 - \varphi(\delta, p))}{\varphi(\delta, p)p + 2 - p}.$$

and ex-post utility is given by:

$$u^*(p, x_1, x_2; \delta) = \frac{\delta x_H^{1-\gamma} + (1 - \delta)x_L^{1-\gamma} - 1}{1 - \gamma},$$

where  $b$  is the belief of the agent and

$$\varphi(\delta, p) = \left( \frac{1 - \delta}{\delta} \frac{p}{2 - p} \right)^{-\frac{1}{\gamma}},$$

$$x_L = \frac{2px_1 + 2x_2}{\varphi p + 2 - p}, \quad x_H = \frac{2p\varphi x_1 + 2\varphi x_2}{\varphi p + 2 - p}.$$

**Proof.** An agent with portfolio  $(x_1, x_2)$  and  $\delta = Pr(S = H|b)$  solves the problem:

$$\max_y E[u(y)|b] = \frac{\delta x_H(y)^{1-\gamma} + (1 - \delta)x_L(y)^{1-\gamma} - 1}{1 - \gamma},$$

where

$$x_H(y) = 2(x_1 - y) + x_2 + py, \text{ and } x_L(y) = x_2 + py.$$

By the concavity of the objective function and the assumption that agents are allowed to take short positions, the F.O.C of this problem obtains the result.

□

One should note that if  $y^* \geq 0$ , then the agent's best response is to sell the risky asset, and if  $y^* < 0$ , the agent's best response is to buy.

Lemma 11 studies the properties of the traders' optimal positions in the centralized market,  $y(p, x_1, x_2; \delta)$ .

**Lemma 11.**  $y^*(p, x_1, x_2; \delta)$  is a strictly increasing function of  $p$  and a strictly decreasing function of  $\delta$ . Thus we have  $y_H < y_M < y_L$ , where

$$\begin{aligned} y_H &= y^*(p, x_1, x_2; \delta_I), \\ y_M &= y^*(p, x_1, x_2; 0.5), \\ y_L &= y^*(p, x_1, x_2; 1 - \delta_I). \end{aligned}$$

**Proof.**

$$\begin{aligned} \frac{\partial y^*}{\partial p} &= \frac{-2\frac{\partial \varphi}{\partial p}(1+p) + (\varphi - 1)(\varphi - 3)}{(\varphi p + 2 - p)^2} \\ &= \frac{1}{(\varphi p + 2 - p)^2} \left( \frac{4(1+p)\varphi}{(2-p)p\gamma} + (\varphi - 1)(\varphi - 3) \right) \\ &> 0. \end{aligned}$$

The second equality is due to  $\frac{\partial \varphi}{\partial p} = -\frac{2\varphi}{\gamma(2-p)p}$ , and the last inequality comes from the fact that  $0 < p < 2$ ,  $0 < \gamma \leq 1$ .

Similarly, we can calculate that:  $\frac{\partial y^*}{\partial \delta} = \frac{\partial y^*}{\partial \varphi} \frac{\partial \varphi}{\partial \delta} < 0$ .

Note that  $\frac{\partial y^*}{\partial \varphi} < 0$ ,  $\frac{\partial \varphi}{\partial \delta} > 0$ .

□

There are two implications of Lemma 11. First, agents' optimal trading volumes largely depend on the price in the centralized market. For example, if the optimal choice for an agent is to sell the risky asset ( $y^* \geq 0$ ) to the market maker, then the selling volume  $y^*$  increases if the price  $p$  increases. In contrast, if the optimal choice for the agent is to buy

the risky asset ( $y^* < 0$ ) from the market maker, then the buying volume  $y^*$  decreases as  $p$  increases. This is consistent with the fact that agents always want to buy assets at a lower price and sell assets at a higher price. Second, if an agent believes that the true state is more likely to be  $H$  (i.e.  $\delta$  is high), which implies that the risky asset has a high probability to achieve a high payoff, then his intentions to buy the risky asset becomes stronger (i.e.  $y^*$  decreases as  $\delta$  increases).

Given the agents' trading strategy in the centralized market, we then focus on their trading strategies in the bilateral trading round. Theorem 3 characterizes their bilateral trading behavior in the benchmark model.

**Theorem 3.** If there is no trading limitation in the centralized market (no restrictions on  $y$ ) and the market maker receives no signals, then, in equilibrium, there is no trade in the bilateral market.

**Proof.** The centralized market is fully open if all traders have an equal opportunity to participate. It is complete when the price is fully observable and all traders trade at one single price, and there are no trading limitations or price spreads. In this case, for trade to happen in the bilateral trading round a mutually beneficial proposal must be found. Suppose a proposal  $(\bar{z}, q)$ ,  $\bar{z} > 0$  is accepted, that is to say, both proposer and responder agree to exchange  $\bar{z}$  units of asset 1 with  $\bar{z}q$  units of asset 2. After such a trade, the responder's portfolio becomes  $(1 + \bar{z}, 1 - \bar{z}q)$ , and his belief about the true state could possibly change, becoming  $\delta$ . Hence his utility later after trading with the market maker is:

$$\tilde{u}(p, 1 + \bar{z}, 1 - \bar{z}q; \delta) = \frac{1}{1 - \gamma} \left\{ \left[ \frac{2(p + 1 + z(p - q))}{\varphi p + 2 - p} \right]^{1-\gamma} (\delta \varphi^{1-\gamma} + 1 - \delta) - 1 \right\},$$

and the proposer's portfolio becomes  $(1 - \bar{z}, 1 + \bar{z}q)$ , with belief  $\tilde{\delta}$ , and utility later of:

$$\tilde{u}(p, 1 - \bar{z}, 1 + \bar{z}q; \tilde{\delta}) = \frac{1}{1 - \gamma} \left\{ \left[ \frac{2(p + 1 - z(p - q))}{\tilde{\varphi} p + 2 - p} \right]^{1-\gamma} (\tilde{\delta} \tilde{\varphi}^{1-\gamma} + 1 - \tilde{\delta}) - 1 \right\}.$$

There are several cases to study.

**Case(i):** The proposer is an informed agent. An informed agent's belief does not change over time. Hence  $\delta = \delta_I \in [1/2, 1]$  if he is an H type and  $\delta = 1 - \delta_I$  if he is an L type. And

because  $(\delta\varphi^{1-\gamma} + 1 - \delta) > 0$ ,  $\varphi(p, \delta) > 0$ , for him to benefit from the trade, both H and L type informed agents will offer  $q \geq p$ . That is, the proposer sells  $\bar{z}$  units of asset 1 at a price  $q$  bigger than the public price  $q$ . But this would be impossible because if the responder is also an informed agent (same type), such a proposal would certainly be against his favor and he would reject it. If the responder is an uninformed agent, he cannot learn any information by observing such a proposal. Hence, his best response would be also to reject the proposal.

**Case (ii):** The proposer is an uninformed agent. The uninformed agent is willing to lose some money in the bilateral trading period only when he can learn the information and hence adjust the right trading volume later in the centralized trading round. Since both H and L type responders' best responses are the same, the uninformed proposer cannot learn by observing the acceptance of his proposal. Hence he would not propose any offer that is against his benefit. And other offers, of course, would be rejected.

Hence there is no trade in the bilateral trading period. □

Theorem 3 reveals one aspect of the relationship between the centralized market and the bilateral trading market: when the centralized market is absolutely liquid, it squeezes out the bilateral trading. The intuition is straightforward: in order for a trade to take place in the bilateral market, the trade must provide incentives for both parties. Yet, when the centralized market is absolutely liquid, it fails the incentive of informed traders to participate in the bilateral market. Specifically, for an uninformed trader, seeking information is the incentive to encourage their activities in the bilateral market. However, for an informed trader, trading in the bilateral market does not increase their informativeness since all informed traders receive the same signal. If the centralized market is absolutely liquid, they can trade any amount with the market maker. Therefore, there is no incentive for them to participate in the bilateral market.

### 2.2.4 The Market Maker's Profit

We now analyze the market maker's profit. The following Lemma and Proposition describes what constitutes the market maker's profit and the conditions for a positive profit.

**Lemma 12.** If there is no trading limit in the centralized market (no restrictions on  $y$ ) and the market maker receives no signals, then, in equilibrium, the market maker's price is bounded by  $2(1 - \delta_I) < p \leq 1$ .

**Proof.** The market maker's profit function is:

$$\begin{aligned}\Pi(p) &= E[E[(\tilde{F} - p)Y|b_a^M]] \\ &= \frac{1}{2} [(2\delta_H - p)Y_H + (2\delta_L - p)Y_L],\end{aligned}$$

where  $\delta_H = Pr(S = H|b_a^M = 1) = \delta_I$ ,  $\delta_L = Pr(S = H|b_a^M = 0) = 1 - \delta_I$ ,

Consider on the equilibrium path,  $b_a^M = 1$  if  $Y = Y_H$ , and  $b_a^M = 0$  if  $Y = Y_L$ , where  $Y_H$  and  $Y_L$  are the aggregate trading volumes when the signal is H and L respectively.

Note that  $Y_H$  and  $Y_L$  are increasing functions of  $p$ , thus  $\Pi(p)$  is a strictly concave function of  $p$  and there exists a unique optimal  $p$ .

By taking a derivative of the objective function we get

$$\Pi'(p) = \frac{1}{2} \left[ (2\delta_H - p) \frac{\partial Y_H}{\partial p} + (2\delta_L - p) \frac{\partial Y_L}{\partial p} - Y_H - Y_L \right].$$

Then we can immediately conclude that  $p \leq 2\delta_H$ . Suppose  $p > 2\delta_H$ , then we have  $\delta(b) = Pr(S = H|b) \leq 1, \forall b \in [0, 1]$ , and hence  $Y_H \geq 0$  and  $Y_L \geq 0$ . Since  $\frac{\partial Y_i}{\partial p} > 0, \forall i = H, L$ ,  $\Pi'(p) < 0$ . Moreover, we could conclude that  $p \leq 1$  because then  $\Pi'(1) \leq 0$ . Also by the same argument we can see that  $p \geq 2(1 - \delta_I)$ .  $\square$

Lemma 12 provides an upper and lower bound for the equilibrium market maker price  $p$ . Then Proposition 6 finds the necessary and sufficient condition for the market maker to make positive profit.

**Proposition 6.** In equilibrium, the market maker loses money to the informed traders and makes money from the uninformed traders. And, for each  $\delta_I \in [0.5, 1]$ , there exists a  $\alpha^*(\delta_I) \geq 0$  such that

$$\Pi \geq 0 \text{ if } \alpha \geq \alpha^*(\delta_I),$$

$$\Pi < 0 \text{ if } \alpha < \alpha^*(\delta_I).$$

where  $\Pi$  is the profit for the market maker.

**Proof.** The market maker's profit function is:

$$\begin{aligned} \Pi &= E[E[(\tilde{F} - p)Y|b_a^M]] \\ &= \frac{1}{2} [(2\delta_H - p)Y_H + (2\delta_L - p)Y_L], \end{aligned}$$

where  $\delta_H = Pr(S = H|b_a^M = 1) = \delta_I$ ,  $\delta_L = Pr(S = H|b_a^M = 0) = 1 - \delta_I$ , and  $Y_H = \alpha y_H + (1 - \alpha)y_M$  and  $Y_L = \alpha y_L + (1 - \alpha)y_M$  where  $y_H, y_M$  and  $y_L$  are trading volumes for type  $H, U, L$  respectively.

Hence

$$\begin{aligned} \Pi(\alpha) &= \frac{\alpha}{2} [(2\delta_H - p)y_H + (2\delta_L - p)y_L] + (1 - \alpha)(1 - p)y_M \\ &= \alpha\Pi_I + (1 - \alpha)\Pi_U. \end{aligned}$$

The first term  $\Pi_I$  is the profit made from informed traders, and the second term  $\Pi_U$  is the profit made from uninformed traders. As price  $p$  is bounded by  $(2(1 - \delta_I), 1)$ , and by Lemma 10, we have that  $y_H < 0, y_L > 0$  and  $y_M > 0$ , hence  $\Pi_I = [(2\delta_H - p)y_H + (2\delta_L - p)y_L] < 0$  and  $\Pi_U = (1 - p)y_M > 0$ . That is, the market maker loses money to the informed traders and earns positive profit from the uninformed traders.  $\Pi$  decreases as  $\alpha$  increases. and  $\Pi(0) > 0$ . Thus, there must exist some  $\alpha^*(\delta_I)$  such that  $\Pi(\alpha^*) = 0$ .

Also, one could notice that as  $\delta_I$  increases,  $\delta_H$  increases and  $\delta_L$  decreases. Thus,  $y_H$  decreases and  $y_L$  increases, and so  $\Pi_I = [(2\delta_H - p)y_H + (2\delta_L - p)y_L] < 0$  decreases. That is to say, as the informed agents procure more accurate information, the market maker loses more money to them. Hence,  $\alpha^*(\delta_I)$  decreases with  $\delta_I$ .  $\square$

Proposition 6 states that, if the fraction of the informed traders is sufficiently small, the monopolistic market maker can make profit. Figure 2.2 captures a numerical example of the market maker’s profit in the benchmark model. As exhibited in the figure, as the accuracy of the private information increases, the profit of the market maker decreases. Also, as the population of the information agents increases, the profit of the market maker decreases. And the break-even points show that the likelihood for the market maker to make positive profits is higher when the private information is less accurate.

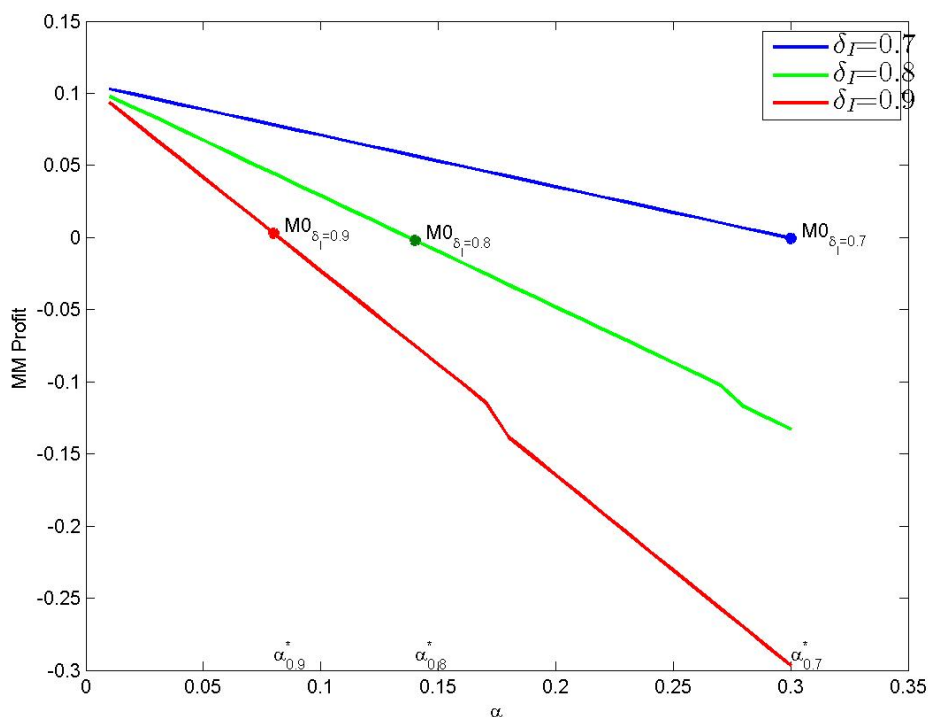


Figure 2.2: **The Market Maker’s profit As  $\alpha$  Varies** (no trading limitation, no information for the market maker).  $M0$  represents the point where the market maker breaks even. The parameters are  $\gamma = 0.8, \bar{z} = 0.3$ .

In this section, we conclude that if the centralized market is “good enough” (too liquid) for the traders, no trades would take place in the bilateral market. Since the bilateral



market is the channel for agents to acquire information, there is no information acquisition and hence no information diffusion in this case. Also, the market maker in the centralized market simply plays the role as a liquidity provider: they take any order from the agents and hence may lose money if the population of informed traders is too large.

Driven by the fact that, in reality, the bilateral market and centralized trading may coexist and market makers make money, we identify a factor that can link the two markets and also influence the profit of the market maker. This is characterized in Section 2.3.

### 2.3 Illiquid Centralized Market and An Uninformed Market Maker

In the benchmark model, we assume that the market maker does not impose any trading limitation on the transactions in the centralized market. That is, we assume that traders can trade any amount of the risky asset with the market maker. In this section, we assume that the market maker imposes a trading limitation and restricts traders from trading too aggressively. We argue that this trading limitation is key to link the bilateral and centralized markets. It also determines the profit of the market maker. Once a trading limit is imposed, the centralized market is not absolutely liquid anymore. We call it an illiquid centralized market. The liquidity level depends on how restrictive the trading limits are.

The assumption of a trading limitation allows us to associate the model with practical issues. In financial practices, all financial institutions incorporate risk management as one of their objectives. A restriction on the trading quantity is one of the commonly used measures to prevent a significant loss. To incorporate that in the model, we assume that, beside the trading price  $p$ , in the morning the market maker also announces a trading limit,  $\underline{y}$  and  $\bar{y}$ , indicating that traders can only be allowed to trade a volume  $\underline{y} \leq y \leq \bar{y}$ . Everything else is the same as in the benchmark model.

### 2.3.1 Equilibrium Characterization

In this section, we characterize the equilibrium of the model when there are trading limits imposed and the market maker has no information about the risky asset value. At this point, we assume that the trading limits,  $\underline{y}$  and  $\bar{y}$ , are exogenous restrictions. In future studies, we can relax this assumption and endogenize the parameter.

Lemma 13 provides us the upper bound and lower bound of the market maker price  $p$ .

**Lemma 13.** Suppose the market maker imposes the trading limits,  $\underline{y}$ ,  $\bar{y}$ , and  $-1 \leq \underline{y} < \bar{y} \leq 1$ , then, in equilibrium, the market maker's price is bounded by  $2(1 - \delta_I) < p \leq 1$ .

**Proof.** Similar with Lemma 12. □

Now we study the relationship between the trading limitation and traders' strategies, bilateral trading, and information diffusion. This is characterized in Theorem 4.

**Theorem 4. (Existence of bilateral trading and information diffusion)**

Suppose the accuracy of the signal  $\delta_I = Pr(S = s | S_I = s)$  is sufficiently large, and  $y_0 < \underline{y} < y'_0 < y'_1 < \bar{y} < y_1$ , where  $y_0, y_1$  solve:

$$y_0 = \max\{y^*(1, 1, 1; 1 - \delta_I), y^*(2(1 - \delta_I), 1, 1; 0.5)\}, \quad (2.1)$$

$$y_1 = \min\{y^*(1, 1, 1; 0.5), y^*(2(1 - \delta_I), 1, 1; \delta_I)\}, \quad (2.2)$$

and  $y'_0, y'_1$  satisfy:

$$\max_{y \in [y'_0, y'_1]} u(p, y, 1 - \bar{z}, 1 + \bar{z}q^U; \delta_I) = u(p, y^*(p, 1, 1; 0.5), 1, 1; \delta_I), \quad (2.3)$$

where  $q^U$  solves

$$\max_{y \in [\underline{y}, \bar{y}]} u(p, y, 1 + \bar{z}, 1 - \bar{z}q^U; \delta_I) = \max_{y \in [\underline{y}, \bar{y}]} u(p, y, 1, 1; \delta_I).$$

then a Perfect Bayesian Equilibrium exists in which bilateral trading is active and therefore the fraction of the informed population increases.

**Proof.** Recall that if both proposers and responders agree to exchange  $\bar{z}$  units of asset 1 with  $\bar{z}q$  units of asset 2, the responder's portfolio becomes  $(1 + \bar{z}, 1 - \bar{z}q)$ . Suppose his belief is  $\delta$ , then his utility later after trading with the market maker is:

$$\begin{aligned} U^r(p, q; \delta) &= \max_{\underline{y} \leq y \leq \bar{y}} u(p, y, 1 + \bar{z}, 1 - \bar{z}q; \delta) \\ &\leq \tilde{u}(p, 1 + \bar{z}, 1 - \bar{z}q; \delta) \\ &= \frac{1}{1 - \gamma} \left\{ \left[ \frac{2(p + 1 + z(p - q))}{\varphi p + 2 - p} \right]^{1-\gamma} (\delta \varphi^{1-\gamma} + 1 - \delta) - 1 \right\}, \end{aligned}$$

and the proposer's portfolio becomes  $(1 - \bar{z}, 1 + \bar{z}q)$ , with belief  $\tilde{\delta}$ , and the utility later being:

$$\begin{aligned} U^p(p, q; \tilde{\delta}) &= \max_{\underline{y} \leq y \leq \bar{y}} u(p, y, 1 - \bar{z}, 1 + \bar{z}q; \tilde{\delta}) \\ &\leq \tilde{u}(p, 1 - \bar{z}, 1 + \bar{z}q; \tilde{\delta}) \\ &= \frac{1}{1 - \gamma} \left\{ \left[ \frac{2(p + 1 - z(p - q))}{\varphi p + 2 - p} \right]^{1-\gamma} (\delta \varphi^{1-\gamma} + 1 - \delta) - 1 \right\}. \end{aligned}$$

One should notice that  $U^r(p, q; \delta)$  is a decreasing function of  $q$ . Also, by Lemma 11,  $y^*(\cdot; \delta_I) < y^*(\cdot; 0.5) < y^*(\cdot; 1 - \delta_I)$ . If  $\underline{y}$  and  $\bar{y}$  satisfy conditions (2.1) and (2.2), then we must have that, for all  $p \in [2(1 - \delta_I), 1]$  (by lemma 13),

$$y^*(p, 1, 1; \delta_I) \leq y_0 < \underline{y} \leq y^*(p, 1, 1; 0.5) \leq \bar{y} < y_1 \leq y^*(p, 1, 1; 1 - \delta_I). \quad (2.4)$$

That is to say the unbounded optimal trading volume for an H type informed trader is below the lower boundary  $\underline{y}$  and above the upper boundary  $\bar{y}$  for the L type. Hence we have that  $\forall p \in [2(1 - \delta_I), 1]$ ,

$$\begin{aligned} u^*(p, 1, 1; \delta_I) &\geq \max_{\underline{y} \leq y \leq \bar{y}} u(p, y, 1, 1; \delta_I) = u_c^*(p, 1, 1; \delta_I), \\ u^*(p, 1, 1; 0.5) &= \max_{\underline{y} \leq y \leq \bar{y}} u(p, y, 1, 1; 0.5) = u_c^*(p, 1, 1; 0.5), \\ u^*(p, 1, 1; 1 - \delta_I) &\geq \max_{\underline{y} \leq y \leq \bar{y}} u(p, y, 1, 1; 1 - \delta_I) = u_c^*(p, 1, 1; 1 - \delta_I). \end{aligned}$$

Because of the limited order constraint, the informed agents have a motivation to trade in the bilateral market. And the bilateral trading in equilibrium is characterized as below:

**Case(i)** If the proposer is an uninformed agent.

In this case the uninformed agent can propose one option  $(\bar{z}, q^U(p))$ , targeting the H type informed agent.  $q^U(p)$  satisfies:

$$(IC1) : U^r(p, q^U(p); \delta_I) = u_c^*(p, 1, 1, \delta_I).$$

According to the argument above, we can conclude that  $q^U(p) \geq p$  by condition (2.4) and the fact that  $U^r(p, q, \delta)$  is a decreasing function of  $q$ . The H type informed will accept the option, and the uninformed and L informed will accept neither. Also if  $\delta_I$  is sufficiently large, and 2.3 is satisfied, we have  $U^p(p, q^U(p), \delta_I) \geq u(p, y^*(p, 1, 1; 0.5), 1, 1; \delta_I))$ . That is, even though the proposer loses money to the informed responder in the bilateral round, he will acquire the right information and gain back in the centralized trading period. Given the actions of the responder, the uninformed agents achieve the best payoff by setting  $q = q^U(p)$ . If he observes a rejection, he cannot tell whether it is because of that the private signals are L, or he simply came across an ignorant agent just like himself.

**Case(ii)**: If the proposer is an informed agent, suppose the proposer is an H type informed trader. If one of his proposals  $(\bar{z}, q^H(p))$  is accepted by the responder, his portfolio becomes  $(1 - \bar{z}, 1 + q^H(p)\bar{z})$ , and his utility becomes:

$$\begin{aligned} U^p(p, q^H(p); \delta_I) &= \max_{y \leq y \leq \bar{y}} u(p, y, 1 - \bar{z}, 1 + \bar{z}q; \delta_I) \\ &< \frac{1}{1 - \gamma} \left\{ \left[ \frac{2(p + 1 - z(p - q))}{\varphi p + 2 - p} \right]^{1-\gamma} (\delta \varphi^{1-\gamma} + 1 - \delta) - 1 \right\}, \end{aligned}$$

by the fact that  $y^* < y_0$ . As  $U^p(p, q; \delta_I)$  is an increasing function of  $q$ , he would be happy as long as  $q \geq q^H(p)$ , where  $q^H(p)$  satisfies:

$$(IC2) : U^p(p, q^H(p); \delta_I) = u_c^*(p, 1, 1; \delta_I).$$

Comparing (IC1) and (IC2), we have  $q^H(p) \geq q^U(p) \geq p$ .

Similarly, suppose the proposer is an L type informed trader, his proposal  $(\bar{z}, q^L(p)\bar{z})$

must satisfy  $q \leq q^L(p)$  where  $q^L(p)$  satisfies:

$$(IC3) : U^p(p, q^L(p); 1 - \delta_I) = u_c^*(p, 1, 1; 1 - \delta_I).$$

And by a similar argument,  $q^H(p) > q^U(p) > p > q^L(p)$ .

As the game itself is common knowledge, proposals of informed agents would be rejected because all informed agents have the same signals and uninformed agents can immediately make an inference about the information and hence understand that such proposals are against them. Thus if the proposer is an informed agent, no responder will accept his proposal. And given that no one accepts, he has no incentive to deviate.

Hence in equilibrium the actions of the agents are:

$$q(\theta, p) = q^\theta(p), \theta = H, U, L. \quad (2.5)$$

$$r(H, p, q) = \begin{cases} 0 & \text{if } q > q^U(p), \\ 1 & \text{if } q \leq q^U(p). \end{cases} \quad (2.6)$$

$$r(U, p, q) = 0, \forall q. \quad (2.7)$$

$$r(L, p, q) = 0, \forall q. \quad (2.8)$$

Agents' beliefs are: for the proposer, before the bilateral trading,

$$b^p(\theta, p) = \begin{cases} 1 & \text{if } \theta = H, \\ 0.5 & \text{if } \theta = U, \\ 0 & \text{if } \theta = L. \end{cases} \quad (2.9)$$

$$b^r(\theta, p, q) = \begin{cases} 1 & \text{if } \theta = H, \text{ or } \theta = U, q \geq q^H(p), \\ 0 & \text{if } \theta = L, \text{ or } \theta = U, q \leq q^L(p), \\ 0.5 & \text{if } \theta = U, q \in (q^H(p), q^L(p)). \end{cases} \quad (2.10)$$

$$b_a(\theta, p, q, \omega) = \begin{cases} 1 & \text{if } \theta = H, \text{ or } \theta = U, q \leq q^U(p), \omega = (1, q, 1) \\ & \text{or } \theta = U, q \geq q^H(p), \omega = (0, q, 0), \\ \frac{0+1}{2} & \text{if } \theta = U, q \leq q^U(p), \omega = (1, q, 0), \\ 0.5 & \text{if } \theta = U, q^H(p) > q > q^U(p), \omega = (0, q, 0), \\ 0 & \text{if } \theta = L, \text{ or } \theta = U, q \leq q^L(p), \omega = (0, q, 0). \end{cases} \quad (2.11)$$

□

Theorem 4 provides us a sufficient condition such that bilateral and centralized trading coexist. Trading limitations create the trading incentives for both informed and uninformed agents in the bilateral market. As we have discussed, the uninformed traders are willing to trade in the bilateral period to acquire information. But the absence of trading limitations fails in providing the other party, the informed traders, incentives to participate in the bilateral market because they don't need to do so. An absolutely liquid centralized market has given them enough opportunity to hedge their positions. However, a reasonable trading limitation limits their actions in the centralized market and encourages them to participate in the bilateral period. Therefore, with a reasonable trading limitation, both parties are willing to trade in the bilateral market. By observing the behavior of one's trading partner, an uninformed trader infers and learns the private information. The information thus diffuses through bilateral trading.

### 2.3.2 Information diffusion and Market Interaction

**Market Interaction.** In this section, we study the interaction between the bilateral and centralized markets. Table 2.1 provides a numerical example analyzing the relationship between the trading limits in the centralized market and bilateral trading activities. First, as is shown in the table, if the centralized market is shut down (i.e.  $\underline{y} = \bar{y} = 0$ ), we observe no trading in the bilateral market, and hence no information diffusion. This result is consistent with the “No Trade Theorem” in [Milgrom and Stokey \(1982\)](#). The intuition is as follows. Since the information structure is common knowledge and all traders have no opportunity to trade later, revealing private information would be against the benefit of an informed trader. This is because no one would accept his offers once the information is induced. Therefore, access to different information cannot be the sole basis for trade. No-trade in the bilateral market still holds if we open the door slightly (for example,  $\underline{y} = -0.05, \bar{y} = 0.05$ ). Note that in this case even though both the informed traders and uninformed traders have incentives

to participate in the bilateral trade, we would still observe no trade because the loss of an uninformed trader to his informed trading partner exceeds the make-up benefit he obtains from trading with the market maker. That is, learning is too costly for an uninformed trader when the centralized market is too illiquid (the door is too narrow). Third, if we continue to increase the liquidity level in the centralized market (i.e. the gap between  $\bar{y}$  and  $\underline{y}$  becomes larger), we would observe active trading behavior in the bilateral markets. In this case both informed and uninformed traders have proper incentive to trade in both markets: informed traders trade to compensate the limit of their hedging positions in the centralized market while uninformed traders trade to learn private information. Finally, bilateral trading continues to be active as we widen the door until it hit another threshold and triggers no trade in the bilateral market again. We have discussed the intuition in the benchmark model.

To conclude my analysis of Table 2.1, whether or not bilateral trading is active depends on how liquid the centralized market is (that is, how restricted the centralized trades are). There is no bilateral trading if the centralized trading is too liquid or too illiquid. Bilateral trading is active only if the liquidity level in the centralized market is within a certain range.

**Information Diffusion.** We also analyze information diffusion through bilateral trade. In equilibrium, an uninformed proposer learns by observing the action of his responder. We have shown that, in order to satisfy traders' incentive constraints, an uninformed proposer can only aim for learning one particular type of signal. For example, an uninformed proposer can aim for learning signal  $H$ . If the true signal is  $H$ , an informed responder will accept his offer while an uninformed responder will reject his offer. Therefore by observing the responder's behavior he will immediately know the signal type if the offer is accepted. Learning is "fast" in this case. However, if the true signal is  $L$ , an informed responder will reject his offer because his signal is "L" and the offer is designed particularly for luring traders with  $H$  signals. An uninformed responder will also reject the offer because he immediately infers that his proposer is also uninformed. In this case the uninformed proposer cannot distin-

	[0,0]	[-.05,.05]	[-.1,.1]	[-.15,.15]	[-.2,.2]	[-.4,.4]	[-.7,.7]	[-1,1]	[-1.5,1.5]	[-2,2]
$\Pi$	0	0.0157	0.0229	0.0321	0.0396	0.0593	0.0466	0.0397	0.0397	0.0397
$p$	\	0.6047	0.6230	0.6420	0.6615	0.7438	0.8064	0.8000	0.7826	0.7826
$q^U$	\	\	0.8584	0.8352	0.8110	0.7488	0.8064	0.8000	0.7826	0.7826
$q^H$	\	\	1.0181	0.9958	0.9726	0.8670	0.8064	0.8000	0.7826	0.7826
$q^L$	\	\	0.3567	0.3702	0.3842	0.4452	0.5462	0.6607	0.7826	0.7826
Bi.T	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Info	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No

Table 2.1: **Markets Interaction.** This table studies the relationship between the trading limits,  $\underline{y}, \bar{y}$ , and bilateral trading activities. Parameters are  $\gamma = 0.8, \alpha = 0.1, \bar{z} = 0.3, \delta_I = 0.7$ . Each column represents a trading limit,  $[\underline{y}, \bar{y}]$ .  $\Pi$  is the profit of the Market maker, while  $p$  is the equilibrium market maker price in the centralized market.  $q^U, q^H, q^L$  are, respectively, the equilibrium bilateral prices that will be proposed by an uninformed trader, high-sigaled informed trader and low-sigaled informed trader. “Bi.T” indicates whether or not there exists trades in bilateral market, and “Info” represents the information diffusion in the bilateral market.

guish whether the rejection is because the signal is  $L$  or because his responder is uninformed. Learning is “partial” or “slower” in this case.

Therefore, in equilibrium, information diffuses in an asymmetric way. Good news spreads faster than bad news when agents are “biased” towards learning good news (signal  $H$ ), while bad news spreads faster than good news if agents are “biased” towards learning bad news (signal  $L$ ). Table 2.2 shows the information diffusion in the case when agents are “biased” towards learning  $H$  signals. As is shown in Table 2.2, when the true signal is  $H$ , an uninformed agent learns immediately when he is matched with an informed agent. But when the true signal is  $L$ , an uninformed agent learns the exact signal only when he is matched as a responder with an informed agent. If he is matched as a proposer, he can only partially learn the signal (i.e. the uninformed proposer’s belief drops from .5 to  $.5(1 - \alpha)$ , instead of 0, after observing the rejection from his responder.)



$S_I$	Fraction	Proposer	Responder	Learning	Belief		Portfolio	
					Before	After	Before	After
$H$	$\alpha^2$	Informed	Informed		1	1	(1,1)	(1,1)
	$\alpha(1 - \alpha)$	Informed	Uninformed	Yes(resp.)	.5	1	(1,1)	(1,1)
	$\alpha(1 - \alpha)$	Uninformed	Informed	Yes(prop.)	.5	1	(1,1)	$(1 - \bar{z}, 1 + q^U(p)\bar{z})$
	$(1 - \alpha)^2$	Uninformed	Uninformed		.5	.5	(1,1)	(1,1)
$L$	$\alpha^2$	Informed	Informed		0	0	(1,1)	(1,1)
	$\alpha(1 - \alpha)$	Informed	Uninformed	Yes(resp.)	.5	0	(1,1)	(1,1)
	$(1 - \alpha)$	Uninformed	I or U	Partial(prop.)	.5	.5(1 - $\alpha$ )	(1,1)	(1,1)

**Table 2.2: Information Diffusion When Uninformed Proposers Are “Biased” Towards Learning Good News.** This table studies the result of information diffusion after an active bilateral trading period, in a case when the uninformed proposers are “biased” towards learning signal  $H$ . “Fraction” is the population fractions of all matching scenarios. Yes(resp.) and Yes(prop.) represent the case that the uninformed agents learn the full signal as a responder and as a proposer, respectively. Partial(prop.) presents the case that the uninformed agents learn the signal partially as a proposer.

### 2.3.3 Market Maker's Profit and Optimal Trading Limits

In this section, we provide a numerical simulation of the market maker's profit as a function of the trading limits. Figure 2.3 shows that the market maker's profit exhibits an inverted U-shape as the trading limits are relaxed (i.e.  $[-\bar{y}, \bar{y}]$  becomes wider). This is consistent with our analysis. On the one hand, if the door is too narrow in the centralized market it prevents loss to the informed trader but also squeezes out uninformed traders, hence the centralized price and the market maker's profit is low. On the other hand, if the door is too wide the market maker would expose too much risk to informed traders which lowers the market maker's profit. It is possible that an optimal volume limitation exists. This is for our future studies.

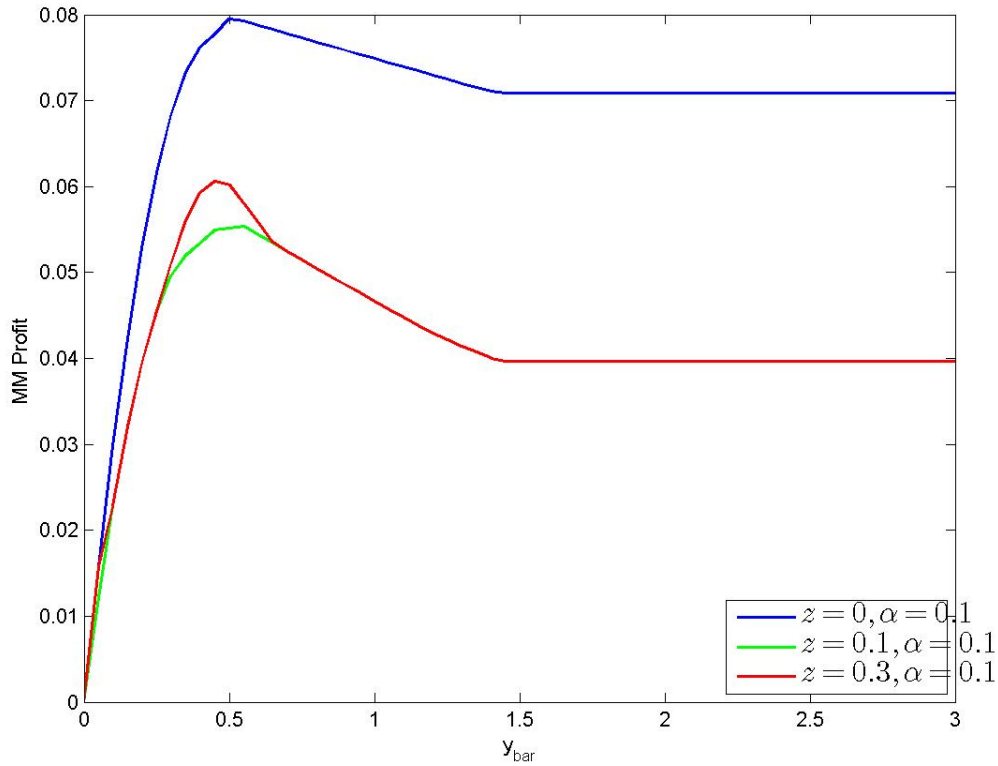


Figure 2.3: **The Market Maker's Profit As  $[-\bar{y}, \bar{y}]$  Varies.** The graph plots how the market maker's profit varies as  $[-\bar{y}, \bar{y}]$  becomes wider.

## 2.4 Illiquid Centralized Market and An Informed Market Maker

In the previous sections we analyzed information diffusion and market interaction between the centralized market and bilateral market, when the market maker obtains no information about the state of the world. In this section, we study the case when the market maker acquires some information about the asset's payoff. Particularly, we examine the following questions.

- (i) Is it in the interest of the market maker to reveal her information to the public?
- (ii) If the market maker decides to provide an informative public price, would this affect bilateral trading and information diffusion?

In this model, we consider a second source of information. We assume that the market maker receives a signal in the morning before she announces the centralized price  $p$ . We denote the signal as  $S_M \in \{H, L\}$  and its accuracy as  $\delta_M = Pr(S = s | S_M = s)$ . We also assume that the market maker's signal is less accurate than the informed agents (i.e.  $1/2 < \delta_M < \delta_I < 1$ ). Everything else stays the same with the model in the last section.

Having an informative market maker does not necessarily correlate to a decrease in the agents' welfare because the market maker may potentially make this privately owned information publicly available (i.e. to announce an informative public price  $p$ ). It is interesting to understand the relationship between public information, private information and social welfare in this case. And those analyses will be provided in our future studies.

Proposition 7 provides the condition when the market maker decides to reveal her own signal to the public by announcing an informative centralized price. That is, we examine the condition when there exist only separating equilibria.

### **Proposition 7. (An Informed Market Maker.)**

Suppose  $\alpha$  is sufficiently small, then there exists no pooling equilibrium. Hence the public

price  $p$  is informative, and in equilibrium  $p_H > p_L$ .

**Proof.** Suppose the market maker pools the price at  $p$ . Let  $\Pi_H(p)$  be the profit of the market maker who receives a signal  $S_M = H$ , and  $\Pi_L(p)$  be the profit of the market maker who receives a signal  $S_M = L$ . Suppose  $\alpha$  is sufficiently small. We now show that any  $p < 1$  fails the intuition criterion. Suppose instead  $p = \hat{p} + \epsilon$ , with  $\epsilon > 0$  and sufficiently small. Then

$$\Pi_H(p + \epsilon) \simeq (1 - \alpha)(1 - p)y_c^*(p, 1, 1; \delta_{HM}),$$

$$\Pi_L(p + \epsilon) \simeq (1 - \alpha)(1 - p)y_c^*(p, 1, 1; \delta_{LM}),$$

$$\Pi_{Pool}(p + \epsilon) \simeq (1 - \alpha)(1 - p)y_c^*(p, 1, 1; 0.5),$$

where

$$\delta_{HM} = Pr[S = H | b(S_M = H) = 1, b(S_I = H) = 0.5],$$

$$\delta_{LM} = Pr[S = H | b(S_M = L) = 1, b(S_I = H) = 0.5].$$

Then

$$\Pi_H(p + \epsilon) < \Pi_{Pool}(p + \epsilon) < \Pi_L(p + \epsilon),$$

by the fact that  $y_c^*(p, 1, 1; \delta_{HM}) < y_c^*(p, 1, 1; 0.5) < y_c^*(p, 1, 1; \delta_{LM})$ . Hence  $p$  fails the intuition criterion. For a separating equilibrium to exist, suppose  $p_H$  and  $p_L$  is the price the market maker poses when receiving an H signal and an L signal respectively. Then  $p_H, p_L$  must satisfy:

$$\Pi_H(p_H) \geq \Pi_H(p_L), \tag{2.12}$$

$$\Pi_L(p_L) \geq \Pi_L(p_H). \tag{2.13}$$

And

$$\Pi'_H(p_H) \simeq (1 - \alpha)((1 - p_H) \frac{\partial y_{cHM}^*}{\partial p} - y_{cHM}^*) = 0, \tag{2.14}$$

$$\Pi'_L(p_L) \simeq (1 - \alpha)((1 - p_L) \frac{\partial y_{cLM}^*}{\partial p} - y_{cLM}^*) = 0. \tag{2.15}$$

by condition (2.14), (2.15) and since  $y_{cHM}^* < y_{cLM}^*$ ,  $\frac{\partial y_{cHM}^*}{\partial p} > \frac{\partial y_{cLM}^*}{\partial p}$ . We have

$$p_H > p_L.$$

□

The market maker faces a trade-off. Revealing his information to the public may potentially decrease his profit because it adds to the accuracy of the informed traders' signals and increases the loss to them. But revealing his information also flattens the gap between the informed and uninformed traders and increases the profit. Proposition 7 states that, if there are not many informed traders, the benefit of revealing the information exceeds the loss.

We are then ready to answer the next question: if the market maker were to reveal his information, do we still observe bilateral trading and information acquisition? Proposition 8 provides us an insight into this question.

**Proposition 8.** If the market maker's signal is sufficiently accurate, then public signaling "squeezes" trade in the bilateral market. That is, there is no bilateral trading.

**Proof.** The unconditional probability that the market maker receives an  $H$  signal is 0.5. There are four cases. They all happen with probability 1/4. Let  $\delta_{S_M, S_I} = Pr(S = H | b(S_M) = 1, b(S_I = H) = 1)$ ,  $\delta_{S_I} = Pr(S = H | b(S_I) = 1)$ . Then we have

$$\delta_{HH} > \delta_{LH} > \delta_{HL} > \delta_{LL}.$$

Hence

$$y_{HH}^* < y_H^* < y_{LH}^* < y_{LH}^* < y_L^* < y_{LL}^*.$$

Consider the lower boundary  $y_0$  (the upper bound  $y_1$  is similar). If the market maker's signal is sufficiently accurate, it is more likely that

$$y_H^* \leq y_0 < y_{LH}^*.$$

If this is the case, no trade will happen in the bilateral period if  $S_M = L, S_I = H$ . Hence the public signal will discourage trade in the decentralized market. □

The result in Proposition 8 is consistent with the general conclusion from the literature regarding the relationship between public signals and private signals. As the public signals become more informative, the uninformed traders have less motivation to acquire private information. Information acquisition is costly: traders pay to learn the information. In this particular model there are costs involved when an uninformed proposer attempts to information from his informed responder's behavior. Therefore, if the public information is good enough, it discourages the uninformed traders' incentive to trade in the bilateral market.

## 2.5 Conclusion

In this paper we study information diffusion through bilateral trading and the interaction between the bilateral market and the centralized market. We find that the liquidity level in the centralized market matters. If the centralized market provides overly sufficient or overly insufficient liquidity (no trading volume limitations), no trade would be observed, and hence no information diffuses in the bilateral market. If the centralized market liquidity is controlled within a reasonable level there would be bilateral trades and information diffusion. While there exists active bilateral trading, private information diffuses in an asymmetric way. That is, good news could disperse faster than bad news, depending on the conjecture of the agents. Finally, when the price in the centralized market contains certain information, such public information could also dry out the bilateral trades.

The results provide empirical implications that we will examine in future studies. Also, the model could be extended in several other directions. Specifically, there are three directions we will explore in the future. First, we will relate our study to the empirical analysis in the pre-IPO and IPO markets, focusing on the traders' intention and behavior, information diffusion and the correlation between liquidity and informational efficiency. Second, we will further study the framework with a more actively managed centralized market, in which the market maker may not commit to his price and instead adapts the price to the market

information. Third, we may incorporate large traders to the model and study their strategy, price impact, and the consequent impact on information diffusion and market interaction.

## CHAPTER 3

# Screening Decisions with Time-Inconsistent Agents

### 3.1 Introduction

A significant amount of evidence in laboratory and the field suggests that, in some markets, consumers deviate from standard preferences and behave time-inconsistently. When evaluating an intertemporal contract, they usually give a higher discount between the present and the next period than between any of the subsequent periods. This discrepancy of the perception between the present and the future implies time inconsistency. Moreover, consumers are not fully aware of this time-inconsistency and exhibit different degrees of naiveté. Some consumers are more sophisticated and can predict more accurately their behavior than others.

Firms can profit from the time-inconsistency among consumers. The seminal paper by [DellaVigna and Malmendier \(2004\)](#) studies the optimal contract by a monopoly firm who interacts with a single type of time-inconsistent consumer. They show that firms set prices above marginal cost for goods with immediate rewards and deferred costs (*leisure goods*) and below marginal cost for goods with immediate costs and deferred rewards (*investment goods*). This result follows from the argument that a commitment device is valued by consumers and firms can profit from providing them.

[DellaVigna and Malmendier \(2004\)](#) does not, however, study the optimal contracting and screening problem between the monopoly firm and multiple types of consumers (unobservable degree of naiveté). In their analysis, consumers are either fully aware of the time-inconsistency (fully sophisticated) or partially aware (partially naive). The firm knows



the consumer's type at the contracting period. Nevertheless, in reality, the consumer type is private information. It raises a question of whether or not and why the firm would screen types of consumers when the information of consumer types is private and unobservable.

This paper studies the question of when and why a monopoly firm chooses to screen the type of a time-inconsistent consumer at the contracting period. The product is an investment good which involves an immediate cost (on the consumer) and a delayed benefit. There are 3 periods in the basic model: the contracting period, the action period, and the benefit period. There are multiple degree of consumer naiveté, i.e. the ability of a consumer to predict her future time-inconsistency. The firm can provide either a menu of two-part tariffs contracts (to screen her degree of naiveté) or a single two-part tariff contract (non-screening), subject to profit maximization.

We show that the firm's decision to screen depends on the uncertainty of the consumer's cost type. When the consumer's cost type is deterministic, the firm provides a non-screening contract, whereas when the consumer's cost type is random, the firm offers a screening contract. Moreover, in the screening contract, the firm provides the naive type a discount price (a price that is less than the firm's marginal cost) and the more sophisticated type a further discounted price. Finally, we extend the basic model to a multi-period model in which a consumer chooses between a cheaper entry fee and longer consumption periods. We show that the more sophisticated type chooses the contract that has a cheaper entry fee but shorter consumption periods.

The results from the basic 3-period model have three implications. First, screening is costly. If consumers have a large choice set, separation of different types becomes expensive and it is optimal for the firm not to screen. Second, the uncertainty for consumers limits their choice set and makes screening cheaper. Therefore, it is optimal for the firm to screen. When a consumer does not know exactly her cost type, the uncertainty deters her from behaving too aggressively. Consequently, it becomes easier for the firm to screen and profit from each type separately. Third, when there are multiple degrees of naiveté, the per-usage price serves as a commitment device. The more sophisticated type often desires a stronger

commitment device (a more discounted price) than the more naive type.

The results cast light on several practical contracting issues. As an example of the deterministic cost, a consumer's physical health condition is relatively stable, observable and therefore more predictable. Therefore, we observe fewer contracts in sports club (typically the club provides everyone a single price for a certain period of membership). In contrast with that, for a random cost case, a consumer's consumption and savings habit largely depends on the environment and is relatively difficult to predict. We observe abundant types of contracts for saving plans, retirement programs and etc.

The paper contributes to the literature on the study of time-inconsistency and self-control. Our paper is closely related with [DellaVigna and Malmendier \(2004\)](#) which studies the contracting theory of a single-type consumer, whereas our paper studies screening and contracting with an unobservable multi-type consumer. Our paper is also closely related with [Eliaz and Spiegel \(2006\)](#) which analyzes the contracting of a principal with diversely naive agents. Our contexts of consumer preferences and time-inconsistency are different from it. Other papers study various issues of time-inconsistency (e.g., [DellaVigna and Malmendier 2004](#), [O'Donoghue and Rabin 1999](#), [Shui 2004](#), [Gottlieb 2008](#), [Sayman and Öncüler 2008](#), [Wong 2008](#), [Incekara-Hafalir 2014](#), [Galperti 2015](#)) and self-control (e.g., [O'Donoghue and Rabin 1999](#), [Esteban et al. 2007](#), [Sanjay 2009](#), [Heidhues and Kszegi 2010](#), [Alexandrov 2016](#)). Different from the previous ones, our paper focuses on the screening and optimal contract problem in the presence of time-inconsistency.

The paper also contributes to the study of contract theory and information asymmetry. For example, [Incekara-Hafalir \(2014\)](#), [Li et al. \(2014\)](#), [Ylmaz \(2015\)](#), [Heidhues and Koszegi \(2014\)](#), [Sandroni and Squintani \(2013\)](#) and [Zhang \(2012\)](#) study the contracting problem issue when there are various types of information asymmetry. Our paper focuses on the role information asymmetry plays in a firm's screening decision.

The organization of the paper is as follows. Section 2 describes the model. Section 3 studies the 3-period screening and contracting problem when the cost incurred is deterministic. Section 4 studies the 3-period screening and contracting problem when the cost is

random. Section 5 extends the model to multiple periods. Section 6 concludes.

## 3.2 The Model

In this section we consider a three period model of a monopoly firm who arranges pricing schemes to sell an investment good, a product with immediate cost and delayed benefit. Consumers are time-inconsistent and (partially) naive. We consider two different cases: a case where the cost of the product is deterministic and a case where the cost of the product is random. We then analyze the firm's optimal pricing and screening strategy in each case.

### 3.2.1 Model Set-up

We follow the standard models of the literature in time-inconsistency (eg. [DellaVigna and Malmendier \(2004\)](#)). There is a monopolist firm. The selling of the product involves 3 periods. In  $t = 0$  the firm proposes a (menu of) two-part tariff contracts  $(L, p)$  to the consumer, where  $L$  represents the entry price and  $p$  the per usage fee. Upon the proposal, the consumer decides to sign or to reject the contract(s). If a menu of contracts had been offered, the consumer may choose one of the contracts in the menu or reject all of them. If he chooses to reject all, the consumer attains the reservation utility 0 at  $t = 1$ . In  $t = 1$ , upon accepting a contract, the consumer pays the entry fee  $L$  as formalized in the contract. Then the consumer chooses to consume or not to consume. If he chooses to consume, he pays the per usage fee  $p$  to the firm and incurs a personal cost  $c$  in  $t = 1$ ; he then receive benefit  $b$  in  $t = 2$ . If he chooses not to consume, he attains payoff 0 in  $t = 1$  and  $t = 2$ . The timeline of the decisions is described in Figure 1.

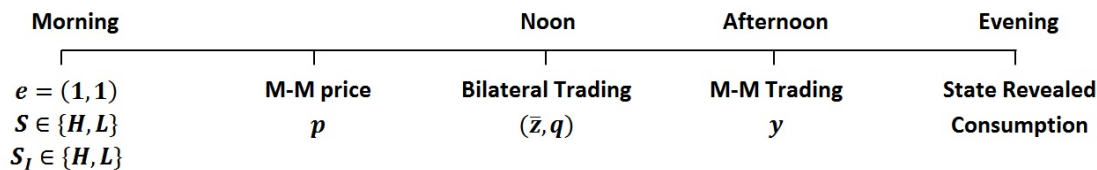


Figure 3.1: Timeline

**Product.** The product we consider in the model is an *investment good*. An investment good is a product which incurs an immediate cost and a delayed benefit. In reality, the cost and benefit of a product do not usually take place at the same time. When a good is costly at present and provides benefits in the future, the good is called an *investment good*. Examples of an investment good include health clubs, sports, education, retirement plans, etc. In contrast with that, a good can provide immediate benefit (pleasure) but incurs cost in the future. Such good is called a *leisure good*. Such goods include, for example, consumption of unhealthy food, expensive cellphone plans, gambling, procrastination, etc<sup>1</sup>. In this paper, we restrict our attention to the screening problem that is associated with an investment good. In the model, upon consumption, the product has a cost  $c$  at  $t = 1$  and a delayed benefit  $b$  at  $t = 2$ .

**Time-inconsistent preferences** Consumers have time-inconsistent preferences. The intertemporal preferences of consumers are assumed to be quasi-hyperbolic (see, e.g., [DellaVigna and Malmendier 2004](#), [O'Donoghue and Rabin 1999](#)). That is, the discount factor of utility flow for time  $s$ , when evaluated at time  $t$ , equals 1 if  $s = t$  and equals  $\beta\delta^{s-t}$  if  $s > t$ , with  $\beta \leq 1, \delta \leq 1$ . Therefore, the present value of the future utility flow  $(u_s)_{s \geq t}$  as of time  $t$  is

$$u_t + \beta \sum_{s=t+1}^{\infty} \delta u_s. \quad (3.1)$$

In correspondence with the standard model of time-inconsistent preferences,  $\beta$  can be interpreted as the short-run discount factor and  $\delta$  the long-run discount factor. The term *time-inconsistency* corresponds to the case when  $\beta$  is less than 1. It reflects the consumers' behavior to perceive differently in the present and in the future—while the discount factor between any adjacent periods in the future is simply  $\delta$ , the discount factor between the present and the next period is  $\beta\delta$ . In comparison with that, the standard *time-consistency* model corresponds to the case when  $\beta$  equals 1, when there is no difference between the short-run and the long-run discount factor.

---

<sup>1</sup>The definition of an investment good and a leisure good is consistent with [DellaVigna and Malmendier \(2004\)](#)

**Consumer types and naiveté.** A *partially naive* consumer is not fully aware of his time-inconsistency and overestimates his time consistency. The (partial) naiveté is reflected by the set of parameters  $(\beta, \hat{\beta})$ , where  $\beta$  is the actual short-run discount factor and  $\hat{\beta}$  is the perceived short-run discount factor. That is, he or she expects to have the discount factors  $1, \hat{\beta}\delta, \hat{\beta}\delta^2, \dots$  with  $\beta \leq \hat{\beta}_1 < \hat{\beta}_2 \leq 1$  in all future periods. The difference between the perceived and actual short-run discount factor,  $\hat{\beta} - \beta$  is the *degree of naiveté*, which measures the degree of how much the consumer could not anticipate his time-inconsistency. There are three special cases. When  $\hat{\beta} = \beta = 1$ , the consumer is an *exponential* consumer who is time consistent ( $\beta = 1$ ) and is aware of it ( $\hat{\beta} = 1$ ). When  $\hat{\beta} = \beta < 1$ , the consumer is called a *sophisticated* consumer, who is time-inconsistent ( $\beta < 1$ ) and is aware of it ( $\hat{\beta} = \beta$ ). When  $\beta < \hat{\beta} < 1$ , the consumer is called a fully *naive* consumer, who is time-inconsistent ( $\beta < 1$ ) and is unaware of it at all ( $\hat{\beta} > \beta$ ).

We assume that there are two types of time-inconsistent consumers (agents) in the market. Type 1 with parameters  $(\beta, \hat{\beta}_1)$  accounts for  $\alpha$  of the population and type 2 with parameters  $(\beta, \hat{\beta}_2)$  accounts for  $1 - \alpha$  of the population. We assume that  $\beta \leq \hat{\beta}_1 < \hat{\beta}_2 \leq 1$ . That is, type 1 is more sophisticated than type 2 consumers. Therefore, in this paper, we refer to type 1 consumers as the "more sophisticated" consumers and type 2 as the "more naive" consumers.

**Monopoly firm, contracts, and screening.** The firm has all the bargaining power and offers the consumer nonnegotiable contract(s) in period 0. A contract  $(L, p)$  is composed of an entry fee,  $L$ , and a per-usage price  $p$ . Since the firm cannot distinguish the types of consumers, the firm can choose either to screen or not to screen the types of consumers at  $t = 0$ , subject to profit maximization. In order to screen, the firm offers a menu of contracts to a consumer at  $t = 0$ . In a successful screening case, the result of the consumer's selection among these contracts reveals his type. When the firm decides not to screen the consumers, she offers a single contract to the consumer.

We assume that the firm has no start-up cost providing the product but has a per-usage cost,  $a$ , when a consumer decide to use the product.

**Information about the cost of consumption ( $c$ ).** If a consumer decides to consume the good at  $t = 1$ , he incurs a personal cost  $c$  and pays the per-usage price  $p$ . For example, if an individual who signs up a gym-club decides to use the equipment in the gym, he not only pays the fees for the equipment per usage ( $p$ ), but also pays personal costs for such usage (time, effort, for example).

Consumers may or may not know his or her personal consumption cost  $c$  prior to signing the contract. We consider two cases. In the first case, the consumption cost is stable over time and the information can be estimated easily through personal experience and knowledge. For example, costs that attribute to physical health (body type, weight, health condition and etc.) seem to be easily evaluated. Therefore the information about the cost of consumption is can be regarded as known before signing the contract. In the second case, however, consumers may find it difficult to estimate accurately his consumption cost type  $c$  at the time of signing the contract. For example, when making retirement plans, consumers' consumption habits may vary over time and largely depends on the realization of the economy.

To summarize the above argument, we analyze the firm's screening problem, respectively, under the following two different assumption.

**(i) Case I (Deterministic cost):** A consumer knows his consumption cost  $c$  prior to signing the contract.

**(ii) Case II (Random cost):** A consumer's cost type  $c$  is drawn from a distribution  $F$ . We assume that  $F$  has a strictly positive density function  $f$  over  $\mathbb{R}$ . At  $t = 0$ , consumers do not know their individual costs  $c$ . A consumer's cost type  $c$  is revealed upon accepting a contract offered by the firm.

Finally, the monopoly firm is fully aware of consumers' time-inconsistency behavior, the distribution of the naiveté within the population, and the distribution of consumers' cost. Section 3.3 summarizes the analysis of case I and Section 3.4 analyzes the screening problem of case II.

### 3.3 Deterministic Cost

In this section we characterize optimal contracts and the firm's screening problem in the 3-period model when the consumer's personal consumption cost is known at the time he signs the contract.

A time-consistent consumer (with parameters  $\beta = 1, \hat{\beta} = 1$ ) evaluates the consumption at  $t = 0$  as follows. Because a time-consistent consumer does not have a discrepancy between the short-run and long-run adjacent discount factor ( $\beta = 1$ ), He discounts by  $\delta$  the total cost of consumption  $c + p$  at  $t = 1$  and by  $\delta^2$  the delayed benefit  $b$  received at  $t = 2$ . Therefore, the present value of the utility of "consume" is  $\delta(\delta b - p - c)$ . The present value of "not consume" is 0. At  $t = 1$ , the consumer chooses to consume if  $c \leq \delta b - p$  and not to consume if  $c > \delta b - p$ . Therefore, at  $t = 0$ , the value of a contract  $(L, p)$  for a time-consistent consumer with parameter  $(\beta, \hat{\beta})$  is

$$U_i((L, T), (1, 1)) = \begin{cases} \delta(-L) & \text{if } c > \delta b - p, \\ \delta(-L + \delta b - p - c) & \text{if } c \leq \delta b - p. \end{cases}$$

A naive time-inconsistent consumer, however, evaluates the consumption differently. First, a time-inconsistent consumer consumes less than he had expected. At the moment of deciding between consuming and not consuming, the net payoff of "consume" equals  $\beta\delta b - p - c$ . Therefore, at  $t = 1$ , he chooses to consume if  $c \leq \beta\delta b - p$ . That is to say, a time-inconsistent consumer chooses to consume less often than a time-consistent consumer. The short-run discount factor,  $\beta$ , determines the inconsistency between the actual and expected likelihood of consumption. Such difference equals 0 if  $\beta = 1$  for a time-consistent consumer. The smaller  $\beta$  is, the larger the issue of time-inconsistency is. Second, a naive time-inconsistent consumer is not fully aware of his time-inconsistency. He discounts by  $\beta$  his next period utility, but he overestimates his future short-run discount factor. That is, at  $t = 0$ , he expects that he will consume at  $t = 1$  if  $\hat{\beta}\delta b - p - c \geq 0$ . However, when it comes to  $t = 1$ , he will consume if  $\beta\delta b - p - c \geq 0$ . Therefore, at  $t = 0$ , a naive consumer overestimates his probability of consumption. The value of a contract  $(L, p)$  for a naive time-inconsistent

consumer  $i$ ,  $i = 1, 2$ , with parameters  $(\beta_i, \hat{\beta}_i)$  is

$$U_i \left( (L, p), (\beta_i, \hat{\beta}_i) \right) = \begin{cases} \beta_i \delta(-L) & \text{if } c > \hat{\beta}_i \delta b - p, \\ \beta_i \delta(-L + \delta b - p - c) & \text{if } c \leq \hat{\beta}_i \delta b - p. \end{cases}$$

The monopoly firm, on the other hand, has complete knowledge about consumers' behavior. The firm correctly expects that the consumer would value the consumption inconsistently in period 0 and period 1. In the spirit of profit-maximization, the firm can decide to screen by offering a menu of contracts or not to screen by offering a single contract. We now proceed to formalize the screening problem and the non-screening problem in the analysis.

### 3.3.1 Deterministic Cost: screening Contract

The monopoly firm can screen the type of a consumer by offer her a menu of contracts. Since there are two types of consumers in the market, with type 1 more sophisticated than type 2, in order to screen the firm can offer two separating contracts,  $(L_1, p_1)$ , and  $(L_2, p_2)$ . The screening holds when type 1 consumers choose the first contract and type 2 consumers choose the second contract.

To satisfy type 1's IC constraint and both agents' IR constraint, we must have

$$p_1 \leq \hat{\beta}_1 \delta b - c < p_2 \leq \hat{\beta}_2 b \delta - c.$$

The first and third inequality must hold because contract 1 and 2 must be designed to make the consumption attractive to type 1 and type 2 consumers respectively. The second strict inequality must hold in order to make the second contract attractive to only type 2 agents.

The firm's profit is then

$$\Pi = \alpha(L_1 + (p_1 - a)1_{\{p_1 < \beta \delta b - c\}}) + (1 - \alpha)L_2. \quad (3.2)$$

Since  $p_2 > \hat{\beta}_1 \delta b - c > \beta \delta b - c$ , in the firm's expectation, type 2 agent will not consume at  $t = 1$ . Whether type 1 agent consumes or not at  $t = 1$  depends on  $p_1$ . The firm's profit



maximization problem is then given by

$$\max_{(L_1, p_1), (L_2, p_2)} \alpha(L_1 + (p_1 - a)1_{\{p_1 < \beta\delta b - c\}}) + (1 - \alpha)L_2$$

s.t.

$$U_1(L_1, p_1, (\beta, \hat{\beta}_1)) \geq U_1(L_2, p_2, (\beta, \hat{\beta}_1)), \quad (\text{IC1})$$

$$U_2(L_2, p_2, (\beta, \hat{\beta}_2)) \geq U_2(L_1, p_1, (\beta, \hat{\beta}_2)), \quad (\text{IC2})$$

$$U_1(L_1, p_1, (\beta, \hat{\beta}_1)) \geq 0, \quad (\text{IR1})$$

$$U_2(L_2, p_2, (\beta, \hat{\beta}_2)) \geq 0. \quad (\text{IR2})$$

We now proceed to prove that screening is implementable and a separating equilibrium exists. We assume first that IR1 and IC2 binds. Note that once the agent accepts the contract, the utility  $U_i(L, p)$  will be independent of  $i$ . Hence IC2 binding will imply IR2 binding. IC1 is clearly satisfied. Hence in optimum we will have, for all  $i = 1, 2$ ,

$$-L_i + \delta b - p_i - c = 0. \quad (3.3)$$

Plugging (3.3) into the profit function (3.2), the firm's problem becomes

$$\max_{p_1, p_2} \alpha(\delta b - p_1 - c + 1_{\{p_1 < \beta\delta b - c\}}(p_1 - a)) + (1 - \alpha)(\delta b - p_2 - c)$$

s.t.

$$p_1 \leq \hat{\beta}_1 \delta b - c < p_2 \leq \hat{\beta}_2 b \delta - c.$$

The result of the optimal separating contracts is summarized as follows.

**Proposition 9. (Deterministic cost, screening contract).** When the consumer's personal consumption cost is deterministic, there exists optimal screening contracts  $(L_1, p_1), (L_2, p_2)$ , in which

$$p_2 = \hat{\beta}_1 \delta b - c + \epsilon, \quad (3.4)$$

$$p_1 = \begin{cases} \beta \delta b - c & \text{if } a \geq \beta \delta b - c, \\ \text{any number } < \beta \delta b - c & \text{if } a < \beta \delta b - c, \end{cases} \quad (3.5)$$

$$L_i = \delta b - p_i - c, i = 1, 2. \quad (3.6)$$

And the least upper bound of firm profit is:

$$\Pi = \begin{cases} \alpha(\delta b - \beta\delta b) + (1 - \alpha)(\delta b - \hat{\beta}_1\delta b) & \text{if } a + c \geq \beta\delta b, \\ \alpha(\delta b - (a + c)) + (1 - \alpha)(\delta b - \hat{\beta}_1\delta b) & \text{if } a + c < \beta\delta b. \end{cases} \quad (3.7)$$

**Remark 3.** We evaluate the firm's maximum profit by the least upper bound because, in order to fully screen types of consumers, the IC constraint for type 2,  $\hat{\beta}_1\delta b - c < p_2$ , must be a strict inequality. Since the firm performs better by lowering  $p_2$ , if we allow weak inequality, in equilibrium IC1 and IC2 will both be binding, but then we will not be able to guarantee the separation.

In an equilibrium when type 1 (more sophisticated) consumers decide to consume in period 1, we have that  $p_2 > p_1$  and  $a \geq \beta\delta b - c = p_1$ . That is, in such equilibrium the more sophisticated consumer prefers a contract with a lower per-usage fee, and, furthermore, the per-usage fee for an sophisticated consumer must be lower than the marginal operation cost of the firm  $a$ .

The reason why  $p_1 \leq a$  is similar to that of [DellaVigna and Malmendier \(2004\)](#): in equilibrium, the more sophisticated consumer is more aware of his time-inconsistency and has less overestimation of his future usage of the product. Therefore, the firm lowers the price  $p_1$  below  $a$  to the extent that the user is conscious about his future time inconsistency. The difference between the per-usage price and the marginal cost,  $p_1 - a$ , can be regarded as a commitment device for the more sophisticated consumer. It provides an additional utility to guarantee a consumption for the more sophisticated consumer.

For the more naive consumer, however, the per usage price  $p_2$  may or may not exceed the firm's marginal cost  $c$ . Also, in the separating equilibrium,  $p_2 > p_1$ . To understand this, we can perceive the firm's profit maximization as decisions of choosing between screening consumers and luring more naive consumers. As we have discussed, a more naive consumer may overestimate his probability of future consumption, thus offering him a discount  $p$  for sure (and an increase in  $L$  relative to the contract) will increase the sign-up profit from the more naive consumers but compromise the efficacy of screening types of consumers. In order

to distinguish consumer types, the firm needs to offer the more naive type a relatively low entry fee and higher per-usage price. Therefore, screening the types of consumers costs the firm profits from signing up the contracts.

Now we analyze the optimal contract and the firm's maximum profit in the pooling equilibrium when the firm offers a single contract.

### 3.3.2 Deterministic Cost: non-screening contract

If the firm decides not to screen types of consumers, she offers a single contract to any consumer. There are two possibilities to consider. First, the firm aims for only type 2 consumers and disregards the other. That is, at  $t = 0$ , the firm provides a contract  $(L, p)$  that yield positive net payoff for type 2 consumers but negative payoff for type 1 consumers. We then have  $\hat{\beta}_1 \delta b - c < p$ . This is not optimal: the monopoly firm can profit by lowering  $p$  and raising  $L$  by the same amount, to the extent that type 1 consumers find it also acceptable.

Therefore, the optimal pooling contract  $(L, p)$  must satisfy

$$p < \hat{\beta}_1 \delta b - c.$$

If both types sign the same contract, the firm's profit function is then

$$\Pi = L + 1_{\{p < \beta \delta b - c\}}(p - a).$$

Note that  $1_{\{p < \beta \delta b - c\}}$  refers to the fact that, at  $t = 1$ , consumption takes place only when  $p \leq \beta \delta b - c$ , for both type 1 and type 2 consumers. The only constraint is the IR constraint,  $-L + \delta b - p - c = 0$ . It is binding under the profit-maximization assumption. Therefore,

$$-L + \delta b - p - c = 0.$$

Substituting this into the firm's profit function, the maximization problem is then given by

$$\max_p \delta b - p - c + 1_{\{p < \beta \delta b - c\}}(p - a).$$

We can then solve for the optimal contract. The result is summarized as follows.

**Proposition 10. (Deterministic cost, non-screening contract).** When consumers' personal consumption cost is deterministic, there exists an optimal non-screening contract  $(L, p)$  in which

$$p = \begin{cases} \text{Any number} < \beta\delta b - c & \text{if } a + c < \beta\delta b, \\ \beta\delta b - c & \text{if } a + c \geq \beta\delta b. \end{cases} \quad (3.8)$$

$$L = \delta b - p - c. \quad (3.9)$$

And the firm profit is

$$\Pi = \begin{cases} \delta b - (a + c) & \text{if } a + c < \beta\delta b, \\ \delta b - \beta\delta b & \text{if } a + c \geq \beta\delta b. \end{cases} \quad (3.10)$$

In an optimal non-screening contract, when  $a + c \leq \beta\delta b$ , the firm offers a discounted  $p$  in which the per-usage price,  $p$ , is lower than the marginal cost,  $a$ . This is consistent with our previous analysis and the argument of [DellaVigna and Malmendier \(2004\)](#). The discounted  $p$  is employed by the firm for two reasons: a) it is a commitment device designed for more sophisticated consumers. The more sophisticated consumers have sufficient awareness of their overconfidence and hence need a lower  $p$  to realize their commitment for the consumption. b) it is a device to lure more naive consumers. The more naive consumers overestimate their probability of consumption and hence desire a lower  $p$ . The firm can then profit from lowering  $p$  and increasing  $L$  relative to the contract.

### 3.3.3 Deterministic Cost: to screen or not to screen?

In this section, we compare the firm's profit from an optimal screening and an optimal non-screening contract. It is summarized as follows.

**Corollary 4. (Deterministic cost).** When the consumers' personal consumption cost is deterministic, offering a non-screening two-part tariff contract yields a higher profit for the firm. Furthermore, in the optimal non-screening two-part tariff contracts,  $p \leq a$ .

Corollary 4 characterizes the phenomenon that, when consumers know their cost type before signing the contract, a monopoly firm offers only a single non-screening contract. This

can be observed in our daily life. For example, in a typical health club contract, users pay the same flat fees but no price per visit. Although there might be multiple levels of naiveté among consumers, the club typically offers a single annual fee contract. Moreover, health club users pay flat fees but no price per visit, despite the club bearing marginal costs per attendance (see, e.g., [DellaVigna and Malmendier 2006](#)).

We can associate this intuition with [Esteban et al. \(2007\)](#) who consider the interaction between a monopoly firm and consumers. Consumers have temptation and it is psychologically costly to exercise self-control. They show that the optimal contract motivates the firm to offer a relatively flat and compact price schedule, serving more customers with low demand. The main variable they study is the variation of the direction of temptation. When a complicated menu of contract is offered (consumers are tempted), some consumers may be tempted upward (desire the goods more) while others are tempted downward (desire the goods less). A narrower selection (simple menu) is more profitable when consumers are tempted in the same direction. Similarly, in our model, when a second contract with lower  $p$  is presented, both types of consumers are tempted in the same way since their cost of consumption is known at the decision point. Such unanimity can be exploited by the firm by providing less selections (single contract) to the consumers.

### 3.4 Random Cost

In this section, we explore the case when the consumer's personal cost of consumption is unknown before signing the contract. We assume that, for all consumers, such cost is drawn from a distribution  $F$ . At  $t = 0$ , a partially naive individual with parameters  $(\beta, \hat{\beta}_i)$  overestimates the probability that his future self will consume the product at  $t = 1$ . He expects that he will consume if  $\beta_i \delta b - p - c \geq 0$ , i.e., with probability  $F(\hat{\beta}_i \delta b - p)$ . The actual probability of consumption, however, is  $F(\beta_i \delta b - p)$ . The difference between the forecasted and actual consumption probability,  $F(\hat{\beta}_i \delta b - p)$  measures the degree of *overconfidence*. A fully sophisticated consumer  $\hat{\beta}_i = \beta_i$  has no overconfidence.

Thus, a naive consumer with parameters  $(\beta, \hat{\beta}_i)$  evaluates a contract  $(L, p)$  using

$$U_i^R \left( (L, p), (\beta, \hat{\beta}_i) \right) = \beta \delta \left( -L + \int_{-\infty}^{\hat{\beta}_i \delta b - p} \delta b - p - c dF(c) \right). \quad (3.11)$$

If the monopoly firm is to screen types of consumers, she offers two different contracts  $(L_1, p_1)$  and  $(L_2, p_2)$ . The firm's profit-maximization problem is given in section 3.3.1 except that the utility functions are under the random cost evaluation as defined in (3.11).

To solve the random cost profit-maximization problem, we first establish the single crossing property. For simplicity, we twist the notation and use  $u_i(L_j, p_j)$  to represent the evaluation of a type  $i$  consumer on a contract  $(L_j, p_j)$ .

**Definition 4. (SCP)** Let  $(L_1, p_1), (L_2, p_2)$  be two contracts such that  $p_1 < p_2$ . The preferences satisfy the single crossing property if

$$u_1(L_1, p_1) \leq u_1(L_2, p_2)$$

implies that

$$u_2(L_1, p_1) < u_2(L_2, p_2).$$

The definition implies that type 2 has a stronger preference over the per-usage price  $p$ .

In order to guarantee existence of a profit-maximizing screening contract, we introduce a technical assumption that we maintain through the rest of the paper.

**Assumption 1.** We assume that the density function  $f(c)$  is decreasing in  $c$ .

Assumption 1 rules out irregularity on the tails of  $f(c)$ . There are many distribution functions that satisfy this property. A typical decreasing density distribution is the exponential distribution function:

$$f(c) = \begin{cases} \lambda e^{-\lambda c} & c \geq 0, \\ 0 & c < 0. \end{cases}$$

Lemma 14 and Lemma 15 establish the SCP.

**Lemma 14.** Let  $f(c)$  be the density of  $F(c)$ . Assume  $f(c)$  is decreasing. Then

$$\phi(p) := \int_{\hat{\beta}_1 \delta b - p}^{\hat{\beta}_2 \delta b - p} -p - c + \delta b dF(c)$$

is increasing in  $p$ .

**Proof.** Note that

$$\phi'(p) = -(\delta b - \hat{\beta}_2 \delta b) f(\hat{\beta}_2 \delta b - p) + (\delta b - \hat{\beta}_1 \delta b) f(\hat{\beta}_1 \delta b - p) - \left( F(\hat{\beta}_2 \delta b - p) - F(\hat{\beta}_1 \delta b - p) \right).$$

By the mean value theorem, there exists  $\tilde{x} \in (\hat{\beta}_1 \delta b - p, \hat{\beta}_2 \delta b - p)$  such that

$$F(\hat{\beta}_2 \delta b - p) - F(\hat{\beta}_1 \delta b - p) = f(\tilde{x})(\hat{\beta}_2 \delta b - \hat{\beta}_1 \delta b).$$

Since  $f$  is decreasing, we then have

$$\begin{aligned} F(\hat{\beta}_2 \delta b - p) - F(\hat{\beta}_1 \delta b - p) &< f(\hat{\beta}_1 \delta b - p)(\hat{\beta}_2 \delta b - \hat{\beta}_1 \delta b) \\ &< -(\delta b - \hat{\beta}_2 \delta b) f(\hat{\beta}_2 \delta b - p) + (\delta b - \hat{\beta}_1 \delta b) f(\hat{\beta}_1 \delta b - p), \end{aligned}$$

which shows that  $\phi(p) > 0$ . □

**Remark 4.** The slope of  $\phi(p)$  can be interpreted as the ability of the firm to distinguish the types of consumers. As  $\phi'(p)$  becomes larger, a small change of the price  $p$  can yield a larger difference of the two types of consumers' utility, and hence a higher ability of the firm to differentiate the consumer types. Thus we expect that a higher  $\phi'(p)$  implies a larger profitability for the firm.

Therefore,

**Lemma 15.** If  $f(c)$  is decreasing, then the preferences satisfy SCP of consumer preferences.

**Proof.**

$$\begin{aligned}
u_2(L_2, p_2) &= \beta\delta \left( -L_2 + \int_{-\infty}^{\hat{\beta}_2\delta b - p_2} -p_2 - c + \delta bdF(c) \right) \\
&= \beta\delta \left( -L_2 + \int_{-\infty}^{\hat{\beta}_1\delta b - p_2} -p_2 - c + \delta bdF(c) + \int_{\hat{\beta}_1\delta b - p_2}^{\hat{\beta}_2\delta b - p_2} -p_2 - c + \delta bdF(c) \right) \\
&\geq \beta\delta \left( -L_1 + \int_{-\infty}^{\hat{\beta}_1\delta b - p_1} -p_1 - c + \delta bdF(c) + \int_{\hat{\beta}_1\delta b - p_2}^{\hat{\beta}_2\delta b - p_2} -p_2 - c + \delta bdF(c) \right) \\
&> \beta\delta \left( -L_1 + \int_{-\infty}^{\hat{\beta}_1\delta b - p_1} -p_1 - c + \delta bdF(c) + \int_{\hat{\beta}_1\delta b - p_1}^{\hat{\beta}_2\delta b - p_1} -p_1 - c + \delta bdF(c) \right) \\
&= u_2(L_1, p_1).
\end{aligned}$$

The third inequality follows from  $u_1(L_1, p_1) \geq u_1(L_2, p_2)$ , and the fourth follows from Lemma 14 and that  $p_1 < p_2$ .  $\square$

Given the SCP of consumer preference, the local downward IC constraint implies the satisfaction of the local upward IC constraint. Moreover, in the optimal contract, downward IC and low type IR are binding. Therefore, the firm's maximization problem can be written as

$$\max_{(L_1, p_1), (L_2, p_2)} \alpha (L_1 + F(\beta\delta b - p_1)(p_1 - a)) + (1 - \alpha) (L_2 + F(\beta\delta b - p_2)(p_2 - a))$$

subject to

$$-L_1 + \int_{-\infty}^{\hat{\beta}_1\delta b - p_1} -p_1 - c + \delta bdF(c) = 0, \quad (\text{IR1})$$

$$-L_2 + \int_{-\infty}^{\hat{\beta}_2\delta b - p_2} -p_2 - c + \delta bdF(c) = -L_1 + \int_{-\infty}^{\hat{\beta}_2\delta b - p_1} -p_1 - c + \delta bdF(c). \quad (\text{IC2})$$

Given the SCP, we are ready to capture the key feature of the firm's optimal contracts in Proposition 11

**Proposition 11. (Random Cost)** When the consumer's personal consumption cost is randomly drawn from a distribution  $F$  and Assumption 1 is satisfied, the firm's optimal two-part tariffs contracts  $(L_1, p_1)$  and  $(L_2, p_2)$  must satisfy



(i)  $L_1 \geq L_2$ , and  $p_1 \leq p_2$ .

(ii)  $p_2 < a$ .

**Proof.** (i): By the SCP.

(ii). The first order condition yields

$$(p_2 - a) = -(1 - \hat{\beta}_2)\delta b \frac{f(\hat{\beta}_2\delta b - p_2)}{f(\beta\delta b - p_2)} - \frac{F(\hat{\beta}_2\delta b - p_2) - F(\beta\delta b - p_2)}{f(\beta\delta b - p_2)}, \quad (3.12)$$

and

$$\begin{aligned} (p_1 - a) + (1 - \hat{\beta}_1)\delta b \frac{f(\hat{\beta}_1\delta b - p_1)}{f(\beta\delta b - p_1)} + \frac{F(\hat{\beta}_1\delta b - p_1) - F(\beta\delta b - p_1)}{f(\beta\delta b - p_1)} = \\ (1 - \alpha) \left[ (p_1 - a) + (1 - \hat{\beta}_2)\delta b \frac{f(\hat{\beta}_2\delta b - p_1)}{f(\beta\delta b - p_1)} + \frac{F(\hat{\beta}_2\delta b - p_1) - F(\beta\delta b - p_1)}{f(\beta\delta b - p_1)} \right]. \end{aligned} \quad (3.13)$$

Since  $0 < \beta < \hat{\beta}_2 \leq 1$ , the RHS of (3.12) is negative,  $p_2 < a$ .

□

In the optimal screening contracts, the firm provides both types of consumers a discounted per-usage price ( $p_1, p_2 < a$ ), with the more sophisticated consumer a higher entry fee ( $L_1 > L_2$ ) and a lower per-usage price ( $p_1 < p_2$ ), compared with the more naive consumer. The intuition is consistent with the deterministic case. Since consumers are time-inconsistent and tend to overestimate their future consumption probability, the firm provides a discounted per-usage price ( $p_1, p_2 < a$ ) to lure the consumers to sign the contracts. Since the more sophisticated consumer recognizes more about her own time-inconsistency, the firm must provide a further discount ( $p_1 < p_2$ ) to serve as a commitment device for the more sophisticated type.

### 3.4.1 Random Cost: to screen or not to screen

The firm solves the profit-maximizing problem described above and decides whether to provide a screening contract or non-screening contract. That is, if the solution of the problem satisfies  $p_1 < p_2$ , then a screening contract is more profitable for the firm. If the solution satisfies  $p_1 = p_2$ , then a non-screening contract is more profitable. While we did not obtain a full analytic condition of whether screening is more profitable, Proposition 12 provides a sufficient condition of when a screening contract dominates a non-screening contract.

**Proposition 12. (Random cost, to screen or not to screen)** When the consumer's personal consumption cost is randomly drawn from a distribution  $F$  and Assumption 1 is satisfied, if

$$G(p) := \left[ F(\hat{\beta}_2 \delta b - p) - F(\hat{\beta}_1 \delta b - p) \right] - \delta b \left[ (1 - \hat{\beta}_1) f(\hat{\beta}_1 \delta b - p) - (1 - \hat{\beta}_2) f(\hat{\beta}_2 \delta b - p) \right]$$

has no root, then a profit-maximizing screening contract yields more profit than a non-screening contract.

**Proof.** If the profit-maximizing contract is a non-screening contract, we have

$$p_1 = p_2 = p.$$

Also, by optimality, (3.12),(3.13) must be satisfied. Combining with  $p_1 = p_2 = p$ , we have that

$$0 = \left[ F(\hat{\beta}_2 \delta b - p) - F(\hat{\beta}_1 \delta b - p) \right] - \delta b \left[ (1 - \hat{\beta}_1) f(\hat{\beta}_1 \delta b - p) - (1 - \hat{\beta}_2) f(\hat{\beta}_2 \delta b - p) \right].$$

□

Proposition 12 characterizes the phenomenon that, under certain conditions, when consumers do not know their cost types before signing the contract, it is optimal for the firm to screen the types of consumers. This is in contrast with the deterministic case in which a non-screening contract is more profitable for the firm. The reason is an uncertainty rationale. While a consumer always desires more discount in the per-usage price  $p$ , the discount

is offset by an increase in the entry fee  $L$ . Because a consumer does not know exactly her cost type, and she will lose too much from his sign-up fee if the draw of the cost type is too high, the uncertainty deters a consumer from behaving too aggressively. Such behavior narrows the consumer's choice set. Therefore it is more convenient for the monopoly firm to screen and profit from each type of consumer separately, because the cost to prevent choice deviation of consumers is cheaper, compared with the case when the cost type is known.

### 3.4.2 Random Cost: a numerical example

In this section, we provide a numerical example of the screening problem when the cost of consumption is unknown at  $t = 0$ .

We consider a specific density function from which the cost is drawn. Suppose the density function satisfies a typical exponential distribution:

$$f(c) = \begin{cases} \lambda e^{-\lambda c} & c \geq 0, \\ 0 & c < 0. \end{cases}$$

The parameter  $\lambda$  controls the tail of the density function  $f(c)$ . Figure 3.2 describes the shape of the exponential distribution when  $\lambda$  has different values. If  $\lambda$  is bigger, consumers have a better chance to be assigned a low cost. This implies that, as  $\lambda$  gets larger, the likelihood that consumers will choose to consume in the next period is higher.

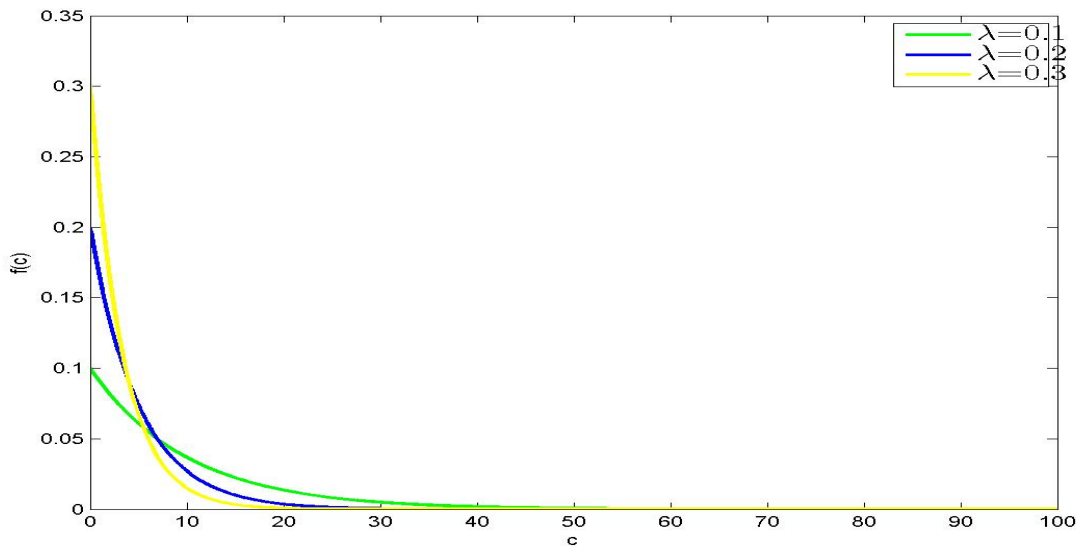


Figure 3.2: An Example of Decreasing Density Functions

In step 1, we check if the function  $\phi(p) = \int_{\hat{\beta}_1 \delta b - p}^{\hat{\beta}_2 \delta b - p} -p - c + \delta b dF(c)$  is increasing in  $p$ . Figure 3.3 demonstrates a numerical example of the function  $\phi(p)$ . As shown in Figure 3.3,  $\phi(p)$  increases in  $p$ . Moreover, the the slope of  $\phi(p)$  increases as  $\lambda$  increases. That is, as  $\lambda$  becomes larger, the likelihood that a consumer obtains a low cost is higher, and the ability of the monopoly firm to distinguish the types of consumers becomes larger (which can be interpreted as a larger slope of  $\phi(p)$ ).

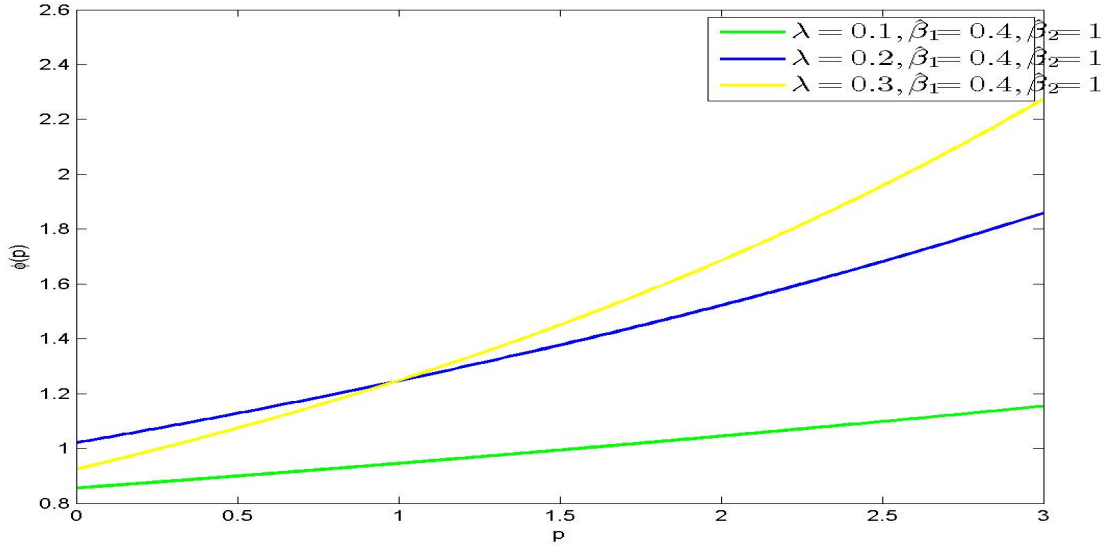


Figure 3.3:  $\phi(p)$ . The plot shows the function  $\phi(p) = \int_{\hat{\beta}_1 \delta b - p}^{\hat{\beta}_2 \delta b - p} -p - c + \delta b dF(c)$ . The distribution is an exponential distribution with parameter  $\lambda$ . The other parameters assigned are:  $\hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1$ .

In step 2, we study whether the Single Crossing Property (SCP) holds in this model. Suppose  $\lambda = 0.1$ , Figure 3.4 shows the indifference curve for both types of consumers in which the firm provides contracts  $(L_1, p_1)$  and  $(L_2, p_2)$ , with  $p_1 < p_2$ . The blue curve represents the indifference curve for type 1 (more sophisticated) agent, whereas the green one represents type 2 (more naive) agents. The Single Crossing Property (SCP) holds in this example. Since the more sophisticated type strictly prefers a low per-usage price  $p$ , the second contract  $(L_2, P_2)$  (with  $p_2 > p_1$ ) represents a strictly lower utility for the more sophisticated consumer.

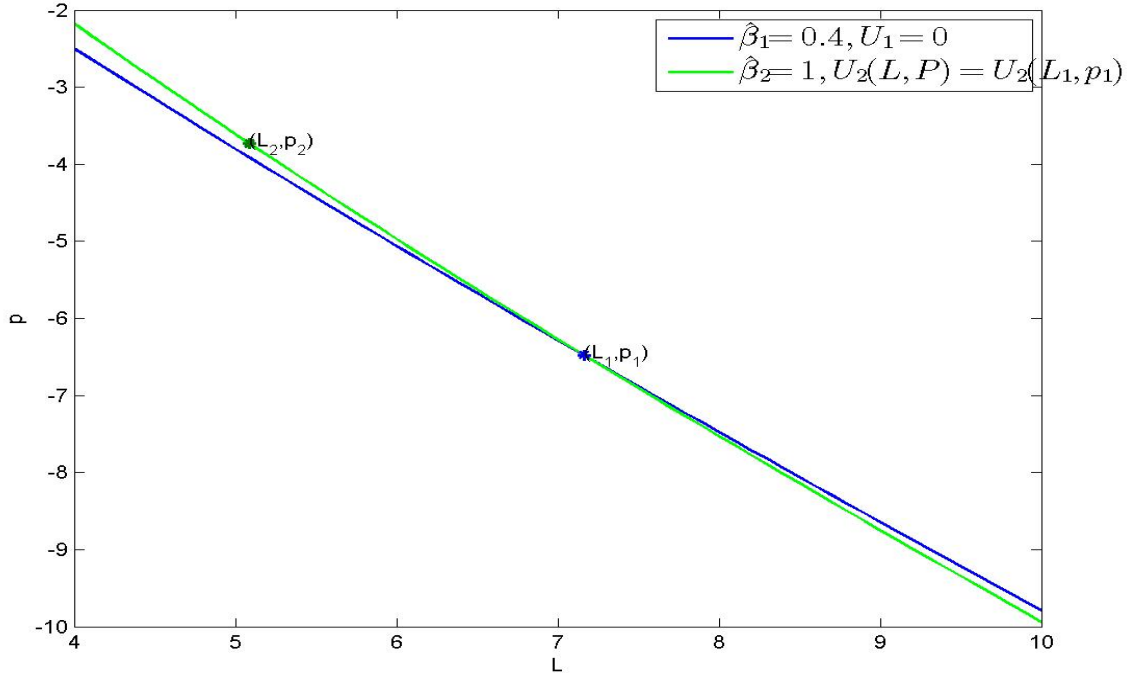


Figure 3.4: **Indifference Curves.** The plot shows the indifferent curves for both type 1 and type 2 consumers. The parameters are  $\lambda = 0.1, \alpha = .4, \beta = 0.3, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1$ .

In step 3, we examine the numerical solution of the firm’s profit-maximization problem. The results are provided in Table 3.1. In Table 3.1,  $u_i(L, P)$  represents the subjective utility for type  $i$  agents, and  $\tilde{u}_i(L, p)$  represent the real utility for the type  $i$  agents. That is,

$$\tilde{u}_i(L, P) = \beta\delta(-L + \int_{-\infty}^{\beta\delta b} (-p - c + \delta b)dF(c)).$$

As shown in Table 3.1, we verify that, for optimality, the firm provides a screening contract in which  $L_1 > L_2$ , and  $p_1 < p_2$ . Moreover, in this particular example,  $L_i > 0$  and  $p_i < 0$ . That is, the firm offers consumers a negative per-usage price  $p$ : it reimburses the consumers when they use the product. Such “seemingly attractive” contracts are observable in reality. For example, a sports club would charge a high entry fee and reimburse the consumers when they start to use the equipments. This trick lures consumers and make it nearly “impossible” for them to understand that, in fact, consumers lose money by accepting such an offer (a consumer’s subjective utility  $u_i$  is always larger than zero, while the real utility

$\tilde{u}_i$  is less than zero). Such a pricing scheme (screening + negative per-usage price) helps the firm to achieve more profit. The naiver consumers (type 2), with higher subjective short-run discount factor  $\hat{\beta}_s$ , are exploited more heavily in the sense that  $\tilde{u}_2 \leq \tilde{u}_1$ . Nevertheless, they subjectively interpret it in the opposite way ( $u_2 > u_1$ ).

We also observe in this example that the firm's profit is increasing in  $\lambda$ . As we argued, screening is costly for the firm. As  $\lambda$  increases, consumer's utility function becomes more sensitive to the variation of  $p$ , hence it is cheaper for the firm to distinguish the types of agents. Therefore, a larger  $\lambda$  yields a higher profit and a heavier exploiting of the more naive type (i.e.  $\tilde{u}_2$  decreases with  $\lambda$ ).

$\lambda$	$(L_1, p_1)$	$(L_2, p_2)$	$u_1$	$\tilde{u}_1$	$u_2$	$\tilde{u}_2$	Profit
$\lambda = 0.1$	(7.1616, -6.4775)	(5.0878, -3.7325)	0	-0.0543	0.1207	-0.1100	2.7718
$\lambda = 0.2$	(10.8572, -6.8976)	(6.8438, -2.6024)	0	-0.0382	0.0692	-0.1849	4.3369
$\lambda = 0.3$	(12.7805, -7.1827)	(7.5151, -1.8175)	0	-0.0193	0.0289	-0.2127	5.2585

Table 3.1: **Optimal contracts and profits as  $\lambda$  varies.** Parameters are  $\alpha = .4, \beta = 0.3, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1$ .

In step 4, we examine the effect on the firm's optimal contract and profit as the population varies. Table 3.2 characterizes such variations. As  $\alpha$  increases, the population of more sophisticated types increases. The firm ends up obtaining less profit and, in the meantime, exploiting more heavily the more sophisticated type ( $\tilde{u}_1$  increases) and less heavily the more naive type ( $\tilde{u}_2$  increases). The reason why the firm's profit declines with  $\alpha$  lies in the fact that the firm mainly extracts profit from the more naive type. As the population of more naive types decreases, it becomes harder for the firm to realize the profit.

$\alpha$	$(L_1, p_1)$	$(L_2, p_2)$	$u_1$	$\tilde{u}_1(L_1, p_1)$	$u_2$	$\tilde{u}_2(L_2, p_2)$	Profit
$\alpha = 0.4$	(7.1616, -6.4775)	(5.0878, -3.7325)	0	-0.0543	0.1207	-0.1100	2.7718
$\alpha = 0.5$	(6.6737, -5.8878)	(5.0246, -3.7325)	0	-0.0576	0.1281	-0.1038	2.7576
$\alpha = 0.6$	(6.4277, -5.5876)	(4.9599, -3.6125)	0	-0.0593	0.1320	-0.1016	2.7478

Table 3.2: **Optimal Contracts and Profits As  $\alpha$  Varies.** Parameters are  $\lambda = 0.1, \beta = 0.3, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1$ .

### 3.5 An Extension: The Multi-period Model

In many realistic cases, including sports club, education program, etc., firms offer series of contracts that cover multiple periods. For example, a typical gym club offers membership privileges for different periods, with different levels of membership fees. Usually, a longer membership costs a higher entry fee. Consumers have to decide which contract to enter: they face a trade-off between entry fee and the time to consume. To examine this case, we extend the 3-period model to a multi-period model.

In this model, a contract  $(L, T)$  is represented by an entry fee,  $L$ , and the time to consume,  $T, T \geq 1$ . Unlike the 3-period model, we do not consider the per-usage fee in this model. Suppose a consumer purchases a  $T$ -period contract and consumes in any period  $t \leq T$ , he then pays the a random cost  $c_t \sim F(c), i.i.d, t = 1, 2, \dots, T$ , and enjoys the benefit in the period  $t + 1$ .

The consumer types, preferences, behaviors, and firm behavior are the same as the 3-period model.

Hence at  $t = T - 1$ , a type  $(\beta, \hat{\beta}_i)$  consumer evaluates a contract  $(L, T)$  as follows,

$$U_{i,T-1} \left( (L, T), (\beta, \hat{\beta}_i) \right) = \beta \delta \int_{-\infty}^{\hat{\beta}_i \delta b} (-c + \delta b) dF(c).$$



At  $t = T - 2$ , she evaluates the contract as

$$\begin{aligned} U_{i,T-2} \left( (L, T), (\beta, \hat{\beta}_i) \right) &= \beta\delta \left( \int_{-\infty}^{\hat{\beta}_i\delta b} (-c + \delta b) dF(c) + u_{T-1} \right) \\ &= \beta\delta(1 + \beta\delta) \int_{-\infty}^{\hat{\beta}_i\delta b} (-c + \delta b) dF(c). \end{aligned}$$

By backward induction, at  $t = 0$ , the consumer evaluates the contract as

$$\begin{aligned} U_{i,0} \left( (L, T), (\beta, \hat{\beta}_i) \right) &= \beta\delta \left( -L + (1 + \beta\delta + \dots + (\beta\delta)^{T-1}) \int_{-\infty}^{\hat{\beta}_i\delta b} (-c + \delta b) dF(c) \right) \\ &= \beta\delta \left( -L + \frac{1 - (\beta\delta)^T}{1 - \beta\delta} \right) \int_{-\infty}^{\hat{\beta}_i\delta b} (-c + \delta b) dF(c). \end{aligned}$$

Similarly, we twist the notation and use  $u_{i,t}(L_j, T_j)$  to represent the type  $i$  consumer's evaluation of contract  $(L_j, T_j)$  at period  $t$ . To study the screening problem and consumer's choices, Lemma 16 establishes the SCP of consumer's preferences and Proposition 13 examines the screening selections.

**Lemma 16.** The preferences satisfies the SCP. That is, if  $T_1 < T_2$  and

$$u_{2,0}(L_2, T_2) \leq u_{2,0}(L_1, T_1),$$

then

$$u_{1,0}(L_2, T_2) < u_{1,0}(L_1, T_1).$$

**Proof.** Function  $\varphi(T) = \frac{1 - (\beta\delta)^T}{1 - \beta\delta} \int_{\hat{\beta}_1\delta b}^{\hat{\beta}_2\delta b} (-c + \delta b) dF(c)$  is increasing in  $T$ , because  $\beta \leq 1$  and  $\delta \leq 1$ .

Suppose  $T_1 < T_2$  and  $u_{2,0}(L_2, T_2) \leq u_{2,0}(L_1, T_1)$ , then

$$\begin{aligned}
u_{1,0}(L_1, T_1) &= \beta\delta(-L_1 + \frac{1 - (\beta\delta)^{T_1}}{1 - \beta\delta}) \int_{-\infty}^{\hat{\beta}_1\delta b} (-c + \delta b)dF(c) \\
&= \beta\delta(-L_1 + \frac{1 - (\beta\delta)^{T_1}}{1 - \beta\delta}) \int_{-\infty}^{\hat{\beta}_2\delta b} (-c + \delta b)dF(c) - \varphi(T_1) \\
&> \beta\delta(-L_2 + \frac{1 - (\beta\delta)^{T_2}}{1 - \beta\delta}) \int_{-\infty}^{\hat{\beta}_2\delta b} (-c + \delta b)dF(c) - \varphi(T_2) \\
&= \beta\delta(-L_2 + \frac{1 - (\beta\delta)^{T_2}}{1 - \beta\delta}) \int_{-\infty}^{\hat{\beta}_1\delta b} (-c + \delta b)dF(c) \\
&= u_{1,0}(L_2, T_2).
\end{aligned}$$

Hence the SCP holds. □

The SCP implies that the low type (more sophisticated type) strictly prefers a lower  $T$ . Given the SCP is satisfied and the fact that  $u_{2,0}(L, T) > u_{1,0}(L, T)$ , consumers' local downward IC constraint implies a local upward IC constraint. In the optimal contract, the downward IC and the low type IR constraint binds. The firm's optimization problem becomes:

$$\max_{(L_1, T_1), (L_2, T_2)} \alpha \left( L_1 - \frac{1 - \delta^{T_1}}{1 - \delta} F(\beta\delta b)a \right) + (1 - \alpha) \left( L_2 - \frac{1 - \delta^{T_2}}{1 - \delta} F(\beta\delta b)a \right)$$

subject to

$$- L_1 + \frac{1 - (\beta\delta)^{T_1}}{1 - \beta\delta} \int_{-\infty}^{\hat{\beta}_1\delta b} (-c + \delta b)dF(c) = 0, \tag{IR1'}$$

$$- L_2 + \frac{1 - (\beta\delta)^{T_2}}{1 - \beta\delta} \int_{-\infty}^{\hat{\beta}_2\delta b} (-c + \delta b)dF(c) = -L_1 + \frac{1 - (\beta\delta)^{T_1}}{1 - \beta\delta} \int_{-\infty}^{\hat{\beta}_2\delta b} (-c + \delta b)dF(c). \tag{IC2'}$$

Proposition 13 characterizes the optimal contract of the monopoly firm.

**Proposition 13. (Multi-period, random cost)** In the multi-period model with random cost, the firm's optimal contract must satisfy  $L_1 \leq L_2$ , and  $T_1 \leq T_2$ .

**Proof.** Suppose that  $T_1 > T_2$ , and the high type's (IC2') constraint holds, then by the SCP the low type's (IC) constraint must be violated because the low type has a stronger

preference over a low  $T$ . Therefore it must be true that  $T_1 \leq T_2$ . And by profit maximization (IR1') and (IC2') must bind,  $L_1 \leq L_2$ .  $\square$

Proposition 13 implies that the more sophisticated type selects a contract with less periods ( $T_1 \leq T_2$ ) and hence pays less ( $L_1 \leq L_2$ ). The intuition lies in the fact that since the more sophisticated type has a more accurate estimation of her overconfidence, she chooses the more realistic contract. The more naive type, however, overestimates his utility from consumption and behaves more aggressively. Nevertheless, the screening success relies on the assumption that consumers are completely unaware of the firm's pricing scheme.

### **The multi-period model: A numerical example.**

We study a numerical example of the multi-period model. Parameters are the same as in section 3.4. Figure 3.5 plots consumers' indifference curves. We can verify in the figure that the SCP holds. In Figure 3.5, the blue line represents the indifference curve for the low type (more sophisticated), and the green line the high type (more naive). As we argued, the more sophisticated type strictly prefers a contract with shorter periods ( $L_1, T_1$ ).

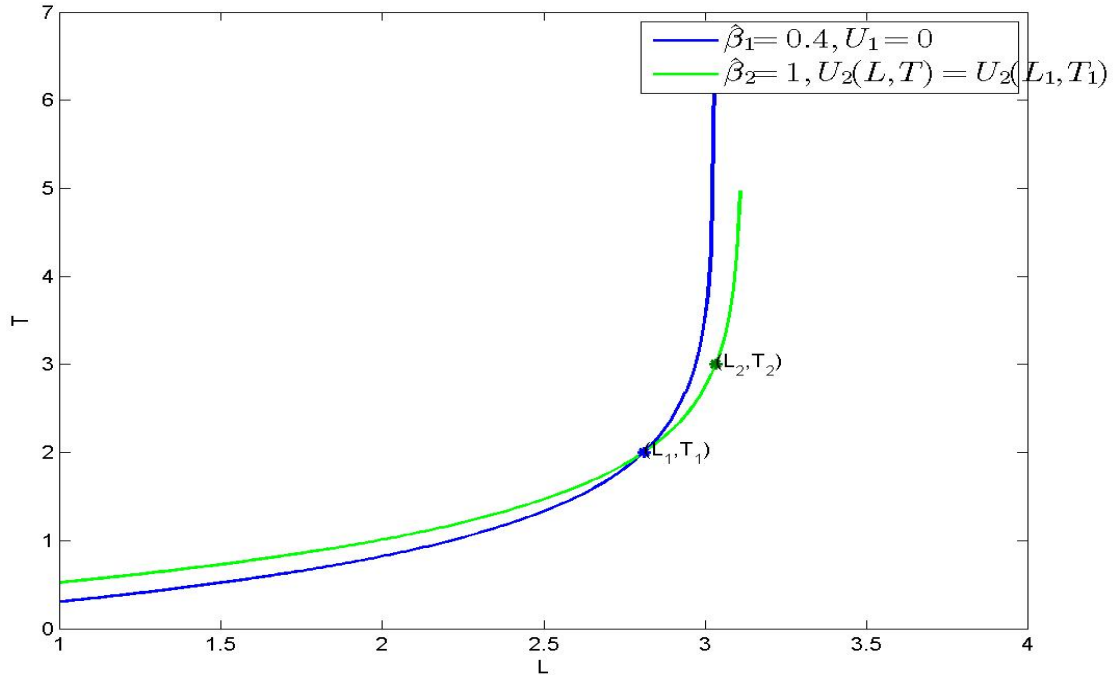


Figure 3.5: **Indifference Curves (Multi-period Model)**. Parameters are  $\lambda = 0.1, \beta = 0.3, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1$ .

Table 3.3 ( $\alpha = 0.4$ ) provides a numerical solution for optimal contracts. Consistent with the prediction in Proposition 13, the more sophisticated type chooses the contract with less periods and less entry fee ( $T_1 \leq T_2$ , and  $L_1 \leq L_2$ ). Both types of agents' perceived utilities ( $u_i$ 's) are positive. However, their actual utilities ( $\tilde{u}_i$ 's) are negative. The monopoly firm exploits the consumer's time-inconsistency (especially on the more naive type) and attains a positive profit. As  $\lambda$  increases, so consumers are more likely to have a low cost  $c$  and hence more likely to consume in the next period, the firm's ability to distinguish the two groups increases. This is reflected by the fact the contracts becomes less distinctive as  $\lambda$  increases. Consequently, the firm obtains more profit as  $\lambda$  grows.

$\lambda$	$(L_1, T_1)$	$(L_2, T_2)$	$u_1$	$\tilde{u}_1$	$u_2$	$\tilde{u}_2$	Profit
$\lambda = 0.1$	(2.8099, 2)	(3.0335, 3)	0	-0.1319	0.2933	-0.1563	2.3784
$\lambda = 0.2$	(4.8401, 2)	(5.1923, 3)	0	-0.1928	0.3497	-0.2240	4.0540
$\lambda = 0.3$	(6.3157, 2)	(6.3157, 2)	0	-0.2116	0.3170	-0.2116	5.2588

Table 3.3: **Optimal Contracts and Profits As  $\lambda$  Varies (Multi-period Model)**.  $u_i$  represents the perceived utility of a type  $i$  consumer and  $\tilde{u}_i$  the actual utility. Parameters are  $\beta = 0.3, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1$ .

Table 3.4 analyzes the variation of the optimal contract and the firm's profit as the population changes. As shown in Table 3.4, the optimal menu of contracts may not change, due to the constraint that  $T$  must be an integer, but the firm is strictly better off as  $\alpha$  decreases. This is in line with the rationale. Since the firm heavily extracts profit from the more naive type, an increase of the population of such type strictly increases the opportunity of exploiting.

$\alpha$	$(L_1, T_1)$	$(L_2, T_2)$	$u_1$	$\tilde{u}_1$	$u_2$	$\tilde{u}_2$	Profit
$\alpha = 0.4$	(2.8099, 2)	(3.0335, 3)	0	-0.1319	0.2933	-0.1563	2.3784
$\alpha = 0.5$	(2.8099, 2)	(3.0335, 3)	0	-0.1319	0.2933	-0.1563	2.3753
$\alpha = 0.6$	(2.8099, 2)	(3.0335, 3)	0	-0.1319	0.2933	-0.1563	2.3721

Table 3.4: **Optimal Contracts and Profits As  $\alpha$  Varies (Multi-period Model)**.  $u_i$  represents the perceived utility of a type  $i$  consumer and  $\tilde{u}_i$  the actual utility. Parameters are  $\lambda = 0.1, \delta = 0.9, \hat{\beta}_1 = 0.4, \hat{\beta}_2 = 1$ .

### 3.6 Conclusion

This paper studied the contract design and screening problem of a monopoly firm which interacts with consumers that are time-inconsistent and exhibits different degrees of so-

phistication. The contracts offered are two-part tariff contracts. The consumer's degree of naiveté is unobservable by the firm. We identify that the consumer's uncertainty on consumption costs is a variable that determines whether the firm would like to screen or not types of consumers. When consumers have deterministic consumption costs, a non-screening contract is optimal for the firm, whereas when consumers have random consumption costs, a screening contract is optimal for the firm. We characterize the optimal screening contract and show that the more sophisticated types accept the contract that has a further discount in the per-usage price. We then extend the model to a multi-period one and show that the more sophisticated types prefer a contract that has shorter periods and lower entry fee.

There are three major implications that can be applied to empirical studies. First, uncertainty on the consumers' side can limit their choice set, force them to perform less aggressively and hence decrease the screening costs for the monopoly firm. Therefore, an increase in the uncertainty increases the likelihood that a screening contract would be offered. Second, a discount per-usage price serves as both a commitment device and a tool to lure a naive consumer (of any degree). A more sophisticated consumer desires more the commitment device and therefore ask for a further discounted price. Third, the firm profits mainly by extracting surplus from the more naive type of consumers. The firm's profit decreases as the population of the naive types decreases.

## BIBLIOGRAPHY

- Admati, A. R. and Pfleiderer, P. (1988). Selling and Trading on Information in Financial Markets. The American Economic Review, 78(2):96–103.
- Alexandrov, A. (2016). Competing for Consumers with Self-Control Problems. Journal of Economics & Management Strategy, 25(1):179–194.
- ASIC (2013). Dark liquidity and high frequency trading. Information sheet, Australian Securities & Investments Commission.
- Babus, A. and Kondor, P. (2013). Trading and information diffusion in over-the-counter markets. Technical report, Working paper, Imperial College and Central European University.
- Baginski, S. P. and Hassell, J. M. (1997). Determinants of Management Forecast Precision. The Accounting Review, 72(2):303–312.
- Bhattacharya, U. and Daouk, H. (2002). The World Price of Insider Trading. The Journal of Finance, 57(1):75–108.
- Boni, L., Brown, D. C., and Leach, J. C. (2013). Dark Pool Exclusivity Matters. SSRN Scholarly Paper ID 2055808, Social Science Research Network, Rochester, NY.
- Botosan, C. A., Plumlee, M. A., and Xie, Y. (2004). The Role of Information Precision in Determining the Cost of Equity Capital. Review of Accounting Studies, 9(2-3):233–259.
- Budsaratragoon, P., Hillier, D., and Lhaopadchan, S. (2012). Applying developed-country regulation in emerging markets: an analysis of Thai insider trading. Accounting & Finance, 52(4):1013–1039.
- Buti, S., Rindi, B., and Werner, I. M. (2011a). Dark Pool Trading Strategies. Working Paper 421, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

- Buti, S., Rindi, B., and Werner, I. M. (2011b). Diving into Dark Pools. SSRN Scholarly Paper ID 1630499, Social Science Research Network, Rochester, NY.
- Chang, C., Chiang, Y.-M., Qian, Y., and Ritter, J. R. (2016). Pre-Market Trading and IPO Pricing. SSRN Scholarly Paper ID 2402806, Social Science Research Network, Rochester, NY.
- Chen, L. and Zhang, C. (2016). Short Selling Before Initial Public Offerings. SSRN Scholarly Paper ID 2623110, Social Science Research Network, Rochester, NY.
- Chen, Q., Francis, J., and Jiang, W. (2005). Investor learning about analyst predictive ability. Journal of Accounting and Economics, 39(1):3–24.
- Chlistalla, M. and Lutat, M. (2011). Competition in securities markets: the impact on liquidity. Financial Markets and Portfolio Management, 25(2):149–172.
- Chowdhry, B. and Nanda, V. (1991). Multimarket trading and market liquidity. Review of Financial Studies, 4(3):483–511.
- Chua, A. (2015). Compensation and Insider Selling During IPO. SSRN Scholarly Paper ID 2585197, Social Science Research Network, Rochester, NY.
- Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? Journal of Accounting and Economics, 27(3):285–303.
- Colla, P. and Mele, A. (2010). Information Linkages and Correlated Trading. Review of Financial Studies, 23(1):203–246.
- Comerton-Forde, C. and Putniņš, T. J. (2015). Dark trading and price discovery. Journal of Financial Economics, 118(1):70–92.
- Conrad, J., Johnson, K. M., and Wahal, S. (2003). Institutional trading and alternative trading systems. Journal of Financial Economics, 70(1):99–134.



- Cornelli, F., Goldreich, D., and Ljungqvist, A. (2006). Investor Sentiment and Pre-IPO Markets. The Journal of Finance, 61(3):1187–1216.
- Dambra, M., Field, L. C., Gustafson, M., and Pisciotta, K. J. (2015). Pre-IPO Communications and Analyst Research: Evidence Surrounding the JOBS Act. SSRN Scholarly Paper ID 2530109, Social Science Research Network, Rochester, NY.
- DeCovny, S. (2008). Dark Pools, Will anonymous trading venues have ripple effects? CFA Magazine, 19(4):28–32.
- Degryse, H., de Jong, F., and Kervel, V. v. (2015). The impact of dark trading and visible fragmentation on market quality. Review of Finance, 19(4):1587–1622.
- Degryse, H., Tombeur, G., Van Achter, M., and Wuyts, G. (2013). Dark Trading. In Baker, H. K. and Kiyamaz, H., editors, Market Microstructure in Emerging and Developed Markets, pages 213–230. John Wiley & Sons, Inc.
- Degryse, H., Van Achter, M., and Wuyts, G. (2009). Dynamic order submission strategies with competition between a dealer market and a crossing network. Journal of Financial Economics, 91(3):319–338.
- DellaVigna, S. and Malmendier, U. (2004). Contract Design and Self-Control: Theory and Evidence. The Quarterly Journal of Economics, 119(2):353–402.
- DellaVigna, S. and Malmendier, U. (2006). Paying Not to Go to the Gym. The American Economic Review, 96(3):694–719.
- Duffie, D., Malamud, S., and Manso, G. (2009). Information Percolation With Equilibrium Search Dynamics. Econometrica, 77(5):1513–1574.
- Duffie, D. and Manso, G. (2007). Information Percolation in Large Markets. American Economic Review, 97(2):203–209.
- Eliaz, K. and Spiegel, R. (2006). Contracting with Diversely Naive Agents. The Review of Economic Studies, 73(3):689–714.

- Esteban, S., Miyagawa, E., and Shum, M. (2007). Nonlinear pricing with self-control preferences. Journal of Economic Theory, 135(1):306–338.
- Fleming, M. and Nguyen, G. (2013). Order flow segmentation and the role of dark trading in the price discovery of U.S. treasury securities. Technical Report 624, Staff Report, Federal Reserve Bank of New York.
- Galperti, S. (2015). Commitment, Flexibility, and Optimal Screening of Time Inconsistency. Econometrica, 83(4):1425–1465.
- Glosten, L. R. and Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics, 14(1):71–100.
- Golosov, M., Lorenzoni, G., and Tsyvinski, A. (2014). Decentralized Trading With Private Information. Econometrica, 82(3):1055–1091.
- Gomber, P. and Pierron, A. (2010). MiFID - Spirit and Reality of a European Financial Markets Directive. SSRN Scholarly Paper ID 1858605, Social Science Research Network, Rochester, NY.
- Gottlieb, D. (2008). Competition over Time-Inconsistent Consumers. Journal of Public Economic Theory, 10(4):673–684.
- Gresse, C. (2006). The effect of crossing-network trading on dealer market’s bid-ask spreads. European Financial Management, 12(2):143–160.
- Gyntelberg, J., Loretan, M., Subhanij, T., and Chan, E. (2010). Private information, stock markets, and exchange rates. In Settlements, B. f. I., editor, The international financial crisis and policy challenges in Asia and the Pacific, volume 52, pages 186–210. Bank for International Settlements.
- Hasbrouck, J. (1995). One Security, Many Markets: Determining the Contributions to Price Discovery. The Journal of Finance, 50(4):1175–1199.

- Hatheway, F., Kwan, A., and Zheng, H. (2013). An Empirical Analysis of Market Segmentation on U.S. Equities Markets. SSRN Scholarly Paper ID 2275101, Social Science Research Network, Rochester, NY.
- He, W. P. and Lepone, A. (2014). Determinants of liquidity and execution probability in exchange operated dark pool: Evidence from the Australian securities exchange. Pacific-Basin Finance Journal, 30:1–16.
- Heidhues, P. and Koszegi, B. (2014). Using information about naivete to price discriminate.
- Heidhues, P. and Kszegi, B. (2010). Exploiting Naivete about Self-Control in the Credit Market. The American Economic Review, 100(5):2279–2303.
- Hendershott, T. and Jones, C. M. (2005). Island goes dark: Transparency, fragmentation, and regulation. Review of Financial Studies, 18(3):743–793.
- Hendershott, T. and Mendelson, H. (2000). Crossing Networks and Dealer Markets: Competition and Performance. The Journal of Finance, 55(5):2071–2115.
- Incekara-Hafalir, E. (2014). Credit Card Competition and Naive Hyperbolic Consumers. Journal of Financial Services Research, 47(2):153–175.
- Jiang, C. X., McInish, T. H., and Upson, J. (2012). Market Fragmentation and Information Quality: The Role of TRF Trades. SSRN Scholarly Paper ID 1960115, Social Science Research Network, Rochester, NY.
- Kang, S. K., Kang, H. C., Kim, J., and Kim, N. (2015). Insiders Pre-IPO Ownership, Underpricing, and Share-Selling Behavior: Evidence from Korean IPOs. Emerging Markets Finance and Trade, 51(sup3):66–84.
- Kim, O. and Verrecchia, R. E. (1994). Market liquidity and volume around earnings announcements. Journal of Accounting and Economics, 17(1):41–67.
- Kwan, A., Masulis, R., and McInish, T. H. (2015). Trading rules, competition for order flow and market fragmentation. Journal of Financial Economics, 115(2):330 – 348.

- Kyle, A. S. (1985). Continuous Auctions and Insider Trading. Econometrica, 53(6):1315–1335.
- Lang, M. and Lundholm, R. (1993). Cross-Sectional Determinants of Analyst Ratings of Corporate Disclosures. Journal of Accounting Research, 31(2):246–271.
- Lffler, G., Panther, P. F., and Theissen, E. (2005). Who knows what when? The information content of pre-IPO market prices. Journal of Financial Intermediation, 14(4):466–484.
- Li, D., Li, G. G., and Li, M. (2012). Information precision and its determinants. Insurance Markets and Companies: Analyses and Actuarial Computations, 3(1):6–14.
- Li, S., Yan, J., and Xiao, B. (2014). Contract Design and Self-Control with Asymmetric Information. Economic Inquiry, 52(2):618–624.
- Louis, H., Lys, T. Z., and Sun, A. X. (2014). Conservatism, Analyst Ability, and Forecast Error: Evidence on Financial Statement Users’ Ability to Account for Conservatism. SSRN Scholarly Paper ID 1031981, Social Science Research Network, Rochester, NY.
- Maffett, M. (2012). Financial reporting opacity and informed trading by international institutional investors. Journal of Accounting and Economics, 54(2):201–220.
- Manela, A. (2014). The value of diffusing information. Journal of Financial Economics, 111(1):181–199.
- Miao, J. (2006). A search model of centralized and decentralized trade. Review of Economic Dynamics, 9(1):68–92.
- Milgrom, P. and Stokey, N. (1982). Information, trade and common knowledge. Journal of Economic Theory, 26(1):17–27.
- Mittal, H. (2008). Are You Playing in a Toxic Dark Pool? The Journal of Trading, 3(3):20–33.

- Næs, R. and Ødegaard, B. A. (2006). Equity trading by institutional investors: To cross or not to cross? Journal of Financial Markets, 9(2):79 – 99.
- Nimalendran, M. and Ray, S. (2014). Informational linkages between dark and lit trading venues. Journal of Financial Markets, 17(C):230–261.
- O’Donoghue, T. and Rabin, M. (1999). Doing It Now or Later. The American Economic Review, 89(1):103–124.
- O’Hara, M. and Ye, M. (2011). Is market fragmentation harming market quality? Journal of Financial Economics, 100(3):459–474.
- Ozsoylev, H. N., Walden, J., Yavuz, M. D., and Bildik, R. (2014). Investor Networks in the Stock Market. Review of Financial Studies, 27(5):1323–1366.
- Peretti, P. d. and Tapiero, O. (2014). “Is there light in dark trading” ? A GARCH analysis of transactions in dark pools. Technical report.
- Preece, R. (2012). Dark Pools, Internalization, and Equity Market Quality. CFA Institute.
- Ray, S. (2010). A Match in the Dark: Understanding Crossing Network Liquidity. SSRN Scholarly Paper ID 1570634, Social Science Research Network, Rochester, NY.
- Ready, M. J. (2014). Determinants of Volume in Dark Pool Crossing Networks. SSRN Scholarly Paper ID 1361234, Social Science Research Network, Rochester, NY.
- Rust, J. and Hall, G. (2002). Middlemen versus Market Makers: A Theory of Competitive Exchange. Working Paper 8883, National Bureau of Economic Research.
- Sandroni, A. and Squintani, F. (2013). Overconfidence and asymmetric information: The case of insurance. Journal of Economic Behavior & Organization, 93:149–165.
- Sanjay, J. (2009). Self-Control and Optimal Goals: A Theoretical Analysis. Marketing Science, 28(6):1027–1045.

- Sarkar, A., Schwartz, R. A., and Klagge, N. (2009). Liquidity Begets Liquidity. Trading, 2009(1):15–20.
- Sayman, S. and Öncüler, A. (2008). An Investigation of Time Inconsistency. Management Science, 55(3):470–482.
- Shui, H. (2004). Time Inconsistency in the Credit Card Market.
- Tang, H.-w., Chen, A., and Chang, C.-C. (2013). Insider trading, accrual abuse, and corporate governance in emerging markets-Evidence from Taiwan. Pacific-Basin Finance Journal, 24:132–155.
- Viswanathan, S. and Wang, J. J. D. (2002). Market architecture: limit-order books versus dealership markets. Journal of Financial Markets, 5(2):127–167.
- Wang, X. and Wu, M. (2011). The quality of financial reporting in China: An examination from an accounting restatement perspective. China Journal of Accounting Research, 4(4):167–196.
- Weaver, D. G. (2014). The Trade-At Rule, Internalization, and Market Quality. SSRN Scholarly Paper ID 1846470, Social Science Research Network, Rochester, NY.
- Wolinsky, A. (1990). Information Revelation in a Market with Pairwise Meetings. Econometrica, 58(1):1–23.
- Wong, W.-K. (2008). How much time-inconsistency is there and does it matter? Evidence on self-awareness, size, and effects. Journal of Economic Behavior & Organization, 68(34):645–656.
- Ye, M. (2010). Non-Execution and Market Share of Crossing Networks. SSRN Scholarly Paper ID 1719016, Social Science Research Network, Rochester, NY.
- Ye, M. (2011). A Glimpse into the Dark: Price Formation, Transaction Cost and Market Share of the Crossing Network. SSRN Scholarly Paper ID 1521494, Social Science Research Network, Rochester, NY.

- Ylmaz, M. (2015). Contracting with a nave time-inconsistent agent: To exploit or not to exploit? Mathematical Social Sciences, 77:46–51.
- Yu, Z. and Lu, W. (2009). Literature Review of the Accounting Information Quality in China. In Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology, 2009. COINFO '09, pages 194–197.
- Zhang, W. (2012). Endogenous Preferences and Dynamic Contract Design. The B.E. Journal of Theoretical Economics, 12(1).
- Zhu, H. (2014). Do dark pools harm price discovery? Review of Financial Studies, 27(3):747–789.