Essays on Commodity Markets

by

Chao Yang

Business Administration
Duke University

Date: ______________________________

Approved:

______________________________
Simon Gervais, Supervisor

______________________________
S. “Vish” Viswanathan

______________________________
Ming Yang

______________________________
Qi Chen

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University

2019
ABSTRACT

Essays on Commodity Markets

by

Chao Yang

Business Administration
Duke University

Date: ____________________________
Approved: ______________________

______________________________
Simon Gervais, Supervisor

______________________________
S. “Vish” Viswanathan

______________________________
Ming Yang

______________________________
Qi Chen

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University

2019
Abstract

This essay is composed of two papers. In the first paper, I build a theoretical model for the double-sided squeeze in the commodity futures market. The model shows that the manipulator can profit from combining standard short squeeze techniques with control of the physical flows in the warehouses. In the second paper, I build a model for commodity middlemen with aggregate demand shocks. The model shows that having more middlemen in the market can increase the spot price volatility and decrease producers’ surplus, all contrary to common wisdom.
Acknowledgements

All my love to my parents and my grandparents, who gave me my life. I can never be too gratitude to them.

My special thanks are given to Simon Gervais. I got familiar with Simon as TA for his MBA Corporate Finance course. I remember clearly that in the end of the semester that year I failed to fulfill my tasks as the grader due to some personal reasons. Simon gave the warmest encouragement and told me that “family is always the most important.” I was moved by his kindness and understanding at the time. Besides, Simon gave me the utmost freedom of control over my PhD life and is always timely and nicely in pointing out my problems. Ming also spent much time talking with me on the model details and gave me many very good advices. I also give my sincere and deep thanks to Vish, and Qi, who formed my committee. I learned a lot from talking with each of them.

Besides, I am very appreciative to Hui Zheng and Lukas Schmid, without whom I couldn’t have entered Fuqua’s PhD program. Prof. Zheng is my undergraduate advisor at Fudan University. He has a lifelong impact on me and has given me many precious suggestions, one of which is to do a PhD. I had never thought about doing a PhD until I talked with Prof. Zheng. I first met Lukas when attending his Finance I course as a master’s student in economics. He was such a nice and patient person that we could just stop by his office and ask questions for half an hour without an appointment. Lukas is one of the nicest persons I have ever met. Besides, Lukas gave me humongous help in my PhD application.

I also sincerely thank Tobias Hurth, who introduced me to the branch of mathematics called iterated function systems. My work in Chapter 3 relies on a result proved in that literature. Although we have never met and have only communicated through emails, Tobias is very nice and patient in answering all my questions on
maths techniques. Yuan Xue, Yilin Jiang, and Yunke Mai all helped me on the latex formatting of this dissertation.

Finally, I am the luckiest to have Qinglong Zeng as my best friend during my PhD life.
Contents

Abstract iv

Acknowledgements v

List of Figures x

List of Tables 1

1 Introduction 2

2 Warehouse Manipulation in Commodity Futures Market 4

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

2.1.1 Related Literature . . . . . . . . . . . . . . . . . . . . . . . . . 5

2.2 Institutional Background and Empirical Findings . . . . . . . . . . . 7

2.2.1 Ownership transferred to FHC . . . . . . . . . . . . . . . . . . . 10

2.2.2 Concentrated Inventory and Canceled Warrants . . . . . . . . . . . 12

2.2.3 Storage Fees and Higher Spot Premium . . . . . . . . . . . . . . 16

2.2.4 Futures Market . . . . . . . . . . . . . . . . . . . . . . . . . . . 20

2.3 The Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

2.4 Finding the Equilibrium . . . . . . . . . . . . . . . . . . . . . . . . . 32

2.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

3 A Model of Commodity Middlemen with Aggregate Shocks 41

3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41

3.1.1 Related Literature . . . . . . . . . . . . . . . . . . . . . . . . . 43

3.2 Model Setup . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44

3.2.1 Agents . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44
List of Figures

2.1 Flow of Warehouse Manipulation ........................................... 10
2.2 LME Inventory in/out of Detroit and Vlissingen ...................... 13
2.3 Merry-go-round Trades in LME Detroit Warehouses ................... 15
2.4 Warehouse Inventory is Concentrated and Cancelled .................. 16
2.5 Spot Premium vs. All-In Price ............................................. 18
2.6 LME Spread vs. Inventory ................................................. 22
2.7 LME Spread vs. Spot Premium ............................................. 23
2.8 Temporal Pattern for LME Trading Volume ............................. 25
2.9 Parameter Ranges for Equilibrium ....................................... 38
3.1 Structure of Market ....................................................... 49
3.2 Transition Mappings ($p = 0.05$ on left and $0.45$ on right) ...... 67
3.3 Value Functions ($p = 0.05$ on left and $0.45$ on right) ............ 68
3.4 Temporal Trading Patterns ($p = 0.05$ on left and $0.45$ on right) . 70
3.5 Special Case when $l = 0$ .................................................. 73
3.6 Special Case when $p = 1$ .................................................. 75
3.7 Special Case when $p = 0.5$ ............................................... 77
3.8 Slope of Transition Mapping ($p$) ....................................... 80
3.9 Slope of Transition Mapping $(1 - p)$ .................................. 80
3.10 Length of the Support of $l_M$ .......................................... 82
3.11 Length of the Support of $l_M$ (in % of $l$) .................................. 82
3.12 Mean $l_M$ ......................................................................................... 84
3.13 $l_M/l$ Ratio ......................................................................................... 85
3.14 Volatility of $l_M$ ............................................................. 86
3.15 Mean $l_S$ ......................................................................................... 87
3.16 Price Volatility in the Submarket $M \rightarrow B$ ......................... 89
3.17 Total Producers Surplus .............................................................. 91
List of Tables

2.1 Ownership transfer of LME-approved warehouse companies . . . . . . 11
2.2 LME warehouse owners by 2011 . . . . . . . . . . . . . . . . . . . . 14
2.3 Warrant Cancellation Cases in Detroit . . . . . . . . . . . . . . . . . 15
2.4 Welfare Distribution among Agents . . . . . . . . . . . . . . . . . . 40
Chapter 1

Introduction

The commodity market represents a large portion of the world economy but remains relatively under-researched compared to equity and fixed income markets. On the one hand, commodity derivative instruments are mainly written under complex contracts traded in relatively illiquid exchanges. These specific derivative markets are usually dominated by large players who have an informational advantage relative to the outside world. On the other hand, the functioning of commodity spot markets is full of institutional details unfamiliar to researchers with little industry experience. There is also a lack of data about the spot transactions, which are generally kept confidential and sometimes considered commercial secrets by the agents in these markets. Moreover, the interactive dynamics of the commodity derivative markets and the spot markets are tied together through the risk management and speculative activities by the large players active in both markets. One crucial category of such large players is the set of so-called global commodity trading firms (CTF). My dissertation describes the behavior of CTFs and analyzes their effects.

In the energy and oil products markets, some big CTF names include Vitol, Gunvor, and Mercuria. In the agricultural markets, there are the famous big four, usually called “ABCD”: ADM, Bunge, Cargill, and Louis Dreyfus. Some even larger CTFs such as Glencore and Trafigura have a wide range of business coverage including energy, minerals and metals, and agriculture products, etc. The CTFs’ activities in the commodity markets are well summarized by Pirrong (2014) in a booklet published by Trafigura:
“commodity trading firms specialize in (1) the production and analysis of information buyers and sellers active in the market, supply and demand patterns, price structures (over space, time, and form), and transformation technologies, and (2) the utilization of this information to optimize transformations.”

The second chapter characterizes the manipulative behavior of CTFs (and some large financial corporations) in the commodity futures market. When getting control over the physical inflows and outflows of the warehouses that are affiliated with the exchanges (mainly the London Metal Exchange (LME)), the manipulator can profit from riding over the high contango structure while threatening the competitors away by either short-squeezing them or forcing delivery of the physical goods that cannot easily get out of the warehouses.

The third chapter builds a trading model of dynamic equilibrium in which producers and middlemen face Bernoulli aggregate demand shocks. This chapter analyzes the effects that the quantity of middlemen and the uncertainty of the aggregate shocks have on equilibrium dynamics. Contrary to intuition, my model shows that having more middlemen in the market can increase the spot price volatility and decrease producer welfare. This is due to two competing facts. The first fact is that the producers benefit from trading indirectly through middlemen when they cannot meet the buyers directly. The second fact is that the producers face competition from middlemen when they do meet the buyers. Producer welfare decreases when the second force dominates the first one.
Chapter 2

Warehouse Manipulation in Commodity Futures Market

2.1 Introduction

The commodity futures market has long been used by the commodity producers and downstream consumers to hedge the price risk of commodities. Thus, the well-functioning of the commodity futures market is vital to the risk management of these firms as well as the global real economy. However, the commodity futures market has experienced dramatic changes since the 2000s.

One change is the large inflow of money to the commodity futures indices, a process usually called “the financialization of commodity markets” (Cheng and Xiong (2014)). Although the changed composition of futures market participants has both risk sharing and informational effects, the underlying institutional function of the commodity markets is not changed. Yet another change to the commodity markets happens at a more institutional level. Since 2000, the large banks’ involvement in the physical commodity market, including acquiring ownership of power plants, warehouse companies, electricity networks, and holding a large inventory of physical commodities, changes the way the futures market functions.

For example, consider the global aluminum market from 2010 to 2014. In 2010, two LME-approved warehouse operators were acquired respectively by Goldman Sachs and Glencore. The subsequent four years saw unprecedented increases of the LME aluminum inventory in their warehouses in Detroit and Vlissingen. Meanwhile, war-
rant cancellation reached as high as 70% of the total inventory in these warehouses, creating a very long load-out queue. In the most extreme case, a warrant holder of Goldman’s warehouse had to wait over two years in the queue before he could load his commodities out. Since the futures market’s physical settlement works through delivering warehouse warrants, the futures cash price is pressured down from the spot metal prices to an extent determined by the length of the load-out queue. The resulting price distortion results in profits for Goldman through futures trading. On the one hand, if the manipulator holds short positions, he could cancel a large number of warrants, increase the load-out queue, and lower the cash price, thus profiting from a deeper realized futures term spread. On the other hand, if the manipulator holds long positions, he could force the shorts to deliver the warrants that they don’t have, creating a temporary demand in the spot market and profiting from the higher cash price. When the small players in the futures market are uncertain about the manipulator’s position, they face the possibility of either being squeezed or being delivered a blocked warrant. These possibilities should be priced into futures and thus will affect the hedging of metal producers and consumers. This chapter models a futures market equilibrium in which both kinds of manipulation happen with positive probability and asks how they impact the welfare redistribution among market participants.

2.1.1 Related Literature

This chapter contributes to the growing literature on the financialization of commodities markets in two respects. First, this chapter speaks to the impact of financial investors through the futures delivery process. Many existing models claim that financial investors affect the commodity price dynamics through different channels,
including the information aggregation role \cite{Sockin2015} and the risk sharing role \cite{Acharya2013} of the futures market. However, both models assume a smooth delivery process when the futures contract matures. This assumption is questionable given the facts in the next section. When the large Wall Street banks participate in the physical markets and warehouse markets, the smooth delivery process is disrupted, and impacting prices in the process.

Second, most of this strand of the literature targets financial investors who take one-sided speculative positions in a trending market \cite{Tang2012, Basak2016, Knittel2016, Hamilton2015}, e.g., commodity index investors. These investors profit from directional change in the commodity price levels. In contrast, this chapter targets another set of financial investors who profit from carrying futures term spreads forward while being immune to the price level risks.

This chapter contributes to the literature on futures market manipulation. This chapter borrows the framework from \cite{Kyle1984}, the first paper to seriously model the corner/squeeze process, followed by \cite{Pirrong1993, Cooper1998} among others. All these papers model the short squeeze process in itself while this chapter emphasizes that the short squeeze can be used in combination with other manipulative behavior such as blocking the warehouse queue. The manipulator in this model squeezes not only to catch the squeeze profits but also to enlarge the gain from the contango term structure he creates by blocking the warehouse physical outflow. To my knowledge, this channel has not been modeled yet.

This model hints at some interesting points for future works. First, the inventory decision is affected by the manipulator’s potential to squeeze. The higher the probability of squeeze, the higher the convenience yield facing the short holders and the higher is the inventory of the squeezer. This possibly predicts a positive correlation
between convenience yield and inventory as in Tang and Zhu (2016), complementary to the classical storage theory of Working (1960), Telser (1958). Second, from a policy/regulation perspective, the model sheds light on the tradeoff between the need to have large traders and the effectiveness of the price discovery process in the futures market.

2.2 Institutional Background and Empirical Findings

In an exchange of commodity trading, the possibility of physical delivery ensures that futures prices remain in line with the physical industries and that futures prices converge to spot prices as the futures contracts get closer to maturity. To ensure the conformity to the exchange’s strict rules on commodity grade, quality, shape, and location in case of a physical delivery, any deliverable commodities underlying a futures contract are required to be stored in some warehouse approved by the exchange. Therefore, the global reach of the approved storage facilities and the smoothness of the inventory inflow/outflow of these storage facilities can greatly impact the futures pricing on the exchange.

The LME has a well-functioning global network of warehouses which contributes in making it the most traded market for industrial metals. By the end of 2016, the LME’s warehouse system was composed of over 600 LME-approved storage facilities across 40 locations in 14 countries. These warehouses are mostly located at the demand or shipment centers. When the metal producers send their LME-approved

---

brands of metals to these warehouses, the warehouse company registers the metals with the LME and issues warrants to the producers. As a result, the inventory of the warehouse increases. Conversely, when a warrant holder wants to get his commodity out of the warehouses, he must first apply to the warehouse company for putting his warrant onto a load-out queue. While in the load-out queue, a warrant is transformed into a “canceled warrant” but the inventory of the warehouse is not yet affected. The inventory of the warehouse is decreased only when the commodity underlying a canceled warrant is indeed loaded out of the warehouse. Any canceled warrant can be re-warranted while in the load-out queue. The warrants issued by the LME-approved warehouses are the only deliverables in an LME futures’ physical settlement.

Essentially, a warrant and the commodities underlying the warrant are considered equivalent in terms of trading, but working with the warrants in physical delivery makes the futures traders more confident in the type, quality, shape, and location of the commodities underlying their futures contracts. In all, a world-wide network of warehouses and the above warrant/physical delivery rules lie at the heart of the credibility of the LME as an exchange. Note that the above rules work under the assumption that the inventory outflow of the warehouses is smooth. Even in times of commodity oversupply when the global inventory is high, given that the LME has so many approved warehouses over the world, the large inventory should be scattered around in small amounts at each warehouse and thus the inventory outflows are ensured.

However, there are two loopholes in the LME rules that render this assumption untrue under certain market conditions. First, the LME does not own or operate warehouses, nor does it own the material they contain. It simply authorizes warehouse companies and the warehouses they operate to store LME-registered brands of metal on behalf of warrant holders and issues LME warrants through its London agent for material
delivered into its approved warehouses\textsuperscript{2}. This makes it possible for large financial holding companies (FHC) or large mercantile traders (LMT) to own the warehouse companies. Owning the warehouses provides them with the knowledge of the newest physical market conditions. For instance, the size of the producers/consumers’ hedging demand is an informational advantage in futures trading. The distribution of warrant holders could be used to time a futures market squeeze.

Second, the LME requires all of the approved-warehouses (when in need) to load out at least a certain amount of inventory (load-out-rule, LOR). This rule turned out to be disastrous as it can be used by manipulative warehouse owners to block warehouses that have accumulated huge inventories. By doing this, the warehouse owners profit from higher storage rents paid by the blocked warrant holders. If these warehouse owners participate in the futures trading, they can also profit from distorting the term structure of the futures price through manipulating the expected queue length. The following are empirical findings from a five-year long LME warehouse manipulation from 2010 to 2014. In 2010, Goldman Sachs and Glencore (a large global commodities trader) respectively bought Metro International Trade Services and Paccorini, two global LME-approved warehouse companies operating mainly in North America and Europe. During the five years after that, the LME aluminum inventory sky-rocketed and concentrated in the warehouses owned by the two companies in Detroit and Vlissingen. Long LME warehouse load-out queues were carried out by an operation called “merry-go-round” trades (discussed in detail later), blocking the outflow of the aluminum inventory into the spot market. The long queues bring various distortions to both the spot and futures market, as well as profits from various sources to those who created the queues. The manipulation framework is shown in the flowchart of Figure 2.1 with the focus of this chapter highlighted in the red box.

\textsuperscript{2}https://www.lme.com/Trading/Warehousing/Approved-warehouses
2.2.1 Ownership transferred to FHC

The LME warehouse manipulation story is just one aspect of the large banks’ involvement in the commodity markets. This trend originates from the passage of the Financial Services Modernization Act of 1999 (commonly referred to as the Gramm-Leach-Bliley Act) and the Commodity Futures Modernization Act of 2000 (CFMA). The two laws combined authorized U.S. banks to engage as “financial holding companies” (FHC) in many financial activities that were denied to them under the Glass-Steagall Act. In particular, the laws gave FHCs the access to physical commodity
markets. The 2007 financial crisis further deepens the banks’ involvement in the commodity markets through the acquisition of failed or bankrupted banks by the ones that survived. In 2008, the takeover of Bear Sterns by JPMorgan and the takeover of Lehman Brothers by Goldman Sachs and Morgan Stanley integrated these large banks’ already wide-ranging commodity businesses.\(^3\)

Before 2009, most of the LME-approved warehouses were owned by the six professional warehouse companies, including CWT Commodities, Henry Bath, Metro, North European Marine Services (NEMS), Pacorini, and Steinweg. In 2010, four of the six warehouse companies were bought by either FHC or LMT. See Table 2.1 for a list of such ownership transfer cases.

There are three reasons explaining the timing of the concentrated ownership transfer. First, the three big investment banks got their FHC identities late in 2008 and it took some time for them to reach an agreement with the warehouse companies. Second, warehousing is an “anti-cyclical” business. As the big banks’ other business lines shrank due to the financial crisis, warehousing flourished when demand for met-

\(^3\)According to a 2012 nonpublic report by the Federal Reserve, the three investment banks’ involvement in physical commodity areas includes power plants, metal warehouses, coal mines, uranium trading business, crude oil storage tank fields, natural gas storage facilities, liquefied natural gas ships, electricity grid, etc.

<table>
<thead>
<tr>
<th>Warehouse Group</th>
<th>Ownership</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWT</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Erus</td>
<td>Barclays</td>
<td>Sep 2011</td>
</tr>
<tr>
<td>GKE</td>
<td>Louis Dreyfus</td>
<td>Jun 2012</td>
</tr>
<tr>
<td>Henry Bath</td>
<td>JPMorgan</td>
<td>Jul 2010</td>
</tr>
<tr>
<td>Metro</td>
<td>Goldman Sachs</td>
<td>Feb 2010</td>
</tr>
<tr>
<td>NEMS</td>
<td>Trafigura</td>
<td>Feb 2010</td>
</tr>
<tr>
<td>Pacorini</td>
<td>Glencore</td>
<td>Nov 2010</td>
</tr>
<tr>
<td>Scale</td>
<td>Macquarie</td>
<td>Jul 2012</td>
</tr>
<tr>
<td>Steinweg</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>WWS</td>
<td>Noble</td>
<td>Oct 2010</td>
</tr>
</tbody>
</table>
als was lacklustre and stockpiles mounted\(^4\). Third, lower storage costs help the big banks gain higher profits in futures carry trade. When a commodity is in oversupply, the futures price of this commodity tends to be in contango. A carry trade in this case consists of holding the physical commodity in warehouses\(^5\) while shorting the far-end futures. The profit margin of a carry trade is the futures’s contango net of the storage and financing cost of holding the commodities. Thus, owning an LME-approved warehouse helps the big banks internalize the storage cost on carry trades\(^6\).

### 2.2.2 Concentrated Inventory and Canceled Warrants

After Goldman Sachs and Glencore bought Metro and Pacorini, the global LME aluminum inventory was redistributed to their Detroit and Vlissingen warehouses. Figure 2.2 shows that from 2010 to 2014 the LME aluminum inventory in Detroit and Vlissingen increased by about 2.7 million metric tons, compared to a 2.2 million metric tons reduction of LME aluminum inventory in all the other locations. Table 2.2 shows that by the end of 2011, most of the LME-approved warehouses in Detroit and Vlissigen were owned by Goldman and Glencore. This trend reached its peak in 2013 when the two cities held nearly 70% of all the LME aluminum inventory out of the 33 warehouse locations around the world.

The astounding increase in the inventory is achieved through “freight incentive”

\(^4\)Financial Times, March 2, 2010, “Goldman and JPMorgan enter metal warehousing”, https://www.ft.com/content/5025f82a-262e-11df-aff3-00144feabdc0

\(^5\)From Senate’s report (Levin (2014)), p116, at the end of 2011, Goldman held 231,000 metric tons of aluminum, 37,000 metric tons of copper, 3,000 metric tons of nickel, etc. From the same report, p183, Goldman’s aluminum holding by the end of 2012 exceeded 1.5 million metric tons worth more than $3.2 billion dollars.

\(^6\)The rent fee of the off-LME warehouses is much cheaper than that of the LME-approved warehouses. However, the fact that the big banks still want to hold their commodities in the LME-approved warehouses may be due to the liquidity requirement of their internal risk management. These commodities are mostly collateralized through financing deals. In this case, the financier may also raise a liquidity requirement on the collateral.
Figure 2.2: LME Inventory in/out of Detroit and Vlissingen

given by the warehouse operators to the metal owners.\footnote{From the Senate’s report, p188, from 2010 to 2013, Goldman paid $37, $79, $103, $129 million in freight incentives to attract “free aluminum” to its warehouses.} After attracting the aluminum inventory into its warehouses, Goldman also paid millions of dollars to a few large metal owners whose aluminum was already stored inside Metro to participate in what is called “merry-go-round” deals (or “yo-yo” trades\footnote{Summary Public Report of the LME Warehouse Consultation, p38.}). The deals include Metro making incentive payments to the financial firms to cancel their metals held at Metro warehouses; join the queue to exit the warehouse; upon reaching the front of the queue, load out the aluminum, send them back to a nearby warehouse, and re-warrant them. A load-out queue can be created, because when loading out the canceled inventories, Metro actually carried out the minimum load-out amount in the LME load-out-rule as the maximum amount that can be loaded out. Further exacerbating the situation is the ruling that the queues are computed warehouse-

13
Table 2.2: LME warehouse owners by 2011

<table>
<thead>
<tr>
<th>City</th>
<th># of Warehouses</th>
<th>GS</th>
<th>JPM</th>
<th>TG</th>
<th>Glencore</th>
<th>control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johor</td>
<td>21</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>12</td>
<td>81%</td>
</tr>
<tr>
<td>New Orleans</td>
<td>56</td>
<td>19</td>
<td>3</td>
<td>1</td>
<td>32</td>
<td>98%</td>
</tr>
<tr>
<td>Antwerp</td>
<td>70</td>
<td>0</td>
<td>6</td>
<td>29</td>
<td>2</td>
<td>53%</td>
</tr>
<tr>
<td>Detroit</td>
<td>37</td>
<td>29</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>84%</td>
</tr>
<tr>
<td>Singapore</td>
<td>41</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>39%</td>
</tr>
<tr>
<td>Rotterdam</td>
<td>170</td>
<td>0</td>
<td>42</td>
<td>0</td>
<td>6</td>
<td>28%</td>
</tr>
<tr>
<td>Vlissingen</td>
<td>55</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>53</td>
<td>96%</td>
</tr>
</tbody>
</table>

company-wise instead of warehouse-wise. By doing this, Goldman can effectively create a load-out queue and block the other aluminum held at its warehouses from getting out. Figure 2.3 plots the dynamics of canceled aluminum held at Metro’s Detroit warehouses. The red lines show the identified large warrant cancellations, detailed in Table 2.3. The type “M” in the table means that the warrant cancellation is due to a merry-go-round deal agreed between Metro and the warrant holder. The type “P” means that the cancellation is due to proprietary cancellations without a deal. As can be seen, Goldman had a role to play in most of the increase in warrant cancellation up to early 2013 with the result of a greatly lengthened load-out queue in Detroit.

Concurrently, huge amounts of LME aluminum warrants were canceled and thus a prolonged load-out queue was created in Glencore’s Vlissingen warehouses. As shown is Figure 2.4 Detroit and Vlissingen held a large fraction of the LME aluminum inventory (as high as 70% in 2014) and the LME-approved warehouses in the two locations were mostly owned by Goldman and Glencore. Thus, it can be said that the efforts of Goldman and Glencore to create long load-out queues in their warehouses shaped the

9From the Senate’s report, p190, Goldman identified six such cases involving 600,000 metric tons of aluminum. Besides, Metro also saw four large proprietary aluminum cancellations involving about 500,000 metric tons of aluminum held by Goldman and JPMorgan.

10It is possible that the increase in canceled warrants in 2013 and 2014 also had some connection to Goldman.
Figure 2.3: Merry-go-round Trades in LME Detroit Warehouses

Table 2.3: Warrant Cancellation Cases in Detroit

<table>
<thead>
<tr>
<th>Time</th>
<th>Warrant Holder</th>
<th>Volume (mt)</th>
<th>Type</th>
<th>Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010 Sept</td>
<td>Deutsche Bank</td>
<td>100,000</td>
<td>M</td>
<td>65</td>
</tr>
<tr>
<td>2012 Jan</td>
<td>JPMorgan</td>
<td>100,000</td>
<td>P</td>
<td>100</td>
</tr>
<tr>
<td>2012 Jan-Mar</td>
<td>Red Kite</td>
<td>250,000</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>2012 May</td>
<td>Goldman Sachs</td>
<td>50,000</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>2012 Jul</td>
<td>Goldman Sachs</td>
<td>45,000</td>
<td>P</td>
<td>85</td>
</tr>
<tr>
<td>2012 Nov</td>
<td>Red Kite</td>
<td>160,000</td>
<td>M</td>
<td>150</td>
</tr>
<tr>
<td>2012 Dec</td>
<td>Goldman, JPMorgan</td>
<td>322,000</td>
<td>P</td>
<td>150</td>
</tr>
</tbody>
</table>

LME warehouse system as a whole. Figure 2.4 shows the relative scale of total LME aluminum inventory, inventory in Detroit and Vlissingen, and total LME inventory on canceled warrants. In sum, as discussed in this subsection, the load-out queue length can be manipulated by the warehouse owner who has accumulated enough inventory in its warehouses.\(^{11}\)

\(^{11}\)In the model below, a simplifying assumption is made that the amount of inventory at the manipulator’s disposal is large enough to create a load-out queue (if all canceled) and that this amount is observable to the market.
2.2.3 Storage Fees and Higher Spot Premium

There are two direct impacts of long load-out queues, corresponding to the two benefiting channels of the warehouse manipulator: increased storage fees and increased spot premium. To understand the two profit channels, some background on the global metal flows is useful. In the years after the financial crisis, the global physical metal market was experiencing a deep oversupply. The metal surplus then became inventory in either LME-approved warehouses or off-LME warehouses.\footnote{From 2010 to 2014, the global primary aluminum production surplus increased from 8 to 15 million metric tons, while the LME inventory stayed roughly at 4.5-5.5 million metric tons. Therefore, during this period, the percentage of LME primary aluminum inventory over the global inventory actually decreased from roughly 56\% to 37\%. The trade-off between the two types of storage facilities is essentially a trade-off between quality and cost.} There are two main
types of metal holders that put their metals in LME-approved facilities: large commodities traders and large banks. The large commodities traders, including Glencore and Trafigura, specialize in channeling metals from producers to end-consumers and providing warehouse, shipment, and custom services. The large banks are comparatively new to the metals markets and, as discussed above, participate mainly through financing deals. With large amounts of free metals at hand, both entities have basically two choices: either to sell the metal immediately though the LME futures market or to work out a financing deal and wait until the end the deal.

The first direct impact of prolonged load-out queue is increased storage fees. Three things are worth noting about this channel. First, the profit is coming from those who take the delivery through the futures market of a warrant issued by the warehouse. The target of the warehouse incentive is those metal holders who plan to immediately sell their metals through the LME market. By futures trading rules, the deliverable warrant in a physical settlement is at the seller’s choice. Therefore, the seller could load his metal into a high-fee warehouse that gives him some incentive in return. Second, the profit from increased storage fees could be used to fund the incentives to attract the metals, making the manipulation game sustainable. Since the warehouse market is roughly a competitive market and the warehouses compete through incentives, as long as the warehouse can bid higher incentives than other warehouses in attracting the free metals, it can keep this game running and concentrate more inventory. Third, a warrant in non-queued warehouses is rarely used in physical delivery (also called the “locked-up warehouses” problem). This is because the queued warrant is always cheaper than a non-queued one. This ensures that every LME buyer

---

13 This is essentially a principal-agent problem in which the seller decides the storage location of the metals on behalf of the buyers. From this logic, it is obvious that it makes no sense to give incentives to those metal owners who plan to take the stored metals out on their own in the future, such as a bank involved in financing deals.

who takes delivery will receive a warrant in the queued warehouses and have to pay the prolonged storage fee.\textsuperscript{19}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{plot.png}
\caption{Spot Premium vs. All-In Price}
\end{figure}

Another direct impact of prolonged load-out queue is an increased spot premium. The spot premium is defined as the difference between the all-in spot price and the cash price in the futures market. The all-in spot price is driven by the supply and demand in the physical metal market and the cash price is driven by large commodities traders’ selling/buying of their free metals into/from the LME futures market. Given an exogenous spot metal price, a selling merchant would take into account the incentive given by the queued warehouse in setting the ask price in the futures

\textsuperscript{19}Even when a seller finds it too costly to load his metals out of a non-queued warehouse into a queued warehouse, he can turn to the OTC warrant market, exchange for a queued warrant, and capture the differential premium.
market. Similarly, in setting the ask price in the futures market, a buying merchant would take into account the warehouse charges affiliated with the delivered warrant, the shipment cost to move the metal out of the warehouse, and possibly the liquidity cost and further rent costs when the warehouse is heavily queued. In formulaic form\footnote{Summary Public Report of the LME Warehouse Consultation, p39.} we have

\[
\text{spot premium} = \text{average spot price} - \text{LME cash price} = \text{average brand differential} + \text{average shipping cost} + \text{free on truck cost (FOT)} + \text{rent payable in queue} + \text{opportunity cost of queue wait.}
\]

The first three terms are the ordinary warrant discount and the last two terms are the queue-related discount. The black line in Figure\footnote{Euro high-grade premium, Japan 3-month premium, Singapore premium, and US midwest premium.} plots the global average aluminum spot premium. It is computed by taking the average of four representative regional spot premiums\footnote{Euro high-grade premium, Japan 3-month premium, Singapore premium, and US midwest premium.} in Europe, Japan, Singapore, and the U.S. The upward trend in 2010-2014 is largely due to the queue-related discount in the above formula \cite{Stevens2016}. The blue line plots the average spot premium as a percentage of all-in aluminum price. The spot premium remained less than 5% for most of the time up until the aforementioned manipulation story started in 2010. The percentage sky-rocketed to as high as 20% in late 2014. This has two direct implications. First, the LME futures market no longer provides as good a hedging as it used to as the spot premium cannot be hedged. Second, the term structure in the LME futures
market is affected by the markets’ expectation of the future spot premium or the future load-out queues, which is discussed in detail in the next subsection.

2.2.4 Futures Market

The warehouse manipulator not only benefits from higher storage rents, but also from futures market trading. Her main profit line comes through what is called a “financing carry trade.” In such a trade, one holds the physical metals financed by large banks and sells the futures in the far-end of the futures term. In a contango market, the futures price converges to the cash price as the futures contract approaches maturity. The term spread can then be captured through offsetting the short positions by taking a long position of the same amount at the front-end. As long as the term spread is greater than the financing cost and the storage cost of holding the physical metals, the carry trades generate profits. At the time of the previous example, the global borrowing cost was extremely low, particularly for firms like Goldman. Besides, since Goldman owns the warehouses, the storage cost could be internalized. Therefore, for Goldman or Glencore, the financing carry trade is always profitable.

However, there are competitors in carry trades. Competitors are divided into two types according to their risk appetite: risk-averse traders and risk-neutral traders. The risk-averse traders are usually large merchants or large banks like Glencore or Goldman. They usually carry out the trade as described above. They can be considered to be term arbitrageurs that determine the term spread of the futures market. Goldman is basically fine with these traders since they won’t bid the spread lower than that determined by LME storage rents, thus doing no harm to Goldman’s profit line. Due to the financiers’ liquidity requirement, the metals these traders hold as collaterals are usually stored in LME warehouses, many of which belong to Goldman, essentially channeling the futures term spread to Goldman in terms of storage
rent. Some of these traders can store their collateral metals in off-LME warehouses if their financing providers are satisfied with “standby” agreement, in which there exists a guarantee from a LME-approved warehouse that the metal can be warranted on demand.[18] Although their storage cost is lower than LME storage cost, the size of these cases is limited and the profit margin of Goldman is not affected much. The risk-neutral type traders[19] are usually hedge funds willing to take on greater risks. They usually do not hold physical metals to hedge the price risk. Instead, they long at the front-end of the term curve (to hedge the changing levels of the term curve). Since they do not face any storage cost, the pressure caused by these traders’ actions can potentially make the term curve flatter than the slope determined by the LME storage fees, and thus greatly reduce the profit margin of Goldman.

In order to keep these traders from pressuring the term curve, Goldman used a long-lasting futures market manipulation scheme called squeeze. In a short squeeze, the large long holds his position up to the final physical delivery stage, forcing the short position holders to deliver the physical metals that they don’t have. This creates a temporary demand for the metals. If the squeezer simultaneously controls the supply of the metals in the market, the squeezed shorts will have to either buy back their positions at a very high futures price or to acquire physical metals at a very high spot price[20] to fulfill their delivery obligations. In expectation of a potential squeeze, the shorts will cease to bid aggressively in order to avoid being squeezed themselves. This effect could have both quantity and price effects, i.e., smaller short positions

---


[19]The risk-neutral types are called this way not because they are indeed risk-neutral but to illustrate that they are relatively less risk-averse than the large banks.

[20]There are many potential reasons for the high price. It could be that the local supply of the good is limited and thus the shorts have to buy the metals in a far-away place and move them to the warehouses. In this case, the transportation cost contributes to the high price (Pirrong (1993)). Or it could be that the worst-deliverable metal is limited locally and the shorts have to deliver higher grades metals (Kyle (1984)). My model follows Kyle’s setting.
from these speculative traders and higher futures price incorporating the potential of a squeeze.

In our story, we do see indications of possible squeezes. Figure 2.6 plots the 3-

![Figure 2.6: LME Spread vs. Inventory](image)

month term spread (cash price minus 3-month futures price) in blue versus the total LME aluminum inventory in black. The shaded area corresponds to the period when Goldman owns and controls Metro. The term spread can be roughly characterized as increasing contango with frequent spikes. According to the storage theory of futures price, the increasing contango reflects the increasing storage and financing costs. Indeed, if the costs from the term spread are subtracted, values close to zero are attained in periods when there are no spikes. However, the spikes are not consistent
with competitive storage theories. According to the theory, the spikes show up only when there is tightness in the spot market, which is not the case in this story since the inventory is huge. Even if the bulk of LME inventory may be blocked by the load-out queue, there is still an even larger off-LME inventory. If there is a squeeze at delivery, the futures cash price will soar due to the artificial tightness and we will see spikes in the term spreads. Therefore, a reasonable explanation is that the spikes represent squeezes.

Besides the squeeze, a warehouse manipulator could time the warrant cancellation to further strengthen his profits from the futures carry trade. As shown above, an increase in load-out queue leads to a higher spot premium and, given an exogenous

![Figure 2.7: LME Spread vs. Spot Premium](image)

Figure 2.7: LME Spread vs. Spot Premium
all-in spot price, generates a negative shock to the futures cash price. If the manipulator takes a short position and then cancels large amounts of the warrants close to maturity, the resulting negative shock to the cash price makes it even cheaper for the manipulator to offset his short position, delivering a higher carry trade profit. Note that the risk-neutral traders’ profits are not affected by this strategy since they usually take long positions at the front-end while shorting at the far-end to hedge the level-risk. The gain in their short positions should be perfectly offset by the loss on their long positions. Only those who hold the physical inventory as the natural hedge to the short positions are able to catch this profit. More specifically, only when the load-out queue as well as the spot premium fall back to their original levels can the metals holders fully capture the profit. The large banks and merchants involved in financing deals fall into this category of traders, if not exclusively.

If a warehouse manipulator is both squeezing and timing to cancel the warrants, then he won’t do both at the same time. Note that to squeeze one needs to be net long and that to capture the negative shock to the cash price one needs to be net short. The direct testable implication is that the spikes in term spread (when the manipulator squeeze the shorts) happen interchangeably with increase in the spot premium (when the manipulator chooses to cancel his warrants). Figure 2.7 plots the term spread in blue and the spot premium in black. The red area highlights the two surges in the average spot premium and they correspond to the only two periods when the futures market were in relatively stable contango.

A complimentary implication of a squeeze is the shrinkage of pre-settlement trading volume. Since a squeezer has to hold his long position until very close to maturity to squeeze, the trading volume prior to the maturity should be relatively lower than when there is no squeeze. However, the counterfactual comparison is not testable. What can be done is comparing the time-series patterns of trading volumes over the
Figure 2.8: Temporal Pattern for LME Trading Volume

life-cycle of non-squeezed contracts to contracts potentially subject to squeeze. Figure 2.8 plots the average trading volume patterns of 3rd-Wednesday contracts. Each line in this figure shows the daily trading volume over the last eighty days up until maturity of the corresponding contracts. The rightmost point on the x-axis represents the day the contract matures and the leftmost point on the x-axis represents eighty days before the maturity date. For each contract, the time series of trading volume is scaled by its last value, so all the lines reach the value one in the last day. The 3rd-Wednesday contracts mature on the 3rd Wednesday of every month; thus, it is a monthly contract. 3rd-Wednesday contracts are the most liquid and most traded contracts and most susceptible to squeezes. The four red thin lines plot the monthly
average trading volume patterns from 2006 to 2009 (the pre-Goldman period). The thick red line is the average of the four thin red lines. The blue lines plot the trading patterns from 2010 to 2016. Both lines are scaled by the last trading volume to make them comparable. There was a huge spike in trading volume around one week before the settlement in the pre-Goldman period while the spike disappeared after 2010. One of many possible explanations is that the squeeze holds the trading volume up until maturity.

In the following section, I propose a model consistent with three empirical findings:

1. Frequent term spike with high inventory;
2. Spike and spot premium happen interchangeably;
3. Lower pre-settlement trading volume.

The main contribution of this model is to provide a welfare analysis among the commodity futures market participants under a warehouse manipulation game as described in this section.

2.3 The Model

Following Kyle (1984), a two-date one-period model is followed. The futures trading take place at the two dates, denoted t1 and t2. Before t1 there is no open interest and all the remaining open interests after the trading in t2 are liquidated through delivery. The deliverable commodity underlying the futures contract can be of two grades. The value of the low-grade commodity is denoted by a random variable $v$. The distribution of $v$ is public information to all traders in t1 and the realization of $v$ is public information to all traders in t2. The value of the high-grade commodity is always $v + d$ where $d$ is a constant known by all traders. The exogenous supply of
low-grade commodity z is public information at t1.

There are three types of traders: hedgers, speculators, and one manipulator. The hedgers trade exogenously. Thus, the model is a game played by the manipulator and the speculators.

The hedgers trade a number \( H \in \{H_0, H_1\} \) of net long contracts in t1. The distribution of \( H \) is independent of all other random variables in the model and is given by

\[
\mathbb{P}(H = H_0) = \lambda,
\]
\[
\mathbb{P}(H = H_1) = 1 - \lambda.
\]

It is assumed that \( 0 < H_0 < H_1 \) and denote \( \Delta H = H_1 - H_0 \). \( H = H_0 \) corresponds to a situation with inactive hedging demand. The distribution of \( H \) is public information to all traders, but the realization of \( H \) in t1 is observed only by the manipulator and not by the speculators. In t2, the hedgers put \( H \) number of short positions to offset their long positions.

The manipulator has \( I_q \) number of commodities stored in the registered warehouse at t1. \( I_q \) is public information to all traders before t1. In t1, after observing the hedgers net long position \( H \), the manipulator takes a position \( X_1(H) \). A positive \( X_1 \) means a long position and a negative \( X_1 \) means a short position. The speculators cannot observe \( X_1(H) \) but they can observe the net open interests of the hedgers and the manipulator combined: \( X_1(H) + H \). After the speculators set the futures price \( P_1(X_1(H) + H) \), the first trading date t1 is over. The pricing setting behavior by the speculators is described by the price function \( P_1(\cdot) \) and not modeled.

In the beginning of t2, the spot market value of the commodity \( v \) is realized and observed by all the traders. The manipulator then takes the futures position \( X_2(H, X_1(H), v) \)
and decides the quantity of physical commodities $I_c(H, X_1(H), v) \in [0, I_q]$ to load out of the warehouse. The fact that both $X_2$ and $I_c$ are functions of $H, X_1(H)$, and $v$ means that the manipulator can adjust the t2 trading quantity and load-out quantity in light of the t1 trading quantity and the realization of spot price $v$ he observes at the beginning at t2. Moreover, $X_1$ and $X_2$ admit randomization over all trading quantities, i.e., they are probability distributions over $\mathbb{R}$.

The waiting-to-be-loaded-out commodities then form a load-out queue at the warehouse, blocking the other stored commodities from getting out of the warehouse immediately. The delayed load-out becomes a liquidity burden on those who are in immediate need of selling the warehouse-stored commodities. This liquidity burden then creates a discrepancy between the spot price and the price of in-warehouse commodities, which is called the spot premium. The spot premium is modeled in reduced form as a function of the length of the queue $g(I_c)$. It is assumed that $g(0) = 0$ and that $g$ is monotonically increasing, thereby capturing the idea that the longer the queue the higher the spot premium the blocked traders are willing to pay. After observing the length of the queue $I_c$, the realization of the spot price $v$, and the net open interest $X_2(H) - H$, the speculators trade as the counter-party to the hedgers’ and the manipulator’s net positions and set the t2 futures prices $P_2$.

The speculators’ price setting behaviors are described by two measurable price functions: $P_1(X_1(H) + H)$ and $P_2(X_1 + H, X_2 - H, v)$. After observing the net open interests by the hedgers and the manipulator, the speculators trade as the counter-parties and competitively set the futures price such that in expectation they earn zero profit. Though the manipulator randomizes his strategy, we assume that the price function is deterministic. These deterministic price functions can be considered as the result of pure market competition.

Now let $\pi(X_1, X_2, I_c, P_1, P_2)$ be the manipulator’s profit function given his strategy.
$X_1$, $X_2$, $I_c$ and the speculators’ price function $P_1, P_2$. Then

$$
\pi(X_1, X_2, I_c, P_1, P_2) = X_1 \cdot [v - P_1] + X_2 [v - P_2]
$$

$$
+ \max \{X_1 + X_2 - z, 0\} \cdot d
$$

$$
- \min \{X_1 + X_2, 0\} \cdot g(I_c),
$$

or, in explicit form,

$$
\pi(X_1, X_2, I_c, P_1, P_2) = X_1(H) \cdot [v - P_1(X_1(H) + H)]
$$

$$
+ X_2(H, X_1(H), v) \cdot [v - P_2(X_1(H) + H, X_2(H, X_1(H), v) - H, I_c(H, X_1(H), v), v)]
$$

$$
+ \max \{X_1(H) + X_2(H, X_1(H), v) - z, 0\} \cdot d
$$

$$
- \min \{X_1(H) + X_2(H, X_1(H), v), 0\} \cdot g(I_c(H, X_1(H), v)).
$$

The first and second terms above are the no-squeeze no-block payoffs from the futures trading in $t_1$ and $t_2$. The third term gives the additional profit when there is a squeeze. If the manipulator’s outstanding long position $X_1 + X_2$ is in excess of the total deliverable supply of the low-grade commodity in the spot market, then he can hold these positions up until the end of $t_2$ when the settlement begins. In this case, the speculators will have to deliver a quantity $X_1 + X_2 - z$ of high grade commodities to fulfill their delivery obligations. The fourth term captures the profit when there is a block. When in $t_2$, the speculators realize that the manipulator has used his inventory to block the outflow of the warehouse, and the value of a maturing futures contract or one unit of the commodity in the warehouse decreases by the amount of the spot premium. Had the manipulator shorted the futures in $t_1$ at a relatively higher price, he would make a profit by longing the same amount of the futures at $t_2$ at a lower price. By blocking the outflow of the warehouse, the manipulator is
essentially trying to deepen the contango in the futures market.

Let \( y_1 = X_1(H) + H \) and \( y_2 = X_2(H) - H \) denote respectively the net open interests the speculators observe after the trading of the hedgers and the manipulator in \( t_1 \) and \( t_2 \). Then \( (y_1) \) and \( (y_1, y_2, I_c, v) \) are the information sets upon which the speculators make their decisions in \( t_1 \) and \( t_2 \). Note that it is written \( (y_1) \) instead of \( (y_1, I_q) \) in \( t_1 \) since \( I_q \) is public information and thus can be considered a parameter of the model.

The definition of the equilibrium of the model is as follows.

**Definition 1.** An equilibrium is a set of manipulator’s strategies \( X_1, X_2, I_c \) and speculators’ price functions \( P_1, P_2 \) satisfying the following conditions:

1. Optimality: Given \( P_1, P_2 \), for all \( X'_1, X'_2, I'_c \), there must be

\[
\mathbb{E}[\pi(X_1, X_2, I_c, P_1, P_2)] \geq \mathbb{E}[\pi(X'_1, X'_2, I'_c, P_1, P_2)].
\]

2. Market efficiency: For any \( y \) such that \( P(X_1(H) + H = y) > 0 \), there must be

\[
P_1(y_1) = \mathbb{E}[P_2(y_1, y_2, I_c, v)|y_1].
\]

3. \( P_2 \) must satisfy:

\[
P_2(y_1, y_2, I_c, v) = \begin{cases} 
  v + d & \text{if } y_1 + y_2 > z \\
  v & \text{if } 0 \leq y_1 + y_2 \leq z \\
  v - g(I_c) & \text{if } y_1 + y_2 < 0.
\end{cases}
\]

A few comments about this definition are warranted. First, note that the price function does not depend on the equilibrium choice of the manipulator, i.e., the idea that the manipulator distorts the speculators’ belief on the state of the economy
through strategic trading is not modeled here. This is more of a Nash equilibrium flavor. There are several merits of defining the equilibrium this way. The dynamic consistency is automatically ensured: when forming the expectation on the t2 trading outcomes, the speculators can use a “backward deduction” logic and take the optimal $X_2$ strategy as given. In other words, the credibility of the manipulator is not an issue in the model. Besides, the Nash concept makes the manipulator randomize over only those choices that generate the optimal profit, with all the other choices being assigned probability zero in equilibrium. This property simplifies the derivation of the equilibrium.

Second, the market efficiency condition requires that the price be given by the conditional expectations only for those net open interests observable on the equilibrium path. This requirement does not fully specify the price function (for those net open interests on off-equilibrium paths). However, it is important because it needs to be set in a way that induces the manipulator to stay on equilibrium paths. We will see later that there may exist many ways to set these off-equilibrium path prices. In the discussion of the uniqueness of equilibrium below, this chapter essentially means that the price functions are unique for the values given by the market efficiency constraint.

Third, $P_2$ implicitly assumes that the speculators as a whole have no bargaining power in buying high grade commodities from the spot market when there is a delivery need in case of a squeeze or have no bargaining power in offsetting their short positions from the futures market in case of a warehouse block. To see this, consider a case when the speculators need to buy $z + \epsilon$ amount of the commodities to fulfill their delivery obligations. The sellers in the spot market are assumed to be able to hold off selling until the spot prices rises to $v + d$. Similarly, consider a case when the speculators need to buy back long positions to offset their short positions traded in t1. The existing shorts in the futures market (the manipulator) is able to hold selling
the contracts and threaten to deliver the blocked in-warehouse physical commodities until the futures price decrease to $v - g(I_c)$. But proceeding to solving the equilibrium, here is the list of notations used in the model:

- $H$: the exogenous hedging demand
- $X_i(\cdot)$: the manipulator’s position in period $i$
- $P_i(\cdot)$: the price function set by speculators in period $i$
- $v$: the random price which will be realized in period 2
- $z$: the amount of low-grade product available in the market
- $d$: the price difference between high-grade and low-grade products
- $\lambda$: the probability of $H$ being low
- $I_c$: the amount of inventory controlled by the manipulator
- $\pi(\cdot)$: the manipulator’s profit function
- $g(\cdot)$: the reduced form function mapping queue length to spot premium
- $y_i$: the amount of net open interest observed by the speculators in period $i$
- $m, a, \mu$: other critical quantities defined later

## 2.4 Finding the Equilibrium

The following proof proceeds in four steps. In the first step (lemma 1), the dimension of the problem is reduced by specifying the optimal strategies in the second period, and thus the problem reduces to finding only the optimal $X_1(\cdot)$ and $P_1(\cdot)$. The second
step (lemma 2 to lemma 4) shows that the support of $X_1(\cdot)$ should live on isolated points. This informs that strategies $X_1(\cdot)$ that have a continuous distribution on some interval can be ruled out. The third step (lemma 5 and lemma 6) finds the possible values of those isolated points in step 2. Combining all the findings above, the final step (lemma 7 and lemma 8) gives a full description of two separate cases of the equilibrium depending whether the manipulator’s profit is positive. The proof of all the following lemmas can be found in the appendix.

The proof proceeds by backward induction, meaning, it starts by finding the equilibrium quantities in the second period. The following lemma characterizes the optimal strategies in $t_2$ for all possible sets of strategies in $t_1$. This allows focus on only the $t_1$ quantities $X_1$ and $P_1$.

**Lemma 1.** For all $X_1$, $X'_2$, $I'_c$, $P_1$, $P_2$, the trading strategies $X^*_2(H, X_1, v) = 0$ and $I^*_c(H, X_1, v) = \max(X_1, 0) \cdot I_q$ have the property that

$$
\pi(X_1, X^*_2, I^*_c, P_1, P_2) \geq \pi(X_1, X'_2, I'_c, P_1, P_2).
$$

Plug $X^*_2$, $I^*_c$, $P_2$ into $\pi(X_1, X^*_2, I^*_c, P_1, P_2)$, $\pi$ becomes a function of $X_1$ and $P_1$:

$$
\pi(X_1, P_1) = X_1(H) \cdot [v - P_1(X_1(H) + H)] + \max(X_1(H) - z, 0) \cdot d - \min(X_1(H), 0) \cdot g(I_q).
$$

The equilibrium then becomes a set of $X_1$ and $P_1$ such that

- $\mathbb{E}[\pi(X_1, P_1)] \geq \mathbb{E}[\pi(X'_1, P_1)]$ for all $X'_1$, and
- $P_1(y) = \mathbb{E}[P_2|y]$.

Note that the conditional expectation constraint is active only on $\{y|y = X_1(H) + H \text{ and } \mathbb{P}(X_1(H) + H) > 0\}$, i.e., the equilibrium futures price is determined uniquely.
up to sets of probability zero. Given the futures price function $P_1(\cdot)$, let $\Pi_P(x, h)$ measure the expected (with respect to $v$) manipulator profit when the manipulator trades $x$ and the hedgers trades $h$:

$$
\Pi_P(x, h) = x \cdot [\mathbb{E}[v] - P_1(x + h)] + \max(x - z, 0) \cdot d - \min(x, 0) \cdot g(I_q).
$$

Let $\Pi_P^*(h)$ measures the maximized manipulator profit:

$$
\Pi_P^*(h) = \max_x \Pi_P(x, h).
$$

The following lemma shows that the above maximized profit is obtainable and thus well-defined in equilibrium, for otherwise one can only write supremum instead of maximum.

**Lemma 2.** In equilibrium $\mathbb{P}(\Pi_P(X_1(H), H) = \sup_x \Pi_P(x, H)) = 1$.

The above lemma is helpful for characterizing the manipulator’s optimal strategy $X_1$. A critical question is upon observing $H_1$ in $t1$, whether the manipulator will take $X_1$ as a degenerate distribution over a finite number of position choices or randomize over a distribution which is continuous on some interval. The following two lemmas answer this question and rule out any continuous distribution from the manipulator’s choice set.

**Lemma 3.** In equilibrium, $\Pi_P^*(H_i) \geq 0$ for $i = 0, 1$, and at least one of them is strictly greater than zero.

**Lemma 4.** In equilibrium there exists some $m$ such that $\mathbb{P}(X_1(H_i) + H_i = m) > 0$ for $i = 0, 1$. Moreover, if $\Pi_P^*(H_i) > 0$, then $X_1(H_i)$ is supported on a set of isolated points.
For now, the proof has roughly described the support of $X_1$. The next question is which exact values should it take. The fact that the support of $X_1$ is isolated means that there must be some value $m$ for which $X_1$ takes with strictly positive probability. The next three lemmas illustrate the properties of such an $m$. The first of these splits the real line into four regions and characterizes the circumstances under which an equilibrium $m$ to falls into each region.

**Lemma 5.** In equilibrium, $X_1(H_0)$ has the following properties.

a. If $\mathbb{P}(X_1(H_0) = m) > 0$ for some $m \in (-\infty, 0)$, then $\Pi^*_p(H_0) = 0$ and $P_1(m + H_0) = \mathbb{E}[v] - g(I_q)$.

b. If $\mathbb{P}(X_1(H_0) = m) > 0$ for some $m \in (z, \Delta H) \setminus \{\sqrt{zd\Delta H/d+g}\}$, then $\Pi^*_p(H_0) > 0$.

c. $\mathbb{P}(X_1(H_0) = m) = 0$ for all $m \in [\Delta H, +\infty)$.

From now on let $m^* = \sqrt{zd\Delta H/d+g}$. From part b of the last lemma, we realize that $m = m^*$ is a special case that needs to be taken care of in discussing the equilibrium, i.e., the last lemma doesn’t show what happens if $\mathbb{P}(X_1(H_0) = m) > 0$ for $m = m^*$. The following lemma complement the above one with this case.

**Lemma 6.** In equilibrium, if either of the following three conditions hold:

1. $m^* \in (z, \Delta H)$ and $\mathbb{P}(X_1(H_0) = m^*) = 0$;

2. $m^* \in (z, \Delta H)$ and $\mathbb{P}(X_1(H_0) = m^*) > 0$ and $\Pi^*_p(H_0) > 0$; or

3. $m^* \notin (z, \Delta H)$,

then $\Pi^*_p(H_i) > 0$ and $\mathbb{P}(X_1(H_i) + H_i \in (H_0, H_1)) = 1$ for $i = 0, 1$.

The above lemmas show that the support of the distribution of $X(\cdot)$ is isolated on possible values of $m$. Combining the last two lemmas we can now fully characterize
the equilibrium. Depending on the value of $\Pi_p^*(H_0)$, two cases are possible. The next lemma describes the case when $\Pi_p^*(H_0) = 0$ and the lemma after that when $\Pi_p^*(H_0) > 0$. From now on let $a = \frac{g}{d}$, a parameter crucial in determining the equilibrium.

**Lemma 7.** If $\Pi_p^*(H_0) = 0$ in equilibrium, then $\lambda \geq \sqrt{\frac{a}{1+a}}$ and $(1 + a)(1 - \lambda)^2 \leq \frac{\Delta H}{z} \leq 1 + 2a - 2\sqrt{a(1+a)}$, the equilibrium (called Type 0) must satisfy:

$$
\mathbb{P}(X_1(H_0) = m^*) = \mu \\
\mathbb{P}(X_1(H_0) \leq z) = 1 - \mu \\
\mathbb{P}(X_1(H_1) = m^* - \Delta H) = 1, \text{ and}
$$

$$
P_1(y) = \begin{cases} 
\mathbb{E}[v] - g(I_q) & \text{if } y < H_0 \\
\mathbb{E}[v] + \frac{m^* - z}{z}d & \text{if } y = m^* + H_0
\end{cases}
$$

where $\mu \equiv \frac{1-\lambda}{\lambda} \left[(1 + a) \frac{m^*}{z} - 1\right]$.

**Lemma 8.** Suppose that in equilibrium $\Pi_p^*(H_i) > 0$ and $\mathbb{P}(X_1(H_i) + H_i \in (H_0, H_1)) = 1$ for $i = 0, 1$, then there exist at most two constants $y_1^*$ and $y_2^*$ both in $(H_0, H_1)$ such that

1. $\mathbb{P}(X_1(H_i) + H_i = y_j^*) > 0$ for $i = 0, 1$ and $j = 1, 2$, and

2. $\sum_j \mathbb{P}(X_1(H_i) + H_i = y_j^*) = 1$ for $i = 0, 1$.

Combining all the above Lemmas together with the thresholds found in Lemmas we have a full description of the equilibrium in the following theorem.

**Theorem 9.** Depending on the parameters of the model including $\lambda, a = \frac{g}{d}, \frac{\Delta H}{z}$, the equilibrium is one of the following three types.
• **Type 0.** (Lemma 7) $\lambda \geq \sqrt{\frac{a}{1+a}}$ and $(1+a)(1-\lambda)^2 \leq \frac{\Delta H}{\Delta H} \leq 1+2a-\sqrt{a(1+a)}$: Let $\mu = \frac{1-\lambda}{\lambda} [(1 + a)^{\frac{m^*}{z}} - 1]$, then

$$P(X_1(H_0) = m^*) = \mu$$

$$P(X_1(H_0) \leq z) = 1 - \mu$$

$$P(X_1(H_1) = m^* - \Delta H) = 1$$

$$P_1(y) = \begin{cases} 
\mathbb{E}[v] - g(I_q) & \text{if } y < H_0 \\
\mathbb{E}[v] + \frac{m^* - z}{m^*} d & \text{if } y = m^* + H_0
\end{cases}$$

• **Type 1.** $\frac{\Delta H}{\Delta H} < 1 - 2\lambda \left(1 + a - \sqrt{a(1+a)}\right)$: Then for $i = 0, 1$,

$$P(X_1(H_i) + H_i = (1 - \lambda)H_1 + \lambda H_0) = 1,$$

$$P(P_1(X_i(H_i) + H_i) = \mathbb{E}[v] - (1 - \lambda)g + \lambda d) = 1.$$  

• **Type 2.** $\lambda < \frac{a}{1+a}$ and $\frac{\Delta H}{\Delta H} > 1 - 2\lambda \left(\sqrt{a(1+a)} - a\right) \frac{1+a}{a}$: Then for $i = 0, 1$,

$$P(X_1(H_i) + H_i = H_1 - \lambda(1 + \frac{1}{a})\Delta H) = 1,$$

$$P(P_1(X_i(H_i) + H_i) = \mathbb{E}[v] - (1 - \lambda)g + \lambda d) = 1.$$  

There are three points worth mentioning here. First, with Theorem 9 we can discuss the uniqueness of the equilibrium within different regions of the parameter
space of \((\lambda, a)\). Let

\[
\begin{align*}
f_1(a) &= \frac{a}{1 + a} \\
f_2(a) &= \sqrt{\frac{a}{1 + a}} \\
f_3(a) &= \frac{1}{2 \left( 1 + a - \sqrt{a(1 + a)} \right)}
\end{align*}
\]

\[
\begin{align*}
K_1(\lambda, a) &= 1 - 2\lambda \left( 1 + a - \sqrt{a(1 + a)} \right) \\
K_2(\lambda, a) &= 1 - 2\lambda \left( \sqrt{a(1 + a)} - a \right) \frac{1 + a}{a}
\end{align*}
\]

Denote the four regions in the following figure from bottom up by I, II, III, and IV, then the equilibrium can be described as:

**Figure 2.9:** Parameter Ranges for Equilibrium

- Region I: If \(\frac{\dot{z}}{\Delta H} < K_2(\lambda, a)\) then there is unique Type 1 equilibrium. If \(K_2(\lambda, a) < \frac{\dot{z}}{\Delta H} < K_1(\lambda, a)\) then there are both Type 1 and 2 equilibria. If \(\frac{\dot{z}}{\Delta H} > K_1(\lambda, a)\) then there is a unique Type 2 equilibrium.
Region II: When \( \frac{\Delta H}{\Delta H} < K_1(\lambda, a) \) there is unique Type 1 equilibrium.

Region III: Mix of Type 0 and 1 equilibrium.

Region IV: Unique Type 0 equilibrium.

Second, for any \( \lambda \) and \( \frac{\Delta H}{\Delta H} \), there exists an \( a^*(\lambda, \frac{\Delta H}{\Delta H}) \) such that one can always find a unique equilibrium of Type 1 for all \( a > a^*(\lambda, \frac{\Delta H}{\Delta H}) \). In other words, as long as the inventory controlled by the warehouse manipulator is sufficiently large, the futures market can be described by Type 1 equilibrium. To see this, note that for any \( \lambda \) we can always choose \( a \) large enough such that \( (\lambda, a) \) falls into region I because \( f_1(a) \rightarrow 1 \) as \( a \rightarrow \infty \). Next note that for any \( \frac{\Delta H}{\Delta H} \), one can always choose \( a \) large enough such that \( \frac{\Delta H}{\Delta H} < K_2(\lambda, a) \) because \( K_2(\lambda, a) \rightarrow 1 \) as \( a \rightarrow \infty \).

Third, in Type 1 equilibrium, the ex-ante expected payoff of the manipulator is

\[
\Pi^*_M = \lambda \Pi^*_P(H_0) + (1 - \lambda) \Pi^*_P(H_1) = \lambda(1 - \lambda)(d + g)\Delta H - \lambda zd.
\]

The ex-ante payoff of the exogenous hedgers is

\[
\Pi^*_H = -\lambda(1 - \lambda)(d + g)\Delta H.
\]

The exogenous hedgers experience a welfare loss that is transferred to the manipulator. Although the hedgers gain when there is a squeeze, their loss is even larger when there is a warrant cancellation. More generally, in all the possible equilibria of the model, the welfare distribution among the agents is as follows: In the above table, each rows shows the profits of each type of agents upon a realization of exogenous hedge shock of \( H_0, H_1 \), and the ex-ante expected profits. A plus sign means a positive profit and a negative sign means a negative profit. Only the manipulators always have positive profits in all contingencies.
Table 2.4: Welfare Distribution among Agents

<table>
<thead>
<tr>
<th>Agents</th>
<th>$H_0$</th>
<th>$H_1$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manipulator</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Speculators</td>
<td>−</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Hedgers</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

2.5 Conclusion

A model is built to show how the commodity futures market manipulator can combine classical squeeze technique with warehouse queue control. This ability helps the manipulator distort market prices and gain profits under all market contingencies.

In my model, when the exogenous hedging demand is low, the manipulator can take a large long position and carry out a classical short squeeze. When the exogenous hedging demand is high, the manipulator can take a short position and force to deliver goods that are blocked behind a long warehouse queue which is created by himself. Since the small speculators only see the net open interest, they are unable to distinguish the two cases and therefore always get caught by the manipulator.

This model’s equilibrium results align with the empirical findings on the LME aluminum markets in the following three aspects. 1) The contangoes are higher than that can be explained by the storage cost. 2) The term spread goes from contango to deep backwardation frequently meanwhile there are huge inventory in the warehouses. 3) The sharp increases in load-out queues only happen in contango markets and never happen in deep backwardation markets.
Chapter 3

A Model of Commodity Middlemen with Aggregate Shocks

3.1 Introduction

This chapter aims to analyze the functioning of Commodity Trading Firms (CTFs) and their effects on price dynamics and welfare in commodity markets. In the following, I will refer to CTFs as middlemen as this is their essential function. Some interesting questions include how will the spot price volatility change when there are more CTFs in the commodity market? Since CTFs promote and bridge trades between producers and consumers, is it always beneficial to have more CTFs in the commodity market?

To answer these questions, this chapter establishes a model of commodity middlemen with aggregate demand shocks. The middlemen in the model are equipped with advantageous information in the sense that they can allocate goods with full knowledge of the realized aggregate shock in each period. In contrast, producers do not see the aggregate shock when they make their allocation decisions and sometimes end up selling the wrong goods at the wrong time or to the wrong buyers. This difference in ability between the two groups is the driving force of the model. Equilibrium quantities including price, trading volume, and the measure of different groups of agents all change along a realization of independent and identically distributed shocks. The dynamics are analyzed under different parameter settings to answer the above two questions. I find the answers to be no to both of them.
The basic idea of the model is as follows. Before a good is consumed, it must be produced and then sold to buyers if and when they get utility from consuming it. While the producers are specialized in production, the middlemen are better at allocating the goods to meet the shifting demand patterns. This is particularly so in the global commodity trading world, as can be seen from the quote in the first chapter. However, middlemen in general not only promote trade by intermediating the goods but also act as competitors in the spot market by selling the goods. The producers, on the one hand, may need the middlemen to enable them to dump their goods when they cannot correctly direct them to where their consumption is valued. On the other hand, when the producers learn the demand patterns or get lucky so that they can direct the goods on their own, they prefer to be the only sellers in the market without competing sales by middlemen. In the presence of aggregate demand shocks, these two forces come into effect interchangeably and generate interesting dynamics of various market quantities.

From a welfare perspective, each good that trades through middlemen takes a longer time to reach consumers, resulting in a delayed cost. However, the middlemen in general can better allocate the goods to the consumers, which is valuable. These two competing forces depend on various factors such as the degree of informational difference facing the two groups and the time discount imposed by the market. It is obvious that the number of middlemen relative to the number of producers will affect the total surplus and the welfare of each market participant. However, it is unclear whether and how an increase in the number of middlemen in the market will affect these quantities.
3.1.1 Related Literature

A seminal paper in the literature on middlemen is that of Rubinstein and Wolinsky (1987). In their paper, the middlemen have an exogenously higher matching rate in a random search framework, thus trading through middlemen can reduce the cost of delayed trading. The subsequent literature then sought to provide a micro-foundation or realistic assumptions explaining why middlemen may have a higher matching rate.

One strand of this literature assumes that middlemen have private information about the quality of goods (Biglaiser (1993), Li (1998)). One example of such a market is the second-hand car market. Used car buyers tend to buy from platforms such as Carmax but not from individual used car sellers because they do not have the ability to assess the quality of the car. Another strand of this literature assumes heterogeneous goods and idiosyncratic preferences of buyers (Shevchenko (2004), Johri and Leach (2002)). This is a story of supermarkets: supermarkets help the buyers and sellers meet their specific needs. Related work by Watanabe (2010) goes one step further by assuming homogeneous goods and differential storage capacity. In his model, the only key feature that distinguishes middlemen from the producers is that they can store more goods. Finally, another strand of the literature focuses on the land-consuming industrial product markets: dispersion increases productivity while concentration promotes trade (Tse (2011)).

My model is different from those in the above papers in the following three aspects. 1) My model is driven by aggregate shocks while all of their models are driven by idiosyncratic shocks. It is generally hard to deal with aggregate shocks because they make the model unstationary. The equilibrium of the model is described by the invariant measure of an iterated function systems, which is a special type of dynamic system. 2) My model incorporates the observation that middlemen have an informational advantage over producers about demand patterns, in contrast to the existing
literature where the informational advantage is over buyers about the supply patterns.

3) My model can also be used to analyze the time series properties of the equilibrium quantities whereas the existing literature focus on cross-sectional differences.

3.2 Model Setup

Let us consider an infinite period model with three type of agents (producers, middlemen, and buyers) and two types of goods ($A$ and $B$).

3.2.1 Agents

Buyers are denoted $B$. The total measure of $B$ is $l_B > 0$. All of the buyers in each period demand one unit of good of the same type. Consider this demand structure as an extreme form of aggregate shock. The demanded type is independent across periods and is drawn from a Bernoulli distribution with probability $p$, i.e., in each period the buyers will all demand one unit of type $A$ good with probability $p$ or will all demand one unit of type $B$ good with probability $1 - p$. In each period, the buyers’ valuation of their demanded type of good is $u > 0$ and their valuation of the type of good not demanded is zero. After the transaction, whether the buyers bought the demanded good or not, they leave the market. At the beginning of each period, buyers come into the market as new buyers. Thus, the buyers in this market can neither predict the next period’s demand nor use storage.

Each producer can hold at most one unit of good. The producers’ valuation of either type of good is normalized to zero. Thus, the producers will not consume their own good; instead, they will seek to sell it to either buyers or middlemen. At the end of each period, if a producer has sold her good then she can produce another unit of good; if instead she failed to sell her good then she cannot produce and must take
that unit of good into the next period. Thus, at the beginning of each period, every producer holds exactly one unit of good. Denote the producers at the beginning of each period by \( P \) and their measure by \( l_P \). The type of good she holds is not relevant (details given below) because she can transform her good type without any cost. After the transformation, the buyers’ preference is revealed. If a producer has transformed her good to the revealed type in a period, she can then possibly transact with buyers. These producers are denoted by \( S \) and the total measure of all \( S \) is \( l_S \). The producers who transformed their good to the wrong type (i.e., different from the buyers’ revealed type) cannot sell to buyers as a result of the type mismatch. These producers are denoted by \( K \) and their total measure is \( l_K \). Thus, the total measure \( l_P \) producers are split into two groups after the buyers’ type is revealed. In each period, \( l_P = l_S + l_K \). The total measure of producers \( l_P \) is assumed to be constant over time. It is implicitly assumed that there is no exit or entry into the producers industry. However, in each period, the decomposition of \( l_S \) and \( l_K \) depends on the producers’ endogenous transformation decision.

Similarly, each middleman can hold at most one unit of good. Unlike producers, middlemen cannot produce, so the only way for them to acquire the good is by buying from the producers. Middlemen’s valuation of both type of goods is also normalized to zero. However, they can profit from buying at a low price from the producers and sell at a high price to the buyers. In each period, there are two types of middlemen in the market: those who have a unit of good in hand and want to sell it, denoted by \( M \), and those who don’t hold any good and want to buy it, denoted by \( N \). As with the producers, the type of good \( M \) holds in the beginning of each period is not relevant because she can transform her good type without cost. Furthermore, at the time of transformation, middlemen are assumed to have perfect information about the buyers’ demanded type. Each middleman is allowed to make only one transaction
in each period. If an \( N \) middleman bought a unit of good, then she has to wait until next period when she becomes an \( M \) middleman to sell it. And vice versa, if an \( M \) middleman sold her good in one period, then she must wait until the next period when she becomes an \( N \) middleman to restore the good. The total measure of \( M \) and \( N \) is held constant across periods; that is, it is implicitly assumed that there is no entry into or exit from the middlemen industry.

For future reference, here is a summary of the notation introduced so far.

- \( P \): all the producers in the market.
- \( S \): the portion of producers \( P \) who successfully transform their good type to the buyers’ demanded type.
- \( K \): the portion of producers \( P \) who fail to transform their good type to the buyers’ demanded type.
- \( M \): the portion of middlemen who hold a unit of good at the beginning of a period and want to sell it.
- \( N \): the portion of middlemen who has do not hold a good at the beginning of a period and want to buy it.
- \( B \): the buyers all of whom demand exactly the same type of good in each period.
- \( l_i \): the measure of participants with \( i \in \{ P, S, K, B, M, N \} \).

### 3.2.2 Goods

There are only two types of goods, \( A \) and \( B \), in the market. For example, consider commodity trading firms (CTF) in the spot markets for various commodities.
The specialty of the CTFs lies in their ability to transform goods in various dimensions, including spacial transformations through shipment and transportation, temporal transformations through storage, and form/purity transformations through refinement, blending or distillation. This model abstracts from the specific dimension of the transformation and instead focuses on the fact that CTFs are better at transforming the goods to forms that are in high demand. Thus, A and B can be thought about as the copper in China or the copper in Chile. Since the Chilean copper miners don’t have that many local representatives and connections in China as Glencore or Vitol do, they are less informed about local demand patterns and less capable of directing their products to the right locations.

In this model, the transformation cost is assumed to be zero. A possible extension is to introduce positive or even a time-varying cost into the model. However, generally speaking, including transformation cost could only strengthen the points made in this chapter. This is because the embedded value of information about aggregate demand shock would increase as the producers would be more reluctant to make the transformation, unless they are certain that the demanded type is different from the type of good in hand. Another reason for assuming zero transformation cost is that this simplifies the analysis and makes the exposition cleaner. With a strictly positive transaction cost, the state of the market in each period is characterized by four quantities: the measures of producers or middlemen holding good in type A or B. As seen below, if free transformations are allowed instead, the state of the market is one-dimensional: the measure \( l_M \) of middlemen with the good.

### 3.2.3 Timeline

At the beginning of each period, on the supply side, there is a measure \( l_P \) of producers and a measure \( l_M \) of middlemen, each with one unit of good in hand. On the demand
side, there is a measure $l_B$ of buyers and a measure $l_N$ of middlemen who hold nothing and seek to buy one unit of good. The state of the market can be characterized by $\{l_P, l_B, l_M, l_N\}$. Because of the assumption that the total mass of producers ($l_P$), middlemen ($l = l_M + l_N$) and buyers ($l_B$) are constant across periods, the state of the market is effectively one-dimensional: $l_M$. Then the aggregate demand type of the buyers is realized and observed by middlemen but not by the producers. This might correspond to a situation in which CTFs have better information about the local demand patterns than the producers do. With this information structure, the suppliers (including $S$ and $M$) make their transformation decisions. Obviously, the middlemen $M$ will always transform to the buyers’ observed type for otherwise they can only sell to $N$, who do not value the good and so are unwilling to pay any positive price for it. On the other hand, due to their lack of information, the producers $P$ can only randomly transform between $A$ and $B$.

After the transformation is done, the buyers’ preference is revealed to all market participants and becomes public information. At the revelation of this piece of information, the producers $P$ get categorized into $S$ and $K$ based on whether their good type matches the buyers’ demand. The state of the market after the revelation of the demand type is two-dimensional: $(l_M, l_S)$. Then trading begins. Note that there are at most three types of sellers in the markets ($S$, $K$, $M$) and two types of buyers in the market ($B$, $N$). As seen from Figure 3.1, the top three nodes are the potential sellers and the bottom two nodes are the potential buyers. The arrows show the direction of good flows. Note that the two lines starting from $S$ are dotted, indicating producers who possess a matched good may or may not sell it in equilibrium, as shown below.
The matching and transaction between sellers and buyers are modeled using a directed search framework. In the directed search framework, each seller posts a price at which she is willing to sell. Upon observing all the posted prices, the buyers pick a seller to transact with. The market is then split into several submarkets identified by the same posted price and the same type of sellers and buyers. The matching between sellers and buyers within each submarket is described by a matching function \( \alpha(\cdot) \) of the market tightness measure defined by the ratio of the total buyers over the measure of total sellers in this submarket. For example, suppose a set of buyers \( n_b \) direct their search toward a set of middlemen \( n_m \), and then the probability a middleman meets a buyer is \( \alpha(n) \) where \( n = n_b/n_m \) is called the market tightness. Similarly, the probability a buyer meets a middleman is \( \alpha(n)/n \). The matching function \( \alpha \) is the same across all submarkets and has the following properties: \( \alpha(\cdot) \) is increasing and concave, \( \alpha(0) = 0, 0 \leq \alpha(\cdot) \leq 1, \lim_{n \to 0} \alpha'(n) = \infty \).

To summarize, the timeline is as follows.

- Each period begins with \( l_P \) and \( l_M \) holding one unit of good and \( l_B \) and \( l_N \) not holding any.

- The buyers’ (\( B \)) demand shock is realized and observed only by the middlemen.

- \( P \) and \( M \) choose whether or not to transform the type of the good they hold.
• The demand shock is revealed to the public; this leads to masses of \((l_S, l_K, l_M)\) on the supply side and \((l_B, l_N)\) on the demand side.

• Search and trading take place within a directed search framework.

• The producers \(P\) who sold their good produce a new one and the producers \(P\) who failed to sell their good carry it to the next period. The middlemen \(M\) who sold their good become \(N\), and the middlemen \(N\) who bought a good become \(M\).

3.3 One-Period Analysis

First consider the agents’ optimization problems in any given period. For the moment, let each type of agents’ measure \(\{l_i\}\) and continuation value \(\{V'_i\}\) with \(i \in \{S, B, M, N, K\}\) be exogenously given. The goal is to solve for the following endogenous quantities.

• The measure of each type of agents in different submarkets: \(\{l_{SB}, l_{SN}, l_{BS}, l_{BM}, l_{NK}, l_{NM}\}\).

• The present value of each type of agent: \(\{V_i\}\) with \(i \in \{S, B, M, N, K\}\).

• The price prevailing in each submarket: \(\{P_{SB}, P_{SN}, P_{MB}, P_{KN}\}\).

This can be done in two steps. In the first step, solve for a particular submarket and then, in the second step, combine all the submarkets and impose market clearing conditions to get the exact quantities.

3.3.1 One Submarket

A submarket \(\mathcal{M}_{ij}\) is described by the tuple \((l_i, l_j, V'_i, V'_j, V'_{T(i)}, V'_{T(j)})\) where \(i \in \{S, M, K\}\) denotes the sellers, \(j \in \{B, N\}\) denotes the buyer in a submarket, and \(T(\cdot)\) is a state
 transforming function upon a successful transaction. For example, the fact that a middleman $M$ who successfully sell her good becomes $N$ next period can be written as $T(M) = N$. Given these quantities, we want to get the equilibrium quantities $(V_i, V_j, P_{ij})$. The sellers solve the following problem when posting prices:

\[
V_i = \max_{n_{ij}, P_{ij}} \alpha(n_{ij}) \left( P_{ij} + rV'_{T(i)} \right) + (1 - \alpha(n_{ij})) \left( 0 + rV'_i \right)
\]

\[
s.t. V_j = \frac{\alpha(n_{ij})}{n_{ij}} \left( u_j - P_{ij} + rV'_{T(j)} \right) + \left( 1 - \frac{\alpha(n_{ij})}{n_{ij}} \right) \left( 0 + rV'_j \right).
\]

In the above, $n_{ij} = l_j/l_i$ is the market tightness in this submarket, and $u_j$ is the private valuation of the good by the buyers of type $j$. Although $V_j$ should be determined in equilibrium, let the individual buyers take it as given for now.\footnote{This is called “market utility approach” and is standard in solving directed search models. See \textit{Wright et al.} (2017).} Solve for $P_{ij}$ from the constraint and substitute it into the objective function, we get the following first-order condition:

\[
\alpha'(n_{ij}) = \frac{V_j - rV'_j}{u_j + r(V'_{T(j)} - V'_j) + r(V'_{T(i)} - V'_i)}.
\]

Denote the above denominator as $Z_{ij}$, which is the total surplus upon a successful transaction over a failed transaction. Now we can solve for $(V_i, V_j, P_{ij})$ in terms of $(l_i, l_j, V'_i, V'_j, V'_{T(i)}, V'_{T(j)})$:

\[
V_j = \alpha'(n_{ij})Z_{ij} + rV'_j,
\]

\[
V_i = (\alpha(n_{ij}) - n_{ij}\alpha'(n_{ij}))Z_{ij} + rV'_i,
\]

\[
P_{ij} = u_j + r(V'_{T(j)} - V'_j) - \frac{n_{ij}\alpha'(n_{ij})}{\alpha(n_{ij})}Z_{ij}.
\]

Define $\xi = n_{ij}\alpha'(n_{ij})/\alpha(n_{ij})$, which is the elasticity of $\alpha(\cdot)$ with respect to $n_{ij}$. Also, note that the price $P_{ij}$ splits the ex post (after meeting) surplus $Z_{ij}$ according to
\( S_i = (1 - \xi)Z_{ij} \) and \( S_j = \xi Z_{ij} \). In other words, the ex-ante payoff \( V_i \) and \( V_j \) can also be written as

\[
V_i = \alpha(n_{ij})(1 - \xi)Z_{ij} + rV_i' \quad \text{and} \quad V_j = \frac{\alpha(n_{ij})}{n_{ij}} \xi Z_{ij} + rV_j'.
\]

### 3.3.2 Combining Submarkets

When multiple types of agents are on both sides of the market, an appropriate definition of equilibrium is required. This model adopts the definition of a two-sided market with heterogeneous agents from Wright et al. (2017), with a mild modification. The game is played by first letting all the potential sellers post their prices simultaneously. After observing all the posted prices, the potential buyers simultaneously choose a type of agent and a specific posted price to search for. For each type of sellers who post the same price, the buyers who choose to trade with this particular type and price choice are matched with them. The matched buyers and sellers form a submarket, which can be uniquely indexed by the pair of posted price and sellers’ type \((p, t)\). Thus defined, a submarket has only one type on the sellers’ side but could possibly have many types of buyers. Within each submarket, buyers and sellers are matched by a search function \( \alpha(\cdot) \), which is an increasing and concave function of the market tightness in this particular submarket.

Let \( t_1 \) and \( t_2 \) denote the types on the sellers’ and buyers’ side of the market, with \( N_1(t_1) \) and \( N_2(t_2) \) being their respective probability distribution on supports \( T_1 \) and \( T_2 \).\(^2\) Let \( n_1(p, t_1) \) be the mass of type \( t_1 \) agents posting prices weakly below \( p \). Similarly, let \( n_2(p, t_1, t_2) \) be the mass of \( t_2 \) agents who choose to search for prices

\(^2\)As defined in Wright et al. (2017), \( T_1 \) and \( T_2 \) are continuum ordered sets. However, in this chapter, these sets are finite sets, with each type of agents having a positive mass. Moreover, the types cannot be easily ordered.
weakly below \( p \) among \( t_1 \) type sellers. Let \( m_2(p, t_1) \) be the marginal of \( n_2 \) over types \( t_2 \).\(^{\text{3}}\) As mentioned above, a key variable is the market tightness in each submarket. If a pair \((p, t_1)\) attracts a mass of agents on both sides, then the tightness is defined in the obvious way. However, it is possible to let the measure on both sides shrink to zero, then the tightness measure is defined as the Radon-Nikodym derivative \( n(p, t_1) = \frac{dm_2(p, t_1)}{dn_1(p, t_1)} \). Note that \( n_1 \) and \( n_2 \) describe the actions of the agents on the equilibrium path. The belief of agents is described by \( \phi(t_2; p, t_1) \), the probability mass function of type \( t_2 \) buyers in the submarket \((p, t_1)\). Obviously, for the submarkets \((p, t_1)\) that are on equilibrium path, \( \phi \) and \( n \) should be consistent with and determined almost everywhere by \( n_1 \) and \( n_2 \). However, for the off-equilibrium path, \( \phi \) and \( n \) are yet to be determined.

Given the actions and beliefs specified above, one can compute the expected value of different agents. The probability that a seller in submarket \((p, t_1)\) matches with a buyer is denoted by \( \alpha_1(p, t_1; n) = \alpha(n(p, t_1)) \). The probability that a buyer in submarket \((p, t_1)\) matches with a seller is denoted by \( \alpha_2(p, t_1; n) = \frac{\alpha(n(p, t_1))}{n(p, t_1)} \). Let \( v_1(p, t_1, t_2) \) denote the payoff of the seller in a successful trade where the seller is \( t_1 \), the buyer is \( t_2 \) and the trading price is \( p \). Let \( v_2(p, t_1, t_2) \) denote the payoff of the seller in a failed trade. Let the subscript 2 denote the payoff of buyers. Then the expected payoffs are

\[
V_1(p, t_1; n, \phi) = \alpha_1(p, t_1; n) \mathbb{E}_{\phi(\cdot; p, t_1)} v_1(p, t_1, t_2) + [1 - \alpha_1(p, t_1; n)] \mathbb{E}_{\phi(\cdot; p, t_1)} v_2(p, t_1, t_2)
\]

\[
V_2(p, t_1, t_2; n) = \alpha_2(p, t_1; n) v_2(p, t_1, t_2) + [1 - \alpha_2(p, t_1; n)] v_2(p, t_1, t_2).
\]

In the above first equation, note that the expectation is taken with respect to the probability mass function \( \phi(\cdot; p, t_1) \) of various \( t_2 \) types on the buyers’ side within the

\(^{3}\)In this chapter, since the set \( \mathbb{T}_2 \) is finite, \( m_2(p, t_1) = \sum_{t_2 \in \mathbb{T}_2} n_2(p, t_1, t_2) \). However, in general, the definition used here also works for continuous type spaces.
submarket \((p, t_1)\). Also, in this equation, all \(t_1\) could be replaced by \(i\) as there is only one type of seller on the sellers’ side within each submarket. For the same reason, the expectation sign can be removed from the second equation.

This chapter uses the concept of subgame perfection to restrict \(n\) and \(\phi\) off the equilibrium paths. Let \(\Gamma_1 = \{(p, t_1) \in \mathbb{R}_+ \times T_1 | \frac{dm_2(p, t_1)}{dn_1(p, t_1)} > 0\}\) denote the set of submarkets that actually exist in equilibrium. Let \(\Gamma_2 = \{(p, t_1, t_2) | (p, t_1) \in \Gamma_1 \text{ and } \phi(t_2;p, t_1) > 0\}\) be the set of all equilibrium submarkets combined with their type of buyers. Let 
\[
U(t_2) = \sup_{(p, t_1) \in \Gamma_1} V_2(p, t_1, t_2; n)
\]
be the maximum utility a type \(t_2\) buyer can get (called market utility). Then it makes sense to assume that \(t_2\) buyers can be expected in a submarket if and only if they can get their market utility \(U(t_2)\).

**Condition P:** If \(n(p, t_1) > 0\) then the support of \(\phi(t_2;p, t_1)\) is nonempty and contains only \(t_2 \in T_2\) such that \(V_2(p, t_1, t_2; n) = U(t_2)\).

A potential brute force way to support an equilibrium is to assume that \(n(p, t_1) = 0\) for all \((p, t_1)\) on off-equilibrium paths. To rule out this case, let \(n(p', t_1) > 0\) for all \(p'\) close enough to some \(p\) on the equilibrium path. This is saying that, if some seller deviates from the equilibrium by moving her posted price by a small amount, then she is still able to attract some buyers.

**Condition N:** For each \((p, t_1) \in \Gamma_1\) in equilibrium, there exists an \(\epsilon > 0\) such that \(n(p', t_1) > 0\) for all \(p' \in (p - \epsilon, p + \epsilon)\).

The equilibrium is defined as the set of actions the price posted by the sellers and the beliefs of buyers and their choices on where to buy the goods. The beliefs describe what will be the composition of a submarket if price is posted some other way.

**Definition 2.** The equilibrium is a set of profiles \((n_1, n_2, n, \phi, V_1, V_2, U)\) such that

- (Maximization-seller) \(V_1(p, t_1; n, \phi) \geq V_1(p', t_1; n, \phi)\) for all \((p, t_1) \in \Gamma_1\) and \(p' \in \mathbb{R}_+ \times T_1\).

\(^4\)In this case, as the market expects no buyers in these submarkets, it is irrelevant how \(\phi(\cdot;p, t_1)\) is defined for these submarkets.
\[ \mathbb{R}_+. \]

- (Maximization-buyer) \( V_2(p, t_1, t_2; n) \geq V_2(p', t'_1, t_2; n) \) for all \((p, t_1, t_2) \in \Gamma_2 \) and \((p', t'_1) \in \mathbb{R}_+ \times \mathbb{T}_1 \).

- (Consistency) \( n \) and \( \phi \) are consistent with \( n_1 \) and \( n_2 \) on \( \Gamma_1 \).

- (Perfection) Condition P and N are satisfied.

- (Feasibility) For any \( \mathbb{T}_1' \subseteq \mathbb{T}_1 \),

\[
\int_{\mathbb{R}_+ \times \mathbb{T}_1'} dn_1(p, t_1) \leq \int_{\mathbb{T}_1'} dN_1(t_1),
\]

with equality if \( \max_{p \in \mathbb{R}_+} V_1(p, t_1; n_1, \phi) > 0 \) for almost all \( t_1' \in \mathbb{T}_1' \). Similarly, for any \( \mathbb{T}_2' \subseteq \mathbb{T}_2 \),

\[
\int_{\mathbb{R}_+ \times \mathbb{T}_1 \times \mathbb{T}_2'} dn_2(p, t_1, t_2) \leq \int_{\mathbb{T}_2'} dN_2(t_2),
\]

with equality if \( \max_{(p, t_1) \in \mathbb{R}_+ \times \mathbb{T}_1} V_2(p, t_1, t_2; n_1, \phi) > 0 \) for almost all \( t_2' \in \mathbb{T}_2' \).

The maximization conditions say the agents participate only in markets that maximize their payoff. The consistency condition says that the strategies \( n_1 \) and \( n_2 \) should be consistent with the belief about the market composition on equilibrium paths. The perfection conditions restrict the belief on the off-equilibrium paths. The feasibility condition specifies that the number of agents of a particular type participating in different submarkets add up to be less or equal to the total mass of such agents. The equality holds only when all the types attain strictly positive payoff in some submarkets.
Assumption 1. Agents’ payoff functions $v_1, v_2, v_1', v_2'$ have the following structure:

\[
\begin{align*}
v_1(p, t_1, t_2) &= d_1(t_1) + p, \\
v_2(p, t_1, t_2) &= d_2(t_1, t_2) - p, \\
v_1'(p, t_1, t_2) &= r_1(t_1), \\
v_2'(p, t_1, t_2) &= r_2(t_2).
\end{align*}
\]

In the first equation, $d_1(t_1)$ can be considered as the non-monetary payoff to the seller upon a successful transaction. Later in this chapter, this will be the continuation value of staying in the market. Moreover, this non-monetary payoff is independent of the buyer’s type, i.e., a seller does not care to whom she sells to. In the second equation above, $d_2(t_1, t_2)$ depends on both the seller’s and buyer’s types. This is to capture the mismatch between the seller’s supply and the buyer’s demand, i.e., the buyer needs a good of type $A$ while the seller holds a good of type $B$. The last two equations above say that the reservation value of an agent is only related to her own type but not the specificities of the failed transaction. The next proposition characterizes the behavior of the equilibrium when $T_2$ is finite, which will be helpful in deriving the equilibrium quantities analytically.

**Proposition 10.** When Assumption 1 holds and $T_2$ is finite, no submarket in equilibrium contains all types of buyers.

Proposition 10 is particularly useful when there are only two types of buyers in the market because it implies that each submarket in equilibrium has only one type of buyer. If there are a finite set of sellers then the number of submarkets in equilibrium is at most finite. Now consider the specific setting in this model with the following characterization. Let $\Omega = \{P, B, M, N\}$ be the set of all agents at the beginning of a period. And let $T_1 = \Omega_S = \{S, M, K\}$ be the set of types of potential sellers and
\( T_2 = \Omega_B = \{ B, N \} \) be set of the types of potential buyers. Given the continuation values \( \bar{V}' = \{ V'_i \}_{i \in \Omega} \) we can map these values to the payoff functions in Assumption 1:

- \( d_1(S) = d_1(K) = V'_p \) and \( d_1(M) = V'_N \)
- \( d_2(S, B) = d_2(M, B) = d \) and \( d_2(K, B) = d_2(M, N) = 0 \) and \( d_2(S, N) = d_2(K, N) = V'_M \)
- \( r_1(S) = r_1(K) = V'_p \) and \( r_1(M) = V'_M \)
- \( r_2(B) = 0 \) and \( r_2(N) = V'_N \)

Now that the payoff functions are defined in the model, the set of \((n_1, n_2, n, \phi)\) for the equilibrium is needed. Since \( n_1 \) and \( n_2 \) characterize behavior of the agents on the equilibrium path and it has been shown that in equilibrium there are only finitely many submarkets with one type of seller and one type of buyer, this means that the structure of \( n_1 \) and \( n_2 \) is relatively simple and degenerate: \( n_1(p, t_1) \) should be a step function over \( p \) for all \( t_1 \in \Omega_S \) and \( n_2(p, t_1, t_2) \) should be a step function over \( (p, t_1) \) for all \( t_2 \in \Omega_B \). If we denote the measures \( \bar{L} = \{ l_i \}_{i \in \Omega} \), we can then reformulate Definition 2 catering to the specific setting of my model as follows:

**Definition 3.** The *one-period market equilibrium* is a set of configurations

- \( \{ l_{ij} \}_{i,j \in \Omega} \) the measure of agents from type \( i \) who trade with type \( j \),
- \( \{ V_i \}_{i \in \Omega} \) the set of present values of all type of agents, and
- \( \{ P_{ij} \}_{i \in \Omega_S, j \in \Omega_B} \) the prices prevailing in the submarket composed of type \( i \) sellers and type \( j \) buyers

---

5 This definition is shared by [Watanabe 2010](#). He assumes that “all buyers use identical mixing strategies for any configurations of the announced prices including those where suppliers deviate from the equilibrium.” Moreover, his description of the buyers search behavior is conditional on a buyer’s expected value in equilibrium, whatever that value turned out to be. This is equivalent to my approach while I am making this search behavior as a constraint in the sellers’ optimization problem.

---

57
such that the following conditions hold:

- For all $i, j$ pair, if the submarket exists in the sense that $l_{ij} > 0$ and $l_{ji} > 0$, then $(V_i, V_j, P_{ij})$ is the one-submarket optimized result as in the last section given $(l_{ij}, l_{ji})$.

- Feasibility condition: $0 \leq l_{ij} \leq l_i$ for all $i, j$ and $\sum_{j \in \Omega} l_{ij} \leq l_i$ for all $i \in \Omega$.

- Participation condition: $V_i \geq 0$.

Note that the above definition implicitly requires that “market utility” is well-defined. That is, if some type $i$ participate in multiple submarkets, then the optimized present value in each submarket is equalized to be $V_i$. This prevents $i$ to deviate from any of the submarkets that they are active in. Also note that there could be at most four submarkets in the equilibrium including $S \rightarrow B$, $S \rightarrow N$, and $M \rightarrow B$ and $K \rightarrow N$. The submarket $K \rightarrow B$ could not happen because the goods $K$ hold is not what $B$ want. The submarket $M \rightarrow N$ cannot happen as we assumed that the middlemen do not trade with themselves as in {Rubinstein and Wolinsky} (1987). However the number of submarkets could be less than four depending on the measures of different types.

To make this point clear, let us look at a smaller market composed of only \{K, S, B, N\}. The key point here is that $K$ can only sell to $N$ while $S$ can sell to both $B$ and $N$. When $l_K$ is very large and $l_N$ is relatively small, since $K$ has to sell to $N$, they will compete on the submarket $K \rightarrow N$ such that $P_{KN}$ is low and the matching rate for $K$ is low and the matching rate for $N$ is very high. This makes $V_N$ very high. Now although $S$ can sell to $N$, in order for them to compete with $K$ they have to post prices so low in order to reach $N$’s high outside valuation $V_N$, which in turn makes the present valuation of $S$ very low. But when it is so hard for $S$ to compete with $K$ for $N$, they would rather sell to $B$ only and not participate in the
submarket $S \to N$. If this is the case, the submarket $S \to N$ breaks down and we have only two submarkets left as opposed to the maximum amount of three.

With the same idea, one can solve for the one-period equilibrium. In the following, only equilibrium market tightness measure $n_{ij}$ will be reported as all the other quantities including $V$ and $P$ are just functions of $n$ derived in the last section. Given the continuation values $\hat{V}' = \{V_i'\}_{i \in \Omega}$ and the measures $\hat{L} = \{l_i\}_{i \in \Omega}$.

**Proposition 11 (One-Period Market Trading Pattern).** Let $\Delta = V'_M - V'_N$, $c_2 = \left(\frac{u}{\Delta} \right)^\frac{1}{b}$, and $c_3 = (1 - \frac{ru}{v})^\frac{1}{1-b}$. Denote $\tilde{l}_B = c_2 l_B$ and $\tilde{l}_M = c_3 l_M$.

- **Type 0:** If $\frac{\tilde{l}_M}{l_S + l_K} \leq \frac{\tilde{l}_B}{l_N} \leq \frac{\tilde{l}_S + \tilde{l}_M}{l_K}$, then all the four submarkets exist and the tightness for each submarket is

  $$n_{SB} = \frac{l_B + l_N/c_2}{l_S + l_K + c_3 l_M}, \quad n_{SN} = n_{KN} = c_2 n_{SB}, \quad n_{MB} = c_3 n_{SB}.$$  

- **Type 1:** If $\frac{\tilde{l}_B}{l_N} > \frac{\tilde{l}_S + \tilde{l}_M}{l_K}$, then submarket $S \to N$ breaks down and the tightness of each submarket is

  $$n_{SB} = \frac{l_B}{l_S + c_3 l_M}, \quad n_{KN} = c_2 n_{SB}, \quad n_{MB} = c_3 n_{SB}.$$  

- **Type 2:** If $\frac{\tilde{l}_B}{l_N} < \frac{\tilde{l}_M}{l_S + l_K}$, then submarket $S \to B$ breaks down and the tightness of each submarket is

  $$n_{SN} = n_{KN} = \frac{l_N}{l_S + l_K}, \quad n_{MB} = \frac{l_B}{l_M}.$$  

These results are intuitive. Consider first $l_B$. If $l_B$ is very large, then $S$ will be attracted to sell to $B$ and give up the submarket $S \to N$. If $l_B$ is very small, then it would be beneficial for $S$ to compete with $K$ for the market of $N$. If $l_B$ is not
extremely large or small to the either ends, then $S$ will participate in both $S \to B$ and $S \to N$ markets. All the other quantities can be viewed this way.

### 3.4 Infinite Period Analysis

Now consider an infinite period model where, in each period, the one-period game is played. Since this is a model with aggregate shocks, it is impossible to have a stationary equilibrium in which all the quantities remain constant. Instead, a Markovian type equilibrium in which states transit to one another within a bounded set is analyzed. As mentioned above, the state of the market is one-dimensional and can be described by $l_M$. Let the total number of middlemen be $l = l_M + l_N$.

Suppose the quantities of producers, buyers, and middlemen ($l_P, l_B$, and $l$) are fixed. Let $S$ denote an interval covering all the possible values of the state $l_M$. Then, an infinite-period equilibrium has the following three components: 1) The set of continuation value functions of all the agent types $\tilde{V} = \{V_i(\cdot) : S \to \mathcal{R}_+\}_{i \in \Omega_V}$, each defined on the set $S$. 2) The transition probability function $\mathcal{P}(\cdot, \cdot) : S \times S \to [0, 1]$ describing stochastic dynamics of the state variable $l_M$. 3) The producers’ symmetric strategy $\sigma(\cdot) : S \to [0, 1]$, representing the probability with which the producers allocate their good to type $A$ at the beginning of each period.

These three components must be consistent with the one-period equilibrium in Definition 3 in the following sense. For any $s \in S$ there exist $s_1 \in S$ and $s_2 \in S$ such that $\mathcal{P}(s, s_1) = p$ and $\mathcal{P}(s, s_2) = 1 - p$. This is by the Bernoulli assumption on the aggregate demand shocks. Suppose at the beginning of a period with a state $s$, i.e.,

---

6In a finite-state Markov chain, this function degenerates to a transition probability matrix, the $ij$th element of which representing the probability of jumping to state $j$ in next period when the current state is $i$. Thus, each row of this matrix sums up to one. In my case, the state space is possibly infinite dimensional, so the transition dynamics is better characterized by a function defined on $S \times S$. 

60
the number of middlemen $M$ is $s$. After the aggregate shock in this period is realized, there could be two trading structures in the market. If the realized shock is of type $A$, then the next period’s state is $s_1$ according to the transition probability function. Let the continuation value and the market participant measures in this scenario be given by:

$$\Delta = V_M(s_1) - V_N(s_1),$$

$$l_S = \sigma(s) \cdot l_P,$$

$$l_K = (1 - \sigma(s)) \cdot l_P,$$

$$l_M = s.$$ 

The above profile defines a one-period market equilibrium. If the realized shock is of type $B$, then the next period’s state is $s_2$. Similarly, let the continuation value and the market participant measures in this scenario be given by:

$$\Delta = V_M(s_2) - V_N(s_2),$$

$$l_S = (1 - \sigma(s)) \cdot l_P,$$

$$l_K = \sigma(s) \cdot l_P,$$

$$l_M = s.$$ 

The above profile also defines a one-period market equilibrium. Let $\{U_i(s)\}_{i \in \Omega_U}$ and $\{l_{ij}(s)\}_{i,j \in \Omega_U}$ be the present values and the agent measures solving the corresponding one-period market equilibrium. Then, these quantities must satisfy the following conditions in equilibrium.

**Definition 4.** An infinite-period equilibrium is a profile includes the set of continuation value functions $\vec{V}$, the transition probability function $\mathcal{P}(\cdot, \cdot)$, and the produc-
ers’ strategy $\sigma(\cdot)$. These quantities must be consistent with the one-period equilibrium in Definition 3 and satisfy the following conditions:

1. $V_P(s) = \sigma(s) (p \cdot U_S(s_1) + (1 - p) \cdot U_K(s_2)) + (1 - \sigma(s)) (p \cdot U_K(s_1) + (1 - p) \cdot U_S(s_2))$

2. $V_i(s) = p \cdot U_i(s_1) + (1 - p) \cdot U_i(s_2)$ for $i \in \{M, N\}$

3. Incentive Compatibility for Producers: If $\sigma(s) \in (0, 1)$ then

   $$p \cdot U_S(s_1) + (1 - p) \cdot U_K(s_2) = p \cdot U_K(s_1) + (1 - p) \cdot U_S(s_2)$$

4. Incentive Compatibility for Producers (corner case): If $\sigma(s) = 1$ (or 0) then

   $$p \cdot U_S(s_1) + (1 - p) \cdot U_K(s_2) > (\prec) \, p \cdot U_K(s_1) + (1 - p) \cdot U_S(s_2)$$

5. Consistent Middlemen Flows: For $i = 1, 2$

   $$s_i = s + \alpha(l_{NK}/l_{KN}) \cdot l_{KN} + \alpha(l_{NS}/l_{SN}) \cdot l_{SN} - \alpha(l_{BM}/l_{MB}) \cdot l_{MB}$$

These first two conditions are rational expectation constraints, claiming the continuation value of producers should be consistent with both the randomness of the producers mixing allocation strategy ($\sigma$) and the independent aggregate shock ($p$). The third and fourth conditions are incentive compatibility constraints for the producers in making their allocation decisions. The last condition says that the amount of $l_M$ should be consistent with the flows in and out of the $M$ in the one-period market equilibrium. The second and third term on the right-hand side of the equation denote the inflow to $M$ from those $N$ who bought the good. The last term denotes the outflow from $M$ to $N$ for those middlemen who sold their good in the equilibrium.
3.4.1 Existence and Uniqueness of Equilibrium

My definition of the infinite-period equilibrium falls into a category of dynamic systems called iterated function systems in the analysis of fractional mathematics. The proof for the existence and uniqueness of the invariant measure, as well as a density function for the Markov processes, can be found in the work of [Dubins and Freedman (1966)].

**Proposition 12** (cited from [Dubins and Freedman (1966)]). Let \( \Omega \) be a set, \( \Gamma \) be a set of mappings of \( \Omega \) into itself, and \( P \) is a probability on \( \Gamma \). Each \( P \) determines a Markov process with \( \Omega \) for state space and this transition mechanism: when at \( \omega \in \Omega \), choose a \( \gamma \in \Gamma \) according to \( P \), and move to the new state \( \gamma(\omega) \). If \( \omega \) itself is random with distribution \( \mu \), then the distribution of the new state is \( P\mu \). If \( P\mu = \mu \), then \( \mu \) is \( P \)–invariant. If \( \Omega \) is compact metric, \( \Gamma \) is a finite set of contraction mappings on \( \Omega \), and \( P \) assigns positive mass to each \( \gamma \in \Gamma \), then

1. There is one and only one invariant probability \( \mu \).

2. \( \mu \) is continuous (unless there is a common fixed point for all \( \gamma \in \Gamma \), in which case \( \mu \) plainly assigns probability 1 to that point).

3. The support of \( \mu \) is all of \( \Omega \) if and only if each point of \( \Omega \) is in the range of some \( \gamma \in \Gamma \).

4. If \( m \) is a probability on \( \Omega \) and for all \( \gamma \in \Gamma \) the distribution of \( \gamma \) under \( m \) is absolutely continuous with respect to \( m \), then \( \mu \) is either absolutely continuous or purely singular with respect to \( m \).

This proposition can be directly applied to my model. Let \( \Omega \) be an interval of \( l_M \) and \( \Gamma \) be the set of two transition mappings of \( l_M \) upon the two different realizations of a Bernoulli aggregate shock. Suppose we can show that the two transition mappings
are contractions. The first point informs that the Markov chain defined above (or the equilibrium dynamics in this model) enjoys an invariant probability $\mu$. This is very helpful for analyzing the equilibrium dynamics because the existence of an invariant measure immunes the long-term dynamic properties of the system from the initial state, i.e., no matter which $l_M$ we begin with the distribution of a simulated time series of states converge to the same measure. The second and third points inform that the support of $\mu$ doesn’t take irregular shape and is just the whole $\Omega$. Since the Markov chain in this model has countably many states, it is hard to intuitively think about the shape of the set collecting all these states. It turns out that the set should be $\Omega$ itself. The fourth point above ensures us of the existence of a density function on the interval $\Omega$. Note that the existence of a measure along doesn’t translate to the existence of a density function. The latter condition requires that such a measure be absolutely continuous with respect to the Lebesgue measure.

Note that we have yet to specify how to pick the interval $\Omega$ above. The condition that must be satisfied is that the two transition mappings $\gamma$ should both be contractions on $\Omega$. It will be shown in the next section that the two transition mappings intercept with the diagonal only once and the slope at each point is less than one. Therefore, an interval should be chosen such that the two ends correspond to the fixed point of the two transition mappings.

### 3.4.2 A Numerical Approach

When $p \neq 0.5$, it is hard to solve the system analytically. Thus, I proceed with a value function iteration approach. The above definition is instructive in how to carry out the program. In the following, I give the framework for the program and two numerical examples. In the next two sections, we will apply this approach to various $p$ and $l$ values and see how the equilibrium changes accordingly. Since the set of
equilibrium states $S \subset [0, 1]$, we start from a grid on $[0, 1]$. Later on, we will restrict the range based on the pattern of computed value functions to get a finer grid. The steps are as follows.

1. Construct a grid on the unit cubic with the three dimensions being $l_M, l_S, l'_M$.

2. Initialize the value functions $\{V_i\}_{i \in \Omega_V}$.

3. For each point $(l_M, l_S, l'_M)$ on the grid, find the one-period market equilibrium type according to Proposition 11.

4. For each point $(l_M, l_S, l'_M)$ on the grid, compute the one-period market equilibrium quantities and check if the resulting next period level of $l_M$ is consistent with $l'_M$. This step ensures that condition 5 in Definition 4 is satisfied.

5. For each point $(l_M, l_S, l'_M(l_M, l_S))$ compute the one-period market equilibrium present values $\{U_i\}_{i \in \Omega_U}$.

6. For each $l_M$, find the $l_S$ such that condition 3 in Definition 4 is satisfied. If no such $l_S$ exists, then it has to be the case that $\sigma(l_M) = 0, 1$. This step ensures that conditions 3 and 4 are satisfied.

7. For each $l_M$, compute the updated continuation value functions $\{V_i\}_{i \in \Omega_V}$ according to condition 1 and 2 in Definition 4.

8. Check if the updated value functions are globally close to the last iteration of value functions. If not, do one more iteration. If so, stop the iteration process.

This process generates the set of value functions sitting on $[0, 1]$. However, this grid is too coarse to be a useful probability transition matrix. Observe the first row of Figure 3.2 These are the iteration results with parameters with $p = 0.05$ on the
left and \( p = 0.45 \) on the right, with all other parameters being \( l = 0.2, a = 0.5, b = 0.5, r = 0.95, u = 1 \). The red line shows the next period state in equilibrium on the contingency that the buyers’ demand is of type \( A \) for each state \( l_M \). Similarly, the green line shows the next period state in equilibrium on the contingency that the buyers demand is of type \( B \) for each state \( l_M \). I call these two lines state transition mappings from now on. These two transition mappings intersect with the diagonal line in black at the two red points. Since these two mappings are contractive\(^7\) on the interval between their intersections with the diagonal line, the whole Markov chain is sure to fall in this interval. Thus, the area between the two blue dashed vertical lines (call it \( S_0 \)) in both plots is the smallest interval covering the equilibrium set of states \( S \). This is because for each \( l_M \in S_0 \), the next period’s \( l'_M \) never goes out of \( S_0 \), however, for each \( l_M \notin S_0 \), let the process go on and there is strictly positive probability in finite time that the state goes into \( S_0 \).\(^8\) Only one thick blue vertical dashed line can be seen in the right plot because the two blue thin lines are very close to each other.

Once the intervals are found with the above numerical approach, a second round of value function iteration can be done with refined interval ranges to achieve faster convergence. The second row of Figure 3.2 gives such second round results. The two transition mappings tend to overlap when \( p \) is close to 0.5. Indeed, the two transition mappings overlap exactly when \( p = 0.5 \) as will be proved in the following section.

\(^7\)If one considers the local linear approximation of these two mappings, since their slope are obviously less than 1, this ensures a contraction mapping.

\(^8\)In other words, let \( T_1 \) be the stopping time for \( \{l_M \in S_0\} \) and \( T_2 \) be the stopping time for \( \{l_M \notin S_0\} \). Then the following properties hold: For any \( s \notin S_0 \), \( P_s(T_1 < \infty) = 1 \) and for any \( s \in S_0 \), \( P_s(T_2 = \infty) = 1 \).
Figure 3.2: Transition Mappings ($p = 0.05$ on left and $0.45$ on right)

Figure 3.3 plots the values function of producers, middlemen $M$, and middlemen $N$ with red, green, and blue lines. In the following sections, we will compare the expected values of different groups of agents under various parameter settings. The expectation is computed by integrating these lines with respect to the density function (examples are plotted in the second row of Figure 3.4) on their respective interval (or support). One feature of Figure 3.3 is that the values of middlemen when $p = 0.45$ on the right are uniformly greater than when $p = 0.05$ on the left. This is because
when $p$ is closer to 0.5, the producers face more uncertainty as to how to allocate their goods, in which case the middlemen can profit from having the correct information about the buyers’ demand. For the same reason, the value for producers on the left is higher than that on the right. Secondly, $V_M$ is uniformly higher than $V_N$ in both plots because $M$ are those who have a unit of good in hands. However, the difference between $V_M$ and $V_N$ varies over different $l_M$ because these values take into account not only the price of the good but also the probability of a successful matching.

Figure 3.3: Value Functions ($p = 0.05$ on left and 0.45 on right)

With a finer grid on $S_0$ we can approximate the stationary distribution of the equilibrium Markov chain, with which we can further compute interesting quantities such as ex ante total surplus and the welfare of different groups of agents. Figure 3.4 gives some sense of the dynamics of equilibrium for the two parameter settings in Figure 3.2. The three rows plot one realized path of quantities of middlemen $l_M$, the density of $l_M$ to the invariant measure of the Markov chain, and the auto-correlation function of the $l_M$ series. As will be shown later, spot prices is roughly linear in $l_M$, thus we can think of Figure 3.4 as showing middlemen quantities as well as scaled

68
prices. Both price series show volatility but the left plot has more spikes in it, which is a critical pattern in commodity prices. When $p$ is close to 0 or 1, the uncertainty in the market should be small in the sense that the producers know how to allocate their goods with higher confidence. However, whenever the small probability event happens, that also has higher impact on the prices. Accordingly, the price distribution of the right case is quite symmetric while the left distribution is highly skewed. The second row in Figure 3.4 shows this pattern. These empirical density functions are calculated from a simulation of a path of one hundred thousand periods. The results are robust to the initial point of the path, which is indicative of a true density to the invariant measure as shown in the last section. In the third row, the right case shows higher orders of autocorrelation than the left case. This is because when the price is low, it is expected to be low for longer periods of time. Essentially, the two transition mappings in Figure 3.2 are trying to pull the state variable $l_M$ toward their corresponding absorption points (their intersection with the diagonal line). If one considers $p$ as the relative magnitude between these two forces, a $p$ close to 1 or 0 breaks the balance of the two forces and gives more power to one over the other.
Figure 3.4: Temporal Trading Patterns ($p = 0.05$ on left and 0.45 on right)
3.5 Some special cases

3.5.1 When \( l = 0 \)

In this section, we look at two benchmarks where there are only producers and buyers but no middlemen in the market. In the first benchmark, the producers have no information about buyers’ preference. They can only randomly allocate their good so that they are indifferent among allocation choices. In the second benchmark, the producers are assumed to have prefect information about buyers’ preference as middlemen do. I assume a form of the matching function \( \alpha(n) = a \cdot n^b \).

**Proposition 13.** Suppose there is a measure \( l_P \) of producers and a measure \( l_B \) of buyers but no middlemen. If the producers have perfect information about the buyers’ preference, then the unique equilibrium is:

\[
\sigma = 1_A, \\
TS_0 = a \left( \frac{l_B}{l_P} \right)^b l_P u.
\]

If the producers have no information about the buyers’ preference, then depending on the parameter values there are three possible equilibrium:

- There always exists a mixed strategy equilibrium in which

\[
\sigma = \frac{1}{1 + \left( \frac{1-p}{p} \right)^b}, \\
TS = \left( p\sigma^{1-b} + (1-p)(1 - \sigma)^{1-b} \right) \cdot TS_0.
\]
• If \( p > \frac{1}{1 + \alpha(l_B/l_P)} \), then exists a corner solution equilibrium in which

\[
\sigma = 1, \\
TS = p \cdot TS_0.
\]

• If \( p < \frac{1}{1 + \alpha(l_B/l_P)} \), then exists a corner solution equilibrium in which

\[
\sigma = 0, \\
TS = (1 - p) \cdot TS_0.
\]

In the above, the indicator function \( 1_A \) takes a value of one when the buyers’ preference is \( A \) and a value of zero when the buyers’ preference is \( B \). In all the above cases, the buyers’ expected surplus is \( b \) share of the total surplus while producers expected surplus is \( 1 - b \) share of the total surplus. Note that when the producers know the buyers’ preference, the probability \( p \) does not matter in equilibrium. However, when the producers are uncertain about the buyers’ preference, the probability \( p \) affects the equilibrium quantities. Specifically, when there is the most uncertainty\(^9\) in the market \( (p = 0.5) \), the total surplus reaches its minimum, which is equivalent to a market where there are only half the number of producers who can always see the buyers’ preference.

\(^9\)Uncertainty can be measured as the entropy of the Bernoulli distribution: \(-p \ln p - (1-p) \ln(1-p)\), which reaches the maximum at \( p = 0.5 \).
Figure 3.5: Special Case when $l = 0$

Figure 3.5 plots the mixing probability of the producers on the left and the total producers’ surplus on the right. Intuitively, the more likely for the buyers to prefer type $A$ then the producers allocate more likely to type $A$, as in the left plot. On the right, the blue line is the total producer surplus for various $p$ values under the mixed strategy equilibrium. The closer $p$ is to either end, the less uncertainty there is in the market, allowing producers to allocate their goods more correctly and enjoy a higher surplus. The red line shows the total producer surplus under the two corner solution equilibria. These equilibria generate lower producer surplus than the mixed equilibrium. From now on when there exist multiple equilibria for some parameter settings the mixed strategy equilibrium is selected. I make this choice for two reasons. First, working with mixed strategy equilibria ensures the continuity of most equilibrium quantities across various parameter settings. Second, since the producer welfare is higher in mixed strategy equilibria, our results that more middlemen decreases producer surplus can never be the result of our choice of equilibria.

In the following sections, middlemen are introduced to the market and the measure of producers is held constant. Intuitively, the total surplus will increase relative to
when there are only producers selling the goods. This is because the middlemen are more efficient in allocating the goods and they can do this by buying from producers and selling to the buyers in the next period with full information of their preference. Thus, the total number of matches in each period increases and so does the total surplus. It is obvious to see that the buyers are strictly better off with middlemen in existence in the market as they always get $b$ share of the surplus in each transaction. However, it is nontrivial to see how the producers and middlemen split the rest $1 - b$ share of the total surplus. It is also interesting to see how the split changes as the measure of middlemen increases.

### 3.5.2 When $p = 0$ or $p = 1$

When $p = 0$ or $p = 1$, there is no uncertainty in the market and the producers all know exactly how to allocate their goods. It is tempting to conclude that there is no role for the middlemen in such a market as they cannot beat the producers in allocating the goods. However, the truth is that they do participate in the market as “liquidity” providers. The producers always want more buyers in the market but don’t want any competitors. These two driving forces determine the total impact on the producers. Figure 3.6 plots on the left the distribution of total surplus among the three market participants and on the right the decomposition of the additional surplus resulting from the larger number of matched transactions. Note that producers’ share of increased surplus is negative on the right plot, which means that the producers are “subsidizing” or losing profits to the middlemen as they enter the market. However, since in this case the middlemen have no superior information over the producers, this effect is small relative to the case when they are better at information collection, to be shown below.
Figure 3.6: Special Case when $p = 1$

3.5.3 When $p = 0.5$

It can be proved that when $p = 0.5$ there always exists a unique stationary equilibrium in the sense that all the quantities remain constant over periods. This section answers the following three questions: 1) How is the total surplus shared among buyers, producers, and middlemen? 2) As we have more and more middlemen in the market, how will the surplus of producers and buyers change both in absolute and relative terms? 3) How do the quantities of different market participants affect these relationships and why?

Proposition 14. When $p = 0.5$, for appropriate parameters (specified in the proof) there exists a unique stationary equilibrium.

Figure 3.7 illustrates the sequence of stationary equilibrium quantities computed with parameters $a = 0.5, b = 0.5, r = 0.95, u = 1, p = 0.5, l_P = l_B = 1$. The upper left plot shows the decomposition of total surplus among buyers, producers, and middlemen. The total surplus is computed as the buyers’ valuation of good $u$
times the total quantity of goods that end up in the buyers’ hands (through trading with either producers or middlemen). The surplus of buyers or producers is just the quantity of goods they sell times the price of goods as their valuation of the good is normalized to zero. As the quantity of middlemen increases in the system, the middlemen’s share of total surplus increases while the buyers and producers’ share of total surplus decreases.

The upper right plot of Figure 3.7 shows the share of the incremental surplus among the three groups of participants as the quantity of middlemen increases. Supposing middlemen are added to the system, it is intuitive that the total surplus will increase as middlemen promote trades. As is shown in the plot, the middlemen take around 60-70% of the increased surplus, but buyers take only half of that, while producers barely benefit from having more middlemen in the system. When the number of middlemen is small (around 0-0.5), having more middlemen even makes the situation worse for producers as their share of the incremental surplus is negative. The lower left plot in Figure 3.7 makes this point clear. This plot shows the decomposition of producers’ surplus between $S$ and $K$. Note that the surplus is measured without any scaling in this figure. The surplus from $K$ is obviously increasing with more middlemen because they can only sell goods to $N$. However, the surplus from $S$ is decreasing with more middlemen as there are more competition from the middlemen in selling to the buyers. These two effects combined get us the aggregate effect on the producers’ surplus. The lower right plot of Figure 3.7 shows that buyers buy more of the goods from middlemen as there are more middlemen in the system.
3.6 Discussion

Using the above numerical approach, I compute the infinite-period equilibrium for a wide range of parameter settings. This section reports some features of the equilibrium which will help in understanding the analysis in the next section. In the following, two dimensions of the model are varied: 1) the probability $p$ of the Bernoulli
aggregate shock and 2) the total measure $l$ of the middlemen in the system (including the measure of sellers $M$ and measure of buyers $N$ at any particular time).

When $p = 0$ or 1, there is no uncertainty in the system and the producers know exactly where to allocate their goods. When $p = 0.5$, there is pure uncertainty in and the producers have the least information about where to allocate their goods. These two cases represent two extremes of uncertainty in the system. However, as shown above, when $p = 0.5$, all equilibrium quantities including prices, volumes, market tightness and so on do not change as all periods are then identical. In this sense, it shares the same equilibrium characteristics with when $p = 0$ or 1. The only difference is that now there is no role for middlemen in the system. It is interesting to see what happens when $p$ takes a value in between, i.e., when the producers know roughly where to allocate their good and there is some role for the middlemen to help.

### 3.6.1 Slopes of Transition Mappings

The previous section showed that the state of the system can be fully characterized by the quantity $l_M$ and that the dynamics of the equilibrium can be described by the stochastic process of $l_m$ over periods, i.e., two transition mappings of $l_M$ each of which mapping this period’s $l_M$ into the next period’s $l_M$ value. This subsection documents the transition mappings under different parameter settings.

Figure 3.8 shows the fitted slopes of the transition mappings corresponding to probability $p$. First, the slopes are uniformly less than one, indicating that these mappings are contractions. The contraction property is necessary for the existence and uniqueness of the invariant measure of the system. Second, in the right plot of Figure 3.8 the slopes are asymmetric around $p = 0.5$, unlike in other figures. This is because the slopes plotted here describe just one of the two transition mappings

$^\text{10}$Recall that $p$ is the probability that in any period, the aggregate demand is of of type $A$. 

78
in each system. The y-value of each dot is the slope of the mapping that describes the transition rule when the buyers’ preference turn out to be of type $A$ (which happens with probability $p$). Third, the slope monotonically increases to one as more middlemen come into the system. The slopes of the transition mappings control the speed of convergence to the absorbing points (the two boundaries or the intercept of the mapping with the diagonal line). The closer the slope is to one the closer is the next period’s $l_M$ to this period’s $l_M$ value. This means that when there are more and more middlemen in the market, the relative quantity of $l_M$ is less affected by an aggregate shock. In other words, more middlemen make the relative quantity of $l_M$ more stable. Fourth, the slope monotonically increases as the probability of the aggregate shock goes from zero to one. This point illustrates that $l_M$ is less affected by a shock which happens with higher probability, i.e., $l_M$ changes much more upon a shock which happens only with probability $p = 0.1$ (the lower left point on the red line which shows a much lower slope of the transition mapping) than upon a shock which happens with probability $p = 0.9$ (the right most point on the red line which shows a higher slope of the transition mapping).
Figure 3.8: Slope of Transition Mapping \((p)\)

Figure 3.9 shows the fitted slopes of the transition mappings corresponding to probability \(1 - p\). The right plot is indeed symmetric to the right plot of Figure 3.8.

Figure 3.9: Slope of Transition Mapping \((1 - p)\)
3.6.2 Support of the Markov Chain

Given one transition mapping, its intercept with the diagonal line is an absorbing state, i.e., the system will converge to the absorbing state if the system is experiencing an infinite number of shocks of the type that corresponds to that particular transition mapping. Given a pair of transition mappings corresponding to the two realizations of the Bernoulli aggregate shock, the states of the system will jump within the interval bounded by the two intercepts. Figure 3.10 describes the length of such interval. The length of the interval describes how different the system will be upon experiencing an infinite number of consecutive shocks of one type (which is a zero-probability event) versus to experiencing an infinite number of consecutive shocks of the other type. However, the two types of the aggregate shocks can only be distinguished by their probabilities of happening. Therefore, when $p = 0.5$ for the Bernoulli shock (0.5 versus 0.5), it can be expected to observe no difference between the two cases, while when $p = 0.1$ for the Bernoulli shock (0.1 versus 0.9), there can be a big difference. This intuition informs why the length of interval is monotonically increasing as $p$ moves towards zero or one, as shown in the right plot of Figure 3.10. Intuitively, the interval increases as the quantity of middlemen increases, as shown in the left plot.
Figure 3.10: Length of the Support of $l_M$

Figure 3.11 scales the length of such interval by the total quantity of the middlemen $l$. Interestingly, the patterns remain roughly the same but the shape of the lines changes significantly.

Figure 3.11: Length of the Support of $l_M$ (in % of $l$)
3.6.3 $l_M$ in Equilibrium

Figure 3.12 shows that $l_M$ increases as the total quantity $l$ of middlemen increases. However, Figure 3.13 shows the share of $l_M$ relative to $l$ decreases in the process, i.e., there is relatively more $l_N$ as buyers than $l_M$ as sellers in the system. This is due to two features of the model. To understand this, it is useful to look at the following formula stating that in equilibrium the inflow and outflow of $M$ should be roughly the same:

$$\alpha\left(\frac{l_{BM}}{l_M}\right)l_M = \alpha\left(\frac{l_N}{l_K}\right)l_k$$

Though this condition does not hold in exact terms in equilibrium, it is useful to think about it in approximate terms. Consider middlemen $M$ and $N$ in their relative submarket. A useful experiment is to see how the trading volume changes as the number of $M$ or $N$ increases in their submarkets while holding the quantity of their trading counterparties constant. For sellers $M$, a higher $l_M$ means more sellers and thus lower market tightness in the submarket. While more sellers increase the trading volume, the lower market tightness and corresponding lower matching rate decreases the trading volume. The combined effect of both is dictated by the exponential factor of $1 - b$. Suppose $l_M$ increases by a factor of $c$ then the trading volume changes by a factor of $c^{1-b}$.[11] Similar analysis indicates that the trading volume changes by a factor of $c^b$ in $N$’s submarket if $l_N$ increases by the same factor $c$. When $b = 0.5$[12], the two scaling factors equalize and the inflow to $M$ from $N$ and the outflow from $M$ to $N$ are

---

[11]To see this, replace $l_M$ by $cl_M$ on the left-hand side of the equation. Then, we get

$$\alpha\left(\frac{l_{BM}}{cl_M}\right)cl_M = c^{1-b}\alpha\left(\frac{l_{BM}}{l_M}\right)l_M.$$  

[12]Note that $b$ is a parameter in the matching function $\alpha(\cdot)$ that controls the concavity of matching probability. It turns out that $b$ also describes the share of the total surplus buyers get in a successful transaction. Thus, $b = 0.5$ means a buyer and a seller split their total surplus in half and neither party has bargaining power over the other.
the same if the relative shares of $M$ and $N$ remain the same and all other quantities are held constant. This is important because the equilibrium concept requires that the inflows and outflows match roughly.

But the key issue here is that, in equilibrium, the quantities besides $l_M$ and $l_N$ will change accordingly. In $M$’s submarket, upon seeing more sellers $M$ in the market, some buyers $B$ who used to trade directly with producers $S$ are now attracted to trade with $M$ instead. Since there are now more buyers, the trading volume or the outflow of $M$ increases more than that prescribed by the scaling factor $c^{1-b}$. Similarly, in $N$’s submarket, upon seeing more buyers $N$ in the market, the producers $P$ feel much more comfortable being $K$ in making their allocation decisions at the beginning of each period. As there are more sellers in the submarket, the inflow from $N$ to $M$ increases more than that described by the scaling factor $c^b$. However, the buyers $B$ are more responsive than the sellers $P$ in this process because the effect on $P$ should be multiplied by a probability factor $1 - p$. Thus, in order to equate the inflow with outflow, the equilibrium quantity of $l_M$ has to be relatively lower than $l_N$.

![Figure 3.12: Mean $l_M$](image)
Another interesting feature in the above two plots is that $l_M$ decreases in both absolute and relative terms when the aggregate shock $p$ tends to be more extreme. Consider the case when $p = 0.9$. In equilibrium, in most of the periods there is always much more $S$ than $K$. A small $l_K$ means that the middlemen $N$ face few sellers and a big $l_S$ means that the middlemen $M$ face much competition. However, the latter effect is of second order while the first effect is of first order. This means that it is relatively harder for $N$ to buy the good and convert into $M$ than it is for $M$ to sell and convert into $M$. Suppose $l_M$ and $l_N$ are roughly the same. Then one can expect to see more outflow from than inflow to $M$ and thus the $l_M$ will decrease until the system is hit by a “rare” aggregate shock. To equate the inflow and outflow, there needs to be less $l_M$ in equilibrium. This conflict between small inflow and big outflow is eased when there are roughly similar quantities of $S$ and $K$ when $p = 0.5$. Therefore, there are relatively more $M$ than $N$ in when the uncertainty is high.
Figure 3.14: Volatility of $l_M$

Figure 3.14 plots the volatility of $l_M$ in equilibrium. There are two main features in this figure. First, intuitively the volatility of $l_M$ increases with $l$. And the volatility is higher for more extreme uncertainty. As seen from Figure 3.8 and Figure 3.9, the closer $p$ is to the ends, the larger is the difference of slopes between the two transition mapping slopes. This means that there are more large “spikes” in these systems and higher volatility. Second, a bimodal shape emerges in the right plot. This is due to the fact that when $p$ is equal to 0, 1, or 0.5 the equilibrium is stationary and volatility is zero. The properties of the volatility of $l_M$ is important as shows below that in the model, quantities such as spot market prices are approximately a linear transformation of $l_M$ in this model.
3.6.4 $l_S$ in Equilibrium

Figure 3.15 plots $l_S$ in equilibrium. This plot is both in absolute and relative terms as throughout the analysis the total quantity of producers is held constant to one. In the left plot, $l_S$ decreases as middlemen enter the market. Recall that at the beginning of each period, the producers $P$ make a decision on how to allocate their goods in one of two types ($A$ and $B$) facing the aggregate shock. The quantity of $l_S$ depends on the expected payoffs from being $S$ and $K$ after the realization of the aggregate shock. This is rooted in the shape of the matching function $\alpha(\cdot)$. When the number of middlemen is small, the fraction $l_N$ is small as well, and the market tightness in $K$’s submarket is small. When the market tightness ($n$) is small, the slope of the $\alpha(n)$ is high; in other words, the marginal increase in the matching probability is high per unit of increase in the market tightness measure. However, in $S$’s submarket, the competition from $M$ in selling to the buyers $B$ is of a lower order and the slope of the matching function is relatively flat. Therefore, $l_S$ has to decrease
as the number of middlemen increases from zero. But in the far end, $l_S$ might increase with $l$. This is because the effect comes more from the change of quantity of agents than from the matching probabilities. When $l$ is large, it is possible that the market tightness in either submarket is far from zero, thus the sensitivity of the matching rate with respect to the change in the market tightness is quite small. In this case, the change in the relative measure of middlemen takes over. The left plot shows that this quantity effect never beats the matching rate effect. Recall from Figure 3.13 that the share of $l_M$ in $l$ decreases in $l$. From the producers’ perspective, the more $N$ makes the conversion to $K$ more attractive while the less $M$ makes conversion to $S$ more attractive. The former attractiveness beats the latter in this case. Another feature in Figure 3.15 is that the producers convert more to $S$ where the uncertainty decreases.

### 3.6.5 Price Volatility in Spot Market

One focus of this chapter is the spot market price volatility facing the end buyers $B$ and how are they affected by the quantity of middlemen $l$ and uncertainty $p$. The buyers $B$ can buy either from $S$ directly or from $M$ indirectly. To get some intuition let us first take a look at the formula for the spot market price derived earlier:

$$P_{ij} = u_j + r(V^T_{T(j)} - V_j') - \frac{n_{ij}\alpha'(n_{ij})}{\alpha(n_{ij})}Z_{ij}.$$  

Applying this formula to the two submarkets where buyers $B$ trade, we get

$$P_{SB} = u + 0 - b(u) = (1 - b)u,$$

$$P_{MB} = u + 0 - b(u - r\Delta) = (1 - b)u + r\Delta.$$
Note that $\Delta$ above represents the difference between the continuation values of $M$ and $N$ ($\Delta = V_M - V_N$). From above, observe that $P_{SB}$ is a constant, and the total variation of spot price comes from the variation of $P_{MB}$. Since $P_{MB}$ is a linear function of $\Delta$ and since $\Delta$ is almost linear in $l_M$, the stochastic properties of the spot price are fully characterized by the stochastic property of $l_M$; specifically, the spot price volatility is approximately a linear transformation of the volatility of $l_M$. It is helpful to view Figure 3.16 in parallel with Figure 3.14.

Note the left plot has a different nonlinear shape than the left plot in Figure 3.14. This is due to the slope of the linear approximation to $\Delta$, which is variant across different parameter settings. The main idea here is that the volatility of the spot market price may increase with number of middlemen in the market. Intuitively, price volatility is closely related to the mismatch between demand and supply. The two driving forces with the function of middlemen include 1) the ability of the middlemen to meet the random demand by transforming their goods and 2) the fact that the
volatility of $l_M$ itself is a contributor to the random supply shock. The left plot shows that the former force dominates when $l$ is small while the latter force dominates as $l$ increases. In other words, the common wisdom that middlemen may ease the price volatility is only true when the number of middlemen is appropriate. In other words, too many middlemen in the market contributes to supply uncertainty.

3.6.6 Producer Welfare

Another focus of this chapter is the producer surplus. Previously, in the special case when $p = 0.5$, the results suggested that a larger number of middlemen in the market can actually decrease the producers’ surplus. This section extends that result to more general cases. Figure 3.17 shows that producer welfare is decreasing in $l$ when $l$ is small. This pattern is quite similar to the pattern of mean $l_S$ shown in Figure 3.15. Let $S_i$ denote the surplus of type $i$, $V_i$ denote the volume of trading in $i$’s submarket and $P_{ij}$ denote the prevailing price in the submarket where $i$ selling to $j$. Then the total surplus of producers $S_P$ is

$$S_P = S_S + S_K$$

$$= \mathbb{E}[V_S P_{SB}] + \mathbb{E}[V_K P_{KN}]$$

$$= \mathbb{E}[l_S \cdot \alpha(n_{SB}) \cdot P_{SB}] + \mathbb{E}[l_K \cdot \alpha(n_{KN}) \cdot P_{KN}] .$$

The above expectation is taken with respect to the distribution of the invariant measure in equilibrium. From above, observe that the producer surplus is composed of the surplus of being $S$ ($S_S$) and the surplus of being $K$ ($S_K$). Consider the case when $l$ is small and the quantity of middlemen $l$ is increased. The decrease in $S_S$ is much more than the increase in $S_K$. To see this, the three terms in the square brackets above must be compared. 1) Note that from Figure 3.15 we know that $l_S$ decreases

90
and decrease in the same amount as the increase in $l_K$. 2) Moreover, $P_{SB}$ is constant to be 0.5 as shown above and $P_{KN}$ should be smaller than $P_{SB}$ because the total surplus of both parties is lower upon a successful trade. 3) The probability of matching in the submarket $B \rightarrow S$ is much higher than that of submarket $K \rightarrow N$ because the market tightness measure should be very low when the total amount of $l$ is small. Combining the three facts, $S_S$ decreases more than the corresponding increase in $S_K$ as middlemen $l$ increases from a small value. In other words, the producers lose more from the competition of $M$ as sellers when they become $S$ than the gains from having $N$ as buyers when they become $K$. This is the main point that this chapter means to convey: more middlemen may decrease producer surplus in absolute terms. When $l$ is large, the reverse is true.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig3.17.png}
\caption{Total Producers Surplus}
\end{figure}
3.7 Conclusion

This chapter builds a model of commodity middlemen with aggregate shock. The model produces two main results. First, more middlemen in the market may increase the spot price volatility. A potential for normative analysis is based on how governments view the relative importance of the middlemen and the producers industries. It might be possible to find an “optimal” number of middlemen in future work. Second, more middlemen in the market may decrease the producers’ total surplus. Middlemen get a chunk of the pie not only from making the total pie larger but also from appropriating an existing piece from the producers’ hands. This results in strong policy implications as to whether and how to regulate the commodity markets. Suppose the upstream production is endogenized in an extended version of my model, it could possibly show that the producers endogenously shrink productions as more middlemen come into the market.
Appendix A

Appendix for Chapter 2

A.1 Proof of the equilibrium

Proof of Lemma 2. Without loss of generality, we can assume that $X_2(H, X_1, v)$ is first determined and then $I_c(H, X_1, X_2, v)$ is determined. Note that here $I_c$ is considered as a function of $X_2$. By doing this, we can reduce the two-dimensional optimization problem into two one-dimensional optimization problems. Let’s first solve backward for the optimal $I_c(H, X_1, X_2, v)$. Since function $g$ is flat over the range $[0, I_0]$, it suffices to compare the change of the manipulator’s profit from choosing $I_c = 0$ to some $I_c = a \in (I_0, I_q]$. Let this change be

$$\Delta \pi(a) = \pi(X_1, X_2, a, P_1, P_2) - \pi(X_1, X_2, 0, P_1, P_2)$$

It can be computed that

$$\Delta \pi(a) = \begin{cases} 
-g(a) \cdot X_1 & \text{if } X_1 + X_2 < 0 \\
-g(a) \cdot (X_1 + X_2) & \text{if } 0 \leq X_1 + X_2 \leq z \\
-g(a) \cdot (X_1 + X_2) - (X_1 + X_2 - z) \cdot d & \text{if } X_1 + X_2 > z 
\end{cases}$$

Since $g$ is monotonically increasing in $a$, we can compute the optimal $I_c(H, X_1, X_2, v)$:

$$I^*_c(H, X_1, X_2, v) = \begin{cases} 
I_q & \text{if } X_1 + X_2 < 0 \text{ and } X_1 < 0 \\
0 & \text{all other cases}
\end{cases}$$
Plug $I^*_c$ into the manipulator’s profit function we get

$$\pi(X_1, X_2, I^*_c, P_1, P_2) = \begin{cases} 
X_1 \cdot (v - P_1) & \text{if } X_1 + X_2 < 0 \text{ and } X_1 > 0 \\
X_1 \cdot (v - P_1) - X_1 \cdot g(I_q) & \text{if } X_1 + X_2 < 0 \text{ and } X_1 < 0 \\
X_1 \cdot (v - P_1) & \text{if } 0 \leq X_1 + X_2 \leq z \\
X_1 \cdot (v - P_1) + (X_1 - z) \cdot d & \text{if } X_1 + X_2 > z 
\end{cases}$$

From above we can see that for any $X_1$ the strategy $X^*_2 = 0$ is always optimal. Furthermore, since this optimal $X^*_2$ does not depend on the value of $X_1$, our assumption that $X_2$ is a function of $X_1$ does not complicate the problem at all. Thus we have solved backwardly for $X^*_2(H, X_1, v) = 0$ and $I^*_c(H, X_1, v) = \max(X_1, 0) \cdot I_q$. It can be easily checked forwardly as well that this is true.

\[ \square \]

**Proof of Lemma** Let $S_i = \sup_x \Pi_P(x, H_i)$. Suppose (to be contradicted) that in equilibrium $\mathbb{P}(\Pi_P(X_1(H_i), H_i) = S_i) < 1$ for some $i$. We prove in two steps:

**Step 1**: Show that $\mathbb{P}(\Pi_P(X_1(H_i), H_i) < S_i - \frac{1}{m}) > \frac{1}{2m}$ for $m$ large enough. To see this, let $E_n = \{\Pi_P(X_1(H_i), H_i) < S_i - \frac{1}{n}\}$. If $\mathbb{P}(E_n) \leq \frac{1}{2m}$ for all $n$, then $\sum_{n=1}^{\infty} \mathbb{P}(E_n) \leq \sum_{n=1}^{\infty} \frac{1}{2m} < \infty$. By Borel-Cantelli lemma we have $\mathbb{P}(\lim \sup_{n \to \infty} E_n) = 0$. In other words, the probability that infinitely many of $E_n$’s occur is zero, or in formula $\mathbb{P}(\exists n_0 \text{ such that } \Pi_P(X_1(H_i), H_i) \geq S_i - \frac{1}{n} \text{ for all } n \geq n_0) = 1$. By definition of $S_i$, this means $\mathbb{P}(\Pi_P(X_1(H_i), H_i) = S_i) = 1$, contradicting to the supposition that $\mathbb{P}(\Pi_P(X_1(H_i), H_i) = S_i) < 1$.

**Step 2**: Show that there exists a dominating strategy $X^*_1$. With $m$ found in step 1, let $A = \{\Pi_P(X_1(H_i), H_i) < S_i - \frac{1}{m}\}$. By definition of $S_i$, there exists an $x_0$ such that
\( \Pi_P(x_0, H_i) > S_i - \frac{1}{2m} \). Let \( X'_1(H_i) = x_0 \cdot 1_A + X_1(H_i) \cdot 1_{A^c} \). Then we have

\[
\mathbb{E}_{X_1,v}[\pi(X'_1, P_1)|H_i] = \mathbb{E}_{X'_1}[\Pi_P(X'_1(H_i), H_i)]
\]

\[
= \mathbb{E}_{X'_1}[\Pi_P(X'_1(H_i), H_i)|A] \cdot \mathbb{P}(A) + \mathbb{E}_{X'_1}[\Pi_P(X'_1(H_i), H_i)|A^c] \cdot \mathbb{P}(A^c)
\]

\[
= \mathbb{E}_{X'_1}[\Pi_P(x_0, H_i)|A] \cdot \mathbb{P}(A) + \mathbb{E}_{X'_1}[\Pi_P(x_1(H_i), H_i)|A^c] \cdot \mathbb{P}(A^c)
\]

\[
> (S_i - \frac{1}{2m}) \cdot \mathbb{P}(A) + \mathbb{E}_{X'_1}[\Pi_P(x_1(H_i), H_i)|A^c] \cdot \mathbb{P}(A^c)
\]

\[
\geq \frac{\mathbb{P}(A)}{2m} + \mathbb{E}_{X'_1}[\Pi_P(x_1(H_i), H_i)] = \frac{\mathbb{P}(A)}{2m} + \mathbb{E}_{X_1,v}[\pi(X_1, P_1)|H_i]
\]

This makes \( X_1 \) impossible to be an equilibrium strategy. Therefore, in equilibrium we have \( \mathbb{P}(\Pi_P(X_1(H), H) = \sup_x \Pi_P(x, H)) = 1 \). Note that since the supremum is obtainable, maximum and supremum can be used interchangeably.

\[
\text{Proof of Lemma 3.} \quad \Pi^*_p(H_i) = \sup_x \Pi_P(x, H_i) \geq \Pi_P(0, H_i) = 0. \text{ Suppose (to be contradicted) that } \mathbb{E}_{H,X_1,v}[\pi(X_1, P_1)] = 0. \text{ By lemma 2}
\]

\[
\mathbb{E}_{H,X_1,v}[\pi(X_1, P_1)] = \mathbb{E}_{H,X_1}[\Pi_P(X_1(H), H)] = \mathbb{E}_H[\Pi^*_p(H)]
\]

\[
= \Pi^*_p(H_0) \cdot \mathbb{P}(H_0) + \Pi^*_p(H_1) \cdot \mathbb{P}(H_1)
\]

Since both \( \Pi^*_p(H_0) \) and \( \Pi^*_p(H_1) \) are nonnegative as is shown above, this requires that \( \Pi^*_p(H_0) = \Pi^*_p(H_1) = 0 \). By definition of \( \Pi^*_p(\cdot) \), this further means that \( \Pi_P(x, H_i) \leq 0 \) for all \( x \) and \( i = 0, 1 \). Let \( x_1 = \epsilon \) for some \( \epsilon \in (0, \min(H_1 - H_0, z)) \) and \( x_2 = \epsilon - (H_1 - H_0) < 0 \). The following two constraints need to be satisfied:

\[
\Pi_P(x_1, H_0) = x_1 \cdot \mathbb{E}[v] - P_1(x_1 + H_0) \leq 0 \quad \Rightarrow P_1(\epsilon + H_0) \geq \mathbb{E}[v]
\]

\[
\Pi_P(x_2, H_1) = x_2 \cdot [\mathbb{E}[v] - g(I_q) - P_1(x_2 + H_1)] \leq 0 \quad \Rightarrow P_1(\epsilon + H_0) \leq \mathbb{E}[v] - g(I_q)
\]
Since the above two inequalities cannot be satisfied simultaneously, we have in equilibrium $E_{H,X_1,v}[\pi(X_1, P_1)] > 0$ and thus at least one of $\Pi_p(H_0)$ or $\Pi_p(H_1)$ is strictly greater than zero.

\[\pi(X_1, P_1) > 0\] and thus at least one of $\Pi_p(H_0)$ or $\Pi_p(H_1)$ is strictly greater than zero. We prove in four steps:

**Step 1:** Show that $Y_0 \cap Y_1 \neq \emptyset$. By lemma 2 we have $Y_i \neq \emptyset$ for $i = 0, 1$. Suppose that $Y_0 \cap Y_1 = \emptyset$. Then on observing any $y \in Y_0 \cup Y_1$, speculators can identify the state $H$ and hence set a futures pricing rendering the manipulator unprofitable. This contradicts to lemma 3 that the manipulator gains profits in at least one state.

**Step 2:** Show that $P(X_1(H_0) + H_0 \in Y_0 \cap Y_1) = 1$. By lemma 2 and the definition of $Y$, we have $P(X_1(H_0) + H_0 \in Y_0) = 1$. Then it suffices to show that $P(X_1(H_0) + H_0 \in Y_0 \setminus Y_1) = 0$. Suppose (to be contradicted) that $P(X_1(H_0) + H_0 \in Y_0 \setminus Y_1) > 0$. On observing any $y \in Y_0 \setminus Y_1$, speculators infer that $H = H_0$ and set futures price rendering manipulator unprofitable in the sense that $\Pi_p(y - H_0, H_0) \leq 0$. Combined with lemma 2 this contradicts the supposition that $\Pi_p(H_0) > 0$. Therefore, we have $P(X_1(H_0) + H_0 \in Y_0 \setminus Y_1) = 0$ and thus $P(X_1(H_0) + H_0 \in Y_0 \cap Y_1) = 1$.

**Step 3:** Show that $Y_0 \cap Y_1$ is a set of isolated points. Suppose (to be contradicted) that there exists some connected set $(a, b) \subseteq Y_0 \cap Y_1$. For any $y \in (a, b)$ the following two equations hold:

\[
\Pi_p(H_0) = \Pi_p(y - H_0, H_0)
\]
\[
\Pi_p(H_1) = \Pi_p(y - H_1, H_1)
\]

We can substitute out $P_1(y)$ in the above two equations and connect $\Pi_p(H_0)$ and $\Pi_p(H_1)$ by some implicit function of $y$. Specifically, let $f(y, h) = \max(y - h - z, 0)$.
\[ d - \min(y - h, 0) \cdot g(I_q). \] Then we have

\[ \Pi_P(H_0) = \frac{y - H_0}{y - H_1} \cdot [\Pi_P^*(H_1) - f(y, H_0)] + f(y, H_1) \]

Now if we pick \( y_1, y_2 \in (a, b) \), \( \Pi_P^*(H_0) \) and \( \Pi_P^*(H_1) \) can then be solved in terms of \( y_1 \) and \( y_2 \) using the above equation. Since we have the flexibility in picking \( y_1 \) and \( y_2 \), this makes a contradiction. Therefore \( \mathcal{Y}_0 \cap \mathcal{Y}_1 \) contains no connected set and thus is a set of isolated points.

**Step 4:** Note that the above claims also hold if we had supposed \( \Pi_P^*(H_1) > 0 \). Therefore, we have proved the latter part of the lemma. Now pick any \( m \in \mathcal{Y}_0 \cap \mathcal{Y}_1 \) such that \( \mathbb{P}(X_1(H_0) + H_0 = m) > 0 \). Since we supposed \( \Pi_P^*(H_0) > 0 \), the speculators cannot infer the state \( H \) from observing the net quantity \( X_1(H_0) + H_0 \) (for otherwise they will set a futures price rendering the manipulator unprofitable and thus contradicting the supposition that \( \Pi_P^*(H_0) > 0 \)). Therefore, it needs to be the case that \( \mathbb{P}(X_1(H_1) + H_1 = m) > 0 \) for this picked \( m \). Hence the first part of the lemma is proved.

**Proof of Lemma 5.**

**Part a:** On observing the net quantity \( m + H_0 \), speculators infer that \( X_1(H) < 0 \) no matter whether \( H = H_0 \) or \( H = H_1 \), and thus they set price \( P_1(m + H_0) = \mathbb{E}[v] - g(I_q) \). Plug this price into squeezer’s profit function we have \( \Pi_P^*(H_0) = 0 \).

**Part b:** Suppose (to be contradicted) that \( \Pi_P^*(H_0) = 0 \). Then we must have \( \Pi_P(x, H_0) \leq 0 \) for all \( x \). Specifically, we have

\[ \Pi_P(x, H_0) = x \cdot (\mathbb{E}[v] - P_1(x + H_0)) + (x - z) \cdot d \leq 0 \]

Pick a small positive \( \epsilon_1 \) such that \( m - \epsilon_1 > z \). Then for any \( x \in (m - \epsilon_1, m + \epsilon_1) \) the
futures price function $P_1$ should satisfy the following condition:

$$P_1(x + H_0) \geq \mathbb{E}[v] + \frac{(x - z) \cdot d}{x}$$

Note that $P(X_1(H_1) = m - \Delta H) > 0$ (since for otherwise on observing the net quantity $m + H_0$ the speculators can infer that $H = H_0$ and set the futures price to be $P_1(m + H_0) = \mathbb{E}[v] + d$, which makes $X_1(H_0) = m$ a losing choice). For $X_1(H_1) = m + H_0 - H_1$ to be optimal, we need $\Pi_P(x, H_1) \leq \Pi_P(m - \Delta H, H_1)$.

Specifically, we have

$$\Pi_P(x, H_1) = x \cdot (\mathbb{E}[v] - g(I_q) - P_1(x + H_1)) \leq \Pi_P(m - \Delta H, H_1)$$

Pick a small positive $\epsilon_2$ such that $m - \Delta H + \epsilon_2 < 0$. Then for any $x \in (m - \Delta H - \epsilon_2, m - \Delta H + \epsilon_2)$ the futures price function $P_1$ should satisfy the following condition:

$$P_1(x + H_1) \leq \mathbb{E}[v] - g(I_q) - \frac{\Pi_P(m - \Delta H, H_1)}{x}$$

Next we show that the above two conditions cannot be satisfied simultaneously. Rewrite the two conditions in terms of $y = x + H$ and let $\epsilon = \min(\epsilon_1, \epsilon_2)$, we need to check if for all $y \in (m + H_0 - \epsilon, m + H_0 + \epsilon)$

$$\mathbb{E}[v] + \frac{(y - H_0 - z) \cdot d}{y - H_0} \leq P_1(y) \leq \mathbb{E}[v] - g(I_q) - \frac{\Pi_P(m - \Delta H, H_1)}{y - H_1}$$

Define $f(y) = -g(I_q) - \frac{\Pi_P(m - \Delta H, H_1)}{y - H_1} - \frac{(y - H_0 - z) \cdot d}{y - H_0}$. It suffices to check if $f(y) \geq 0$ for
all \( y \in (m + H_0 - \epsilon, m + H_0 + \epsilon) \). We have

\[
 f'(y) = \frac{\Pi_P(m - \Delta H, H_1)}{(y - H_1)^2} - \frac{z \cdot d}{(y - H_0)^2} \\
 f''(y) = -\frac{2\Pi_P(m - \Delta H, H_1)}{(y - H_1)^3} + \frac{2z \cdot d}{(y - H_0)^3} > 0
\]

\[
\Pi_P(m - \Delta H, H_1) = (g(I_q) + \frac{m - z}{m} \cdot d) \cdot (H_1 - H_0 - m) > 0
\]

Since \( f(m + H_0) = 0 \), \( f(y) \) is continuous at \( y = m + H_0 \) and \( f''(y) > 0 \), it suffices to check if \( f'(m + H_0) = 0 \). Solve for the zero point of \( f'(m + H_0) = 0 \) we get \( m^* = \sqrt{z_0 \Delta H/d + g} \). By our assumption that \( m \neq \sqrt{z_0 \Delta H/d + g} \), we have \( f'(m + H_0) \neq 0 \) and thus, we can always find some \( y_0 \in (m + H_0 - \epsilon, m + H_0 + \epsilon) \) such that \( f(y_0) < 0 \), which further means that the two conditions on the futures price function \( P_1 \) cannot hold simultaneously. Therefore we must have \( \Pi_P^*(H_0) > 0 \) if \( \mathbb{P}(X_1(H_0) = m) > 0 \) for some \( m \in (z, \Delta H) \setminus \{ \sqrt{z_0 \Delta H/d + g} \} \).

**Part c:** We split the proof into three cases: \( [\Delta H, +\infty) = \{ \Delta H \} \cup (\Delta H, \Delta H + z] \cup (\Delta H + z, +\infty) \). Now we prove case by case.

**Case 1:** Suppose (to be contradicted) that \( \mathbb{P}(X_1(H_0) = \Delta H) > 0 \). Then it must be that \( \mathbb{P}(X_1(H_1) = 0) > 0 \) (for otherwise the speculators would know that there will be a squeeze, set the futures price \( P_1(H_1) = \mathbb{E}[v] + d \), and the squeezer loses).

From this we know that \( \Pi_P^*(H_1) = 0 \), i.e. the squeezer has to have zero profit when \( H = H_1 \). This requires that \( P_1(y) = \mathbb{E}[v] - g(I_q) \) for \( y < H_1 \). Also from \( \mathbb{P}(X_1(H_1) = 0) > 0 \) we have \( P_1(H_1) \in (\mathbb{E}[v], \mathbb{E}[v] + d) \). Consider the strategy \( X'_1 \) defined by \( \mathbb{P}(X'_1(H_0) = \Delta H - \epsilon) = 1 \) and \( \mathbb{P}(X'_1(H_1) = 0) = 1 \) for some \( \epsilon \) small.

Since the price function \( P_1 \) is left-discontinuous at \( H_1 \) as is shown above, strategy \( X'_1 \) dominates strategy \( X_1 \), i.e. any strategy \( X_1 \) with \( \mathbb{P}(X_1(H_0) = H_1 - H_0) > 0 \) cannot be an equilibrium strategy. In other words, in equilibrium \( \mathbb{P}(X_1(H_0) = \Delta H) = 0 \).

**Case 2:** Suppose (to be contradicted) that \( \mathbb{P}(X_1(H_0) = m) > 0 \) for some \( m \in
Proof of Lemma 6. Let’s prove the lemma under the first condition, the proof under the other two conditions follows immediately. Suppose (to be contradicted) that $X_m > z$ for all $(\triangle H, \triangle H + z]$. Then $\mathbb{P}(X_1(H) + H_1 = m + H_0) > 0$ (since for otherwise the speculators can infer from observing the net quantity $m + H_0$ that $H = H_0$ and hence set price $P_1(m + H_0) = \mathbb{E}[v] + d$, making $\Pi^*_P(H_0) = -z \cdot d < 0$). Since the speculators cannot tell the state $H$ from observing the net quantity $m + H_0$, they will set a price $P_1(m + H_0) \in (\mathbb{E}[v], \mathbb{E}[v] + d)$. But such a price would make $\Pi^*_P(H_1) = (m + H_0 - H_1) \cdot (\mathbb{E}[v] - P_1(m + H_0)) < 0$. Therefore $\mathbb{P}(X_1(H_0) = m) = 0$ for all $m \in (\triangle H, \triangle H + z]$.

Case 3: Suppose (to be contradicted) that $\mathbb{P}(X_1(H_0) = m) > 0$ for some $m \in (\triangle H + z, +\infty)$. On observing the net quantity $m + H_0$ the speculators can infer that $X_1(H) > z$ regardless of $H$ and set the futures price to be $P_1(m + H_0) = \mathbb{E}[v] + d$. This price renders the squeezer unprofitable as $\Pi^*_P(H_0) = -z \cdot d < 0$. Hence $\mathbb{P}(X_1(H_0) = m) = 0$ for all $m \in (\triangle H + z, +\infty)$. 

Proof of Lemma 6. Let’s prove the lemma under the first condition, the proof under the other two conditions follows immediately. Suppose (to be contradicted) that $\Pi^*_P(H_0) = 0$. By lemma 3 we have $\Pi^*_P(H_1) > 0$. By lemma 4 $\mathbb{P}(X_1(H_1) + H_1 = m) > 0$ for some $m$. Following we show that such an $m$ doesn’t exist.

1. Suppose $m \in (-\infty, H_0)$. Note that the supposition $\Pi^*_P(H_0) = 0$ requires $P_1(y) = \mathbb{E}[v] - g(I_0)$ for $y < H_0$, for otherwise $\Pi_P(x, H_0) > 0$ for some $x < 0$. Under such $P_1$, $\Pi_P(m - H_1, H_1) = 0$, contradicting to $\Pi^*_P(H_1) > 0$.

2. Suppose $m \in [H_0, H_0 + z]$. Then we have $P_1(m) \in (\mathbb{E}[v] - g(I_0), \mathbb{E}[v])$. But note that the supposition $\Pi^*_P(H_0) = 0$ requires $P_1(y) \geq \mathbb{E}[v]$ for $y \in [H_0, H_0 + z]$, for otherwise $\Pi_P(x, H_0) > 0$ for some $x < 0$.

3. Suppose $m \in (H_0 + z, H_1)$. It has to be that $\mathbb{P}(X_1(H_0) + H_0 = m) > 0$. By part b of lemma 5 we have $\Pi^*_P(H_0) > 0$, contradicting to the supposition $\Pi^*_P(H_0) = 0$. Thus $m \notin (H_0 + z, H_1)$.
4. Suppose $m \in [H_1, +\infty)$. Obviously, $\Pi_P(m - H_1, H_1) \leq 0$ since the speculators can always infer the state $H = H_1$.

The above arguments show that $\Pi_P(H_0) > 0$. By lemma 4, $X_1(H_0)$ is supported on a set of isolated points. For any $m$ such that $\mathbb{P}(X_1(H_0) = m) > 0$, by part a and c of lemma 5, $m \notin (-\infty, 0) \cup [\Delta H, +\infty)$. Also $m \neq 0$ since $\Pi_P(0, H_0) = 0$ contradicts to $\Pi_P(H_0) > 0$. Therefore we have $m \in (0, \Delta H)$, i.e. $\mathbb{P}(X_1(H_0) + H_0 \in (H_0, H_1)) = 1$. Now pick any $m$ such that $\mathbb{P}(X_1(H_0) = m) > 0$, for $\Pi_P(m, H_0) > 0$ it has to be the case that $\mathbb{P}(X_1(H_1) + H_1 = m + H_0) > 0$ and thus $P_1(m + H_0) \in (\mathbb{E}[v] - g(I_q), \mathbb{E}[v] + d)$. Since $m \in (0, \Delta H)$, we have $X_1(H_1) = m - \Delta H < 0$ and $\Pi_P(H_1) = \Pi_P(m - \Delta H, H_1) > 0$. Since for each point $m$ in the support of $X_1(H_1)$ we must have $\mathbb{P}(X_1(H_0) + H_0 = m + H_1) > 0$ and that $\mathbb{P}(X_1(H_0) + H_0 \in (H_0, H_1)) = 1$, we must have $\mathbb{P}(X_1(H_1) + H_1 \in (H_0, H_1)) = 1$.

**Proof of Lemma 7.** Since $\Pi_P(H_0) = 0$, by lemma 6 it has to be the case that $m^* \in (z, \Delta H)$ (i.e. $\frac{z}{\Delta H} < \frac{1}{1 + \alpha}$) and $\mathbb{P}(X_1(H_0) = m^*) > 0$ (for otherwise, one of the three conditions in lemma 6 is satisfied and thus $\Pi_P(H_0) > 0$). Since $\Pi_P(H_0) = \Pi_P(m^*, H_0) = m^*(\mathbb{E}[v] - P_1(m^* + H_0)) + (m^* - z)d = 0$, we have $P_1(m^* + H_0) = \mathbb{E}[v] + \frac{m^* - z}{m^*}d > \mathbb{E}[v]$. This means $\mathbb{P}(X_1(H_1) + H_1 = m^* + H_0) > 0$ and thus $\Pi_P(H_1) = (m^* - \Delta H)(\mathbb{E}[v] - g - P_1(m^* + H_0)) = (\Delta H - m^*)(g + \frac{m^* - z}{m^*}d) > 0$. By the optimality condition, $\Pi_P(x, H_0) \leq 0$ for all $x$, which means

$$P_1(y) = \begin{cases} \mathbb{E}[v] - g(I_q) & \text{if } y < H_0 \\ \geq \mathbb{E}[v] & \text{if } H_0 < y \leq H_0 + z \\ \geq \mathbb{E}[v] - \frac{zd}{y - H_0} + d & \text{if } y > H_0 + z \end{cases}$$

We next show that if in equilibrium $P_1$ satisfies above constraints, then $\mathbb{P}(X_1(H_1) = m^* - \Delta H) = 1$. Since (by lemma 4) $X_1(H_1)$ is supported on a set of isolated points,
it suffices to show that $P(X_1(H_1) = x) = 0$ for all $x \neq m^* - \Delta H$. We show case by case:

1. $x \notin (-\infty, -\Delta H)$ since for otherwise we would have $P(x, H_0) = 0$, contradicting to the fact that $P^*_p(H_1) > 0$.

2. $x \neq -\Delta H$. If $P(X_1(H_1) = -\Delta H) > 0$. By market efficiency condition, $P_1(H_0) < E[v]$. But the above constraint on $P_1$ requires that $P_1(y) \geq E[v]$ for $y > H_0$. Thus, $P_1$ is right-discontinuous at $H_0$. The optimality condition requires that $P(-\Delta H, H_1) = P^*_p(H_1)$. But the right-discontinuity of $P_1$ at $H_0$ makes it possible to achieve a strictly higher profit: $P(-\Delta H + \epsilon, H_1) > P^*_p(H_1)$ for $\epsilon$ small. Indeed, this logic provides us with a constraint on the existence of type 0 equilibrium. The requirement that $P(-\Delta H, H_1) \leq P^*_p(H_1)$ gives us

$$\frac{z}{\Delta H} \leq 1 + 2a - 2\sqrt{a(1+a)}$$

It can be shown that the above upper bound is smaller than $\frac{1}{1+a}$.

3. $x \notin (-\Delta H, -\Delta H + z]$. The argument is basically the same as above, i.e. the optimality condition requires that $P_1(x + H_1) \geq E[v]$, which is inconsistent with the market efficiency condition.

4. $x \notin (-\Delta H + z, 0) \setminus \{m^* - \Delta H\}$. Let $B(H_0) = \{(y, p) | (y-H_0)(E[v] - p) + (y-H_0 - z)d = 0 \text{ and } H_0 + z < y < H_1 \}$ and $B(H_1) = \{(y, p) | (y-H_1)(E[v] - p - g(I_0)) = P^*_p(H_1) \text{ and } H_0 + z < y < H_1 \}$. These are the iso-profit contours in the space of $(y, p)$. Since these two lines intersect at only one point $(m^* + H_0, P_1(m^* + H_0))$, for any point $x$ in this region the price $P_1(x + H_1)$ making $P^*_p(x, H_1) = P^*_p(H_1)$ will make $P(x + \Delta H, H_0) < 0$. But then $P(X_1(H_0) = x + \Delta H) = 0$. The market efficiency condition should be violated.

102
5. \( x \notin [0, +\infty) \). By part c of lemma\[5\] we have \( \mathbb{P}(X_1(H_0) = x + H_1) = 0 \). Therefore, for any \( x \) in this region, the speculators can refer that the state \( H = H_1 \) on observing the net quantity \( x + H_1 \), which makes it impossible to generate a positive profit for the manipulator.

Thus, we have shown that if \( \mathbb{P}(X_1(H_0) = m^*) > 0 \) and \( \Pi_p^*(H_0) = 0 \) then \( \mathbb{P}(X_1(H_1) = m^* - \Delta H) = 1 \) and \( \Pi_p^*(H_1) > 0 \). Let \( \mu = \mathbb{P}(X_1(H_0) = m^*) \), then by the market efficiency condition

\[
P_1(m^* + H_0) = \frac{\lambda \mu}{\lambda \mu + 1 - \lambda} (\mathbb{E}[v] + d) + \frac{1 - \lambda}{\lambda \mu + 1 - \lambda} (\mathbb{E}[v] - g)
\]

Solve for \( \mu \) we get

\[
\mu = \frac{1 - \lambda}{\lambda} \left( \frac{(1 + a) m^*}{z} - 1 \right)
\]

To ensure \( \mu \in (0, 1] \) we need

\[
\frac{z}{\Delta H} \geq (1 + a)(1 - \lambda)^2
\]

Since from above \( \frac{z}{\Delta H} \leq 1 + 2a - 2\sqrt{a(1 + a)} \) this requires

\[
\lambda \geq \sqrt{\frac{a}{1 + a}}
\]

Proof of Lemma\[8\]. Define the profit function

\[
\Pi(x, p) = x \cdot [\mathbb{E}[v] - p] + \max(x - z, 0) \cdot d - \min(x, 0) \cdot g(I_q)
\]
Define the boundary sets $B(h, \pi)$ and interior sets $I(h, \pi)$ of iso-profit levels:

$$B(h, \pi) = \{(y, p) \mid H_0 < y < H_1 \text{ and } \Pi(y - h, p) = \pi\}$$

$$I(h, \pi) = \{(y, p) \mid H_0 < y < H_1 \text{ and } \Pi(y - h, p) > \pi\}$$

For equilibrium profit levels $\Pi^*_P(H_i)$ define the sets $B(H_i) = B(H_i, \Pi^*_P(H_i))$ and $I(H_i) = I(H_i, \Pi^*_P(H_i))$ for $i = 0, 1$. Note that the set $I(H_0)$ has the property that if $(y, p) \in I(H_0)$ then $(y, p') \in I(H_0)$ for all $p' < p$. Similarly, the set $I(H_1)$ has the property that if $(y, p) \in I(H_1)$ then $(y, p') \in I(H_1)$ for all $p' > p$. We first show that in equilibrium $I(H_0) \cap I(H_1) = \emptyset$. Suppose there exists $(y, p) \in I(H_0) \cap I(H_1)$. Then no matter how $P_1(y)$ is defined $P_1(y) \in I(H_0) \cup I(H_1)$. Thus, we can always find a strictly dominating strategy. To see this, without loss of generality let $P_1(y) \in I(H_0)$, then the strategy $X'_1(H_0)$ defined by $P(X'_1(H_0) + H, P_1(y) + H) = 1$ generates a strictly higher profit $\Pi(y - H_0, H_0) > \Pi^*_P(H_0)$.

Note that 1) $I(H_1)$ is convex and 2) $I(H_0)$ is convex on $y \in (H_0, H_0 + z)$ and $H_0 + z, H_1$, we have that $B(H_0)$ and $B(H_0)$ intersects at most two points. Next, we show that given the conditions in this lemma,

$$\mathbb{P}(\langle X_1(H) + H, P_1(X_1(H) + H) \rangle \in B(H_0) \cap B(H_1)) = 1$$

Since $\Pi^*_P(H_i) > 0$, by lemma it suffices to show that for all $y$ such that $\mathbb{P}(X_1(H) + H = y) > 0$, $(y, P_1(y)) \in B(H_0) \cap B(H_1)$. Again, since $\Pi^*_P(H_i) > 0$, to induce a positive profit, the speculators cannot infer the state $H$ on observing the net quantity $y$, i.e. $\mathbb{P}(X_1(H_0) + H_0 = y) > 0$ and $\mathbb{P}(X_1(H_1) + H_1 = y) > 0$. Clearly, by optimality condition of the equilibrium, we have $\Pi(y - H, P_1(y)) = \Pi^*_P(H)$. Therefore $(y, P_1(y))$ belongs to both $B(H_0)$ and $B(H_1)$. Since $B(H_0) \cap B(H_1)$ contains at most two points, lemma proved. \qed
Proof of Theorem 9. Since the conditions in lemma 6 and lemma 7 provide a complete partition of the possible cases under equilibrium, next we follow these cases. In short, lemma 7 characterizes the equilibrium (Type 0) when $\Pi_p(H_0) = 0$ and lemma 8 characterizes the equilibrium when $\Pi_p(H_0) > 0$. By lemma 8 when $\Pi_p(H_0) > 0$, there are three possible cases depending on the number and location of the intersection(s) between $B(H_0)$ and $B(H_1)$:

1. Type 1: They intersect solely at some point with $y^* \in (H_0 + z, H_1)$.
2. Type 2: They intersect solely at some point with $y^* \in (H_0, H_0 + z)$.
3. Type 3: They intersect at two points in both ranges as above. We will prove that this case cannot be an equilibrium.

Next we first characterize the equilibrium strategies and the futures prices on equilibrium paths. Then we fill in the price function values on off-equilibrium paths and check that both the optimality condition and the market efficiency condition hold.

Type 0. See lemma 7 for details. To show the existence of an equilibrium we still need to fully describe $X_1(H_0)$ and the price function such that they satisfy the two equilibrium conditions (optimality and market efficiency). Consider the following strategy and the price function:

$$X_1(H_0) = \begin{cases} 
    m^* \text{ with prob } \mu \\
    0 \text{ with prob } 1 - \mu
\end{cases}$$

$$P_1(y) = \begin{cases} 
    \mathbb{E}[v] - g(I_q) \quad \text{if } y < H_0 \\
    \mathbb{E}[v] \quad \text{if } H_0 \leq y \leq H_0 + z \\
    \mathbb{E}[v] + d - \frac{zd}{y - H_0} \quad \text{if } H_0 + z < y < H_1 \\
    \mathbb{E}[v] + d \quad \text{if } y \geq H_1
\end{cases}$$
**Type 1.** This is the case when $B(H_1)$ and $B(H_0)$ uniquely intersect at some point with $y^* \in (H_0 + z, H_1)$. First find locus of the contact curve between $B(H_1)$ and $B(H_0)$ for $y \in (H_0 + z, H_1)$. Let $(y^*, p^*)$ be a point on this curve, then

$$(y^*, p^*) \in \arg\max_{(y, p)} \{ (y - H_1)(\mathbb{E}[v] - g - p) : (y - H_0)(\mathbb{E}[v] - p) + (y - H_0 - z)d \geq \Pi_p^*(H_0) \}$$

Compute the Lagrangian we get

$$L = (y - H_1)(\mathbb{E}[v] - g - p) - \alpha[(y - H_0)(\mathbb{E}[v] - p) + (y - H_0 - z)d - \Pi_p^*(H_0)]$$

From the first order conditions we get the contact curve (call it L-II):

$$\Delta H(\mathbb{E}[v] - p) + dH_1 + gH_0 = (d + g)y$$

In this case $p^* = \mathbb{E}[v] - (1 - \lambda)g + \lambda d$. Plug in to the above equation we get $y^* = (1 - \lambda)H_1 + \lambda H_0$. Since $y^* \in (H_0 + z, H_1)$, we need to have

$$\frac{z}{\Delta H} < 1 - \lambda$$

With $(y^*, p^*)$ we can compute the equilibrium profits:

$$\Pi_p^*(H_1) = \lambda^2 (d + g) \Delta H$$
$$\Pi_p^*(H_0) = (1 - \lambda)^2 (d + g) \Delta H - zd$$

To make sure that both $\Pi_p^*(H_0) > 0$ and $\Pi_p^*(H_1) > 0$, we need

$$\frac{z}{\Delta H} < (1 + a)(1 - \lambda)^2$$
Finally, we need to find the condition under which \((y^*, p^*)\) is the unique intersection of \(B(H_1)\) and \(B(H_0)\). We split the discussion into three ranges of values that \(\Pi^*_p(H_0)\) could possibly take defined by 
\[
\pi_1 = ((d + g)\frac{z}{\Delta H} - d)z, \quad \pi_2 = \frac{gz^2}{\Delta H}, \quad \pi_3 = gz.
\]
To understand these values, note that \(B(H_0, \pi_1)\) is the iso-profit level that crosses the intersection between \(L-II\) and \(y = H_0 + z\), \(B(H_0, \pi_2)\) is the iso-profit level that crosses the intersection between \(L-I\) and \(y = H_0 + z\), and \(B(H_0, \pi_3)\) is the iso-profit level that crosses the point \((H_0 + z, E[v] - g)\). Note that the condition \(\frac{z}{\Delta H} < 1 - \lambda\) requires that \(\Pi^*_p(H_0) > \pi_1\). Next we discuss the cases on the value of \(\Pi^*_p(H_0)\):

1. If \(\pi_1 < \Pi^*_p(H_0) \leq \pi_2\). Let \((y', p')\) be the intersection between \(B(H_1)\) and \(L-I\). Let \(\pi'\) be the profit level such that \(B(H_0, \pi')\) is tangent to \(B(H_1)\) at \((y', p')\). We claim that \((y^*, p^*)\) is the unique intersection of \(B(H_1)\) and \(B(H_0)\) iff \(\pi' < \Pi^*_p(H_0)\). Given any point \((y, p)\) on \(L-II\), it can be computed that

\[
\pi' = \frac{\Delta H}{g} \left( g + \frac{y - H_1}{\Delta H} \sqrt{g(d + g)} \right)^2
\]

\[
\Pi^*_p(H_0) = \frac{d + g}{\Delta H} (y - H_0)^2 - zd
\]

For \(\pi' < \Pi^*_p(H_0)\) we have

\[
y > H_1 - \frac{d(\Delta H - z)}{2(d + g - \sqrt{g(d + g)})}
\]

Plug in \(y^*\) above we get

\[
\frac{z}{\Delta H} < 1 - 2\lambda \left( 1 + a - \sqrt{a(a + 1)} \right)
\]

2. If \(\pi_2 < \Pi^*_p(H_0) \leq \pi_3\). Let \(B_I(H_0)\) and \(B_{II}(H_0)\) separately denote the part of the iso-profit curve \(B(H_0)\) in the region \(\{(y, p) | y \in (H_0, H_0 + z)\}\) (we call it
R-I) and \( \{(y,p) | y \in (H_0 + z, H_1) \} \) (we call it R-II). We claim that \( B_I(H_0) \) has no tangency point with the family of iso-profit curves \( \{B(H_1, \pi)\}_{\pi \geq 0} \). To see this, suppose \( B_I(H_0) \) is tangent to some \( B(H_1, \pi_0) \) for some \( \pi_0 > 0 \) at \((y_0, p_0)\). Then \( B(H_1, \pi_0) \) must intersect with \( B(H_1, \pi') \), the iso-profit curve crosses the intersection between L-I and the extended part of \( B_I(H_0) \) in R-II, contradicting to the monotonicity (in \( \pi \)) of the family of \( \{B(H_1, \pi)\}_{\pi \geq 0} \). Thus we know that \( B_I(H_0) \) intersects with the highest iso-profit curve \( B(H_1, \pi^*) \) in the family of \( \{B(H_1, \pi)\}_{\pi \geq 0} \) for some \( \pi^* \) at its corner \( C \) (the intersection between \( B_I(H_0) \) and \( y = H_0 + z \)) and that \( B(H_1, \pi^*) \) is higher than \( B_I(H_0) \) in R-I. Also we have \( B(H_1) \) is higher than \( B(H_1, \pi^*) \) since \( B(H_1) \) is tangent to \( B_{II}(H_0) \) at \((y^*, p^*)\) and \( B(H_1, \pi^*) \) intersects with \( B_{II}(H_0) \) at \( C \). Combing together, we get that \( B(H_1) \) does not intersect with \( B_I(H_0) \) and therefore \((y^*, p^*)\) is the unique intersection of \( B(H_1) \) and \( B(H_0) \). To ensure \( \pi_2 < \Pi^*_p(H_0) \), we need

\[
\frac{z}{\Delta H} \leq \frac{-d + \sqrt{d^2 + 4g(1 - \lambda)^2(d + g)}}{2g}
\]

3. If \( \Pi^*_p(H_0) \geq \pi_3 \). In this case, the iso-profit curve \( B(H_0) \) has no part in the region \( \{(y,p) | y \in (H_0, H_0 + z)\} \). Therefore \((y^*, p^*)\) has to be the unique intersection of \( B(H_1) \) and \( B(H_0) \).

Combining the above analysis, for Type 1 equilibrium to exist, we need

\[
\frac{z}{\Delta H} < \min \left\{ 1 - \lambda, (1 + a)(1 - \lambda)^2, 1 - 2\lambda \left(1 + a + \sqrt{a(a + 1)}\right) \right\}
\]

Since it can be shown that the third term in the above curved bracket is always the smallest one of the three irrespective of the parameters chosen, the condition can be
reduced to
\[
\frac{z}{\Delta H} < 1 - 2\lambda \left(1 + a + \sqrt{a(a+1)}\right)
\]

**Type 2.** This is the case when \(B(H_1)\) and \(B(H_0)\) uniquely intersect at some point with \(y^* \in (H_0, H_0 + z)\). First find locus of the contact curve between \(B(H_1)\) and \(B(H_0)\) for \(y \in (H_0, H_0 + z)\). Let \((y^*, p^*)\) be a point on this curve, then

\[
(y^*, p^*) \in \arg \max_{(y, p)} \{(y - H_1)(\mathbb{E}[v] - g - p) : (y - H_0)(\mathbb{E}[v] - p) \geq \Pi^*_p(H_0)\}
\]

Compute the Lagrangian we get

\[
\mathbb{L} = (y - H_1)(\mathbb{E}[v] - g - p) - \alpha[(y - H_0)(\mathbb{E}[v] - p) - \Pi^*_p(H_0)]
\]

From the first order conditions we get the contact curve (call it L-I):

\[
\Delta H(\mathbb{E}[v] - p) - g(y - H_0) = 0
\]

In this case \(p^* = \mathbb{E}[v] - (1 - \lambda)g + \lambda d\). Plug in to the above equation we get \(y^* = \frac{g - (d + g)\lambda}{g}\Delta H + H_0\). Since \(y^* \in (H_0, H_0 + z)\), we need to have

\[
\lambda < \frac{a}{1 + a} \\
\frac{z}{\Delta H} > 1 - \frac{1 + a}{a}\lambda
\]

With \((y^*, p^*)\) we can compute the equilibrium profits:

\[
\Pi^*_p(H_1) = \lambda^2 \frac{(d + g)^2}{g} \Delta H \\
\Pi^*_p(H_0) = \frac{(g - (d + g)\lambda)^2}{g} \Delta H
\]
Obviously, both $\Pi^*_{P}(H_1) > 0$ and $\Pi^*_{P}(H_0) > 0$. Finally, we need to find the condition under which $(y^*, p^*)$ is the unique intersection of $B(H_1)$ and $B(H_0)$. Let $(y', p')$ be the intersection between $B(H_1)$ and $L-I$. Let $\pi'$ be the profit level such that $B(H_0, \pi')$ is tangent to $B(H_1)$ at $(y', p')$. We claim that $(y^*, p^*)$ is the unique intersection of $B(H_1)$ and $B(H_0)$ iff $\pi' < \Pi^*_{P}(H_0)$. Given any point $(y, p)$ on $L-I$, it can be computed that

$$\pi' = \frac{\Delta H}{d + g} \left( d + g + \frac{y - H_1}{\Delta H} \sqrt{g(d + g)} \right)^2 - zd$$

$$\Pi^*_{P}(H_0) = \frac{g}{\Delta H}(y - H_0)^2$$

For $\pi' < \Pi^*_{P}(H_0)$ we have

$$y < H_1 - \frac{d(\Delta H - z)}{2(\sqrt{g(d + g)} - g)}$$

Plug in $y^*$ above we get

$$\frac{z}{\Delta H} > 1 - 2\lambda \left( \sqrt{a(1 + a)} - a \right) \frac{1 + a}{a}$$

Combing the above conditions we have

$$\lambda < \frac{a}{1 + a} \text{ and } \frac{z}{\Delta H} > \max \left\{ 1 - 2\lambda \left( \sqrt{a(1 + a)} - a \right) \frac{1 + a}{a}, 1 - \frac{1 + a}{a} \lambda \right\}$$

Since it can be shown that the first term in the above curved bracket is always greater than the second term irrespective of the parameters chosen, the conditions can be reduced to

$$\lambda < \frac{a}{1 + a} \text{ and } \frac{z}{\Delta H} > 1 - 2\lambda \left( \sqrt{a(1 + a)} - a \right) \frac{1 + a}{a}$$
Type 3. This is the case when $B(H_1)$ and $B(H_0)$ are tangent to each other at two points: $(y_i^*, p_i^*)$ with $y_i^* \in (H_0, H_0 + z)$ and $(y_{II}^*, p_{II}^*)$ with $y_{II}^* \in (H_0 + z, H_1)$. From the analysis of Type 1 and Type 2 we know that the two cut points are

$$y_i^* = H_1 - \frac{d(\Delta H - z)}{2\left(\sqrt{g(d + g)} - g\right)}$$

$$y_{II}^* = H_1 - \frac{d(\Delta H - z)}{2\left(d + g - \sqrt{g(d + g)}\right)}$$

Since $(y_i^*, p_i^*)$ is on L-I and $(y_{II}^*, p_{II}^*)$ is on L-II, plug in we get

$$p_i^* = \mathbb{E}[v] - g + \frac{(1 - \frac{z}{\Delta H})dg}{2\left(\sqrt{g(d + g)} - g\right)}$$

$$p_{II}^* = \mathbb{E}[v] - g + \frac{(1 - \frac{z}{\Delta H})d(d + g)}{2\left(d + g - \sqrt{g(d + g)}\right)}$$

To ensure that $y_i^* \in (H_0, H_0 + z)$ and $y_{II}^* \in (H_0 + z, H_1)$ we have

$$\frac{z}{\Delta H} > 1 - 2\left(\sqrt{a(1 + a)} - a\right)$$

To find the strategy consistent with these quantities, let $\mu_0 = \mathbb{P}(X_1(H_0) + H_0 = y_i^*)$ and $\mu_1 = \mathbb{P}(X_1(H_1) + H_1 = y_i^*)$. From the market efficiency condition we have

$$p_i^* = \frac{\lambda\mu_0}{\lambda\mu_0 + (1 - \lambda)\mu_1} \mathbb{E}[v] + \frac{(1 - \lambda)\mu_1}{\lambda\mu_0 + (1 - \lambda)\mu_1} (\mathbb{E}[v] - g)$$

$$p_{II}^* = \frac{\lambda(1 - \mu_0)}{\lambda(1 - \mu_0) + (1 - \lambda)(1 - \mu_1)} (\mathbb{E}[v] + d) + \frac{(1 - \lambda)(1 - \mu_1)}{\lambda(1 - \mu_0) + (1 - \lambda)(1 - \mu_1)} (\mathbb{E}[v] - g)$$

111
Solve for $\mu_0$ and $\mu_1$ we get

$$
\mu_0 = \frac{\lambda k_2 - (1 - \frac{z}{\Delta H})d}{\lambda(k_2 - k_1)}
$$

$$
\mu_1 = \frac{(1 - \frac{z}{\Delta H})d - k_1}{(1 - \lambda)(1 - \frac{z}{\Delta H})d(k_2 - k_1)}
$$

To ensure that $\mu_0 \in (0, 1)$ we have

$$
1 - 2\lambda \left(1 + a - \sqrt{a(1 + a)}\right) < \frac{z}{\Delta H} < 1 - 2\lambda \left(\sqrt{a(1 + a)} - a\right)
$$

To ensure that $\mu_1 \in (0, 1)$ we have if $\lambda < \sqrt{\frac{a}{1 + a}}$ then

$$
1 - 2 \left(\sqrt{a(1 + a)} - a\right) < \frac{z}{\Delta H} < 1 - 2\lambda \left(1 + a - \sqrt{a(1 + a)}\right)
$$

and if $\lambda > \sqrt{\frac{a}{1 + a}}$ then

$$
1 - 2\lambda \left(1 + a - \sqrt{a(1 + a)}\right) < \frac{z}{\Delta H} < 1 - 2 \left(\sqrt{a(1 + a)} - a\right)
$$

Combining above conditions, we see that Type 3 equilibrium does not exist. \qed
Appendix B

Appendix for Chapter 3

B.1 Proof of Proposition 10

The idea of the proof is following: Suppose there exist a submarket \((p, t_1)\) with all types of buyers in equilibrium, then the sellers in this submarket always have incentive to deviate by posting a close enough price, in which submarket there is only one type of buyers. The proof proceeds with three steps. In the first step I show that there exists a \(p'\) close enough to \(p\) such that there is only on type buyer in the submarket \((p', t_1)\). In the second step, the type of buyer in the submarket \((p', t_1)\) is characterized. In the third step, I show that the expected value of the seller \(V_1(p', t_1; n, \phi) > V_1(p, t_1; n, \phi)\). Now suppose \((p, t_1)\) is one submarket in equilibrium where all types of buyers exist. Then the set of market utility of all the buyers \(U(t_2)\) for \(t_2 \in T_2\) is given. Let \(\delta_p(\epsilon) = (p - \epsilon, p + \epsilon)\).

**Step 1** Claim: There are at most finitely many points \(p' \in \delta_p(\epsilon)\) such that the submarket \((p', t_1)\) has more than one type of buyers. Thus, we can always pick a \(p' \in \delta_p(\epsilon)\) for which there are only one type buyers.

To see this, consider any submarket \((p', t_1)\) with multiple types of buyers. Let \(T_2\) denote the set buyer types in this submarket. Then by the perfection condition \((P)\) of equilibrium we need \(U(t_2) = \alpha_2(p', t_1; n)(d_2(t_1, t_2) - p') + (1 - \alpha_2(p', t_1; n))r_2(t_2)\) for all \(t_2 \in T_2\). Since this is a system of equations with only one variable \(\alpha_2\), the system in general has no solution unless the given parameters take specific structures. There are at most finitely many such submarkets because \(T_2\) is finite and thus \(T_2\) has
finitely many possible cases. Therefore, we can always pick a $p' \in \delta_p(\epsilon)$ such that the submarket $(p', t_1)$ has only one type buyers.

**Step 2** Claim: Let $F(t_1, t_2) = d_2(t_1, t_2) - r_2(t_2)$

\[
t_{2,\text{min}} = \arg \min_{t_2 \in T_2} F(t_1, t_2)
\]

\[
t_{2,\text{max}} = \arg \max_{t_2 \in T_2} F(t_1, t_2)
\]

then for $p' < p$ the only type of buyer is $t_{2,\text{min}}$ and for $p' > p$ the only type of buyer is $t_{2,\text{max}}$.

First consider the case $p' < p$. From Step 1 we know that there is only one type of buyer in the submarket $(p', t_1)$. I prove by contradiction. Suppose the only type of buyer is $t_2 \neq t_{2,\text{min}}$, then we can pick a $t'_2$ such that

\[
F(t_1, t'_2) < F(t_1, t_2)
\]

I will show that $V_2(p', t_1, t'_2) > U(t'_2)$, i.e. any $t'_2$ buyer has incentive to participate in the submarket $(p', t_1)$, which contradicts to the supposition that $t_2$ is the only type of buyer in this submarket. To see this, from condition (P) of the equilibrium we have

\[
U(t_2) = V_2(p, t_1, t_2) = \alpha_2(p, t_1; n)(F(t_1, t_2) - p) + r_2(t_2)
\]

\[
= V_2(p', t_1, t_2) = \alpha_2(p', t_1; n)(F(t_1, t_2) - p') + r_2(t_2)
\]

thus

\[
\frac{\alpha_2(p', t_1; n)}{\alpha_2(p, t_1; n)} = \frac{F(t_1, t_2) - p}{F(t_1, t_2) - p'}
\]
and we have

\[
\frac{V_2(p', t_1, t'_2) - r_2(t'_2)}{U(t'_2) - r_2(t'_2)} = \frac{V_2(p', t_1, t'_2) - r_2(t'_2)}{V_2(p, t_1, t'_2) - r_2(t'_2)} = \frac{\alpha_2(p', t_1; n)(F(t_1, t'_2) - p')}{\alpha_2(p, t_1; n)(F(t_1, t'_2) - p)} = \frac{F(t_1, t_2) - p}{F(t_1, t'_2) - p} * \frac{F(t_1, t'_2) - p'}{F(t_1, t'_2) - p}
\]

Define \( G(p) = \frac{F(t_1, t_2) - p}{F(t_1, t'_2) - p} \) then the above equation is equal to \( G(p)/G(p') \). If we can show \( G'(p) > 0 \), then since \( p' < p \) we have \( G(p') < G(p) \) and thus \( V_2(p', t_1, t'_2) > U(t'_2) \).

To check this

\[
G'(p) = \left( \frac{F(t_1, t_2) - p}{F(t_1, t'_2) - p} \right)' = \frac{F(t_1, t_2) - F(t_1, t'_2)}{(F(t_1, t'_2) - p)^2} > 0
\]

Thus we have shown that for \( p' < p \) the only type of buyer in the submarket \((p', t_1)\) has to be \( t_{2, \text{min}} \). By the very similar logic we can also show that for \( p' > p \) the only type of buyer in the submarket \((p', t_1)\) has to be \( t_{2, \text{max}} \).

**Step 3** Claim: Type \( t_1 \) sellers always have incentive to deviate from posting \( p \).

In this model, since each submarket has only one type of sellers, whichever buyer goes to the submarket \((p, t_1)\) is willing to trade at price \( p \) with the seller of type \( t_1 \) upon matching, for otherwise the buyer will not participate in this submarket. This means that the expected value of the seller is

\[
V_1(p, t_1; n, \phi) = \alpha_1(p, t_1; n)(d_1(t_1) - r_1(t_1) + p) + r_1(t_1)
\]

To prove this claim, it suffices to show that \( V_1(p', t_1; n) > V_1(p, t_1; n) \) for some \( p' \). Let
\( F_1(t_1) = d_1(t_1) - r_1(t_1). \) Note that for either \( t_2 \in \{ t_{2,\max}, t_{2,\min} \} \) we have

\[
U(t_2) = \frac{\alpha(n(p', t_1))}{n(p', t_1)} (F(t_1, t_2) - p') + r_2(t_2) \\
= \frac{\alpha(n(p, t_1))}{n(p, t_1)} (F(t_1, t_2) - p) + r_2(t_2)
\]

thus we have

\[
p' = F(t_1, t_2) - \frac{n(p', t_1)}{n(p, t_1)} \frac{\alpha_1(p, t_1; n)}{\alpha_1(p', t_1; n)} (F(t_1, t_2) - p)
\]

and therefore

\[
V_1(p', t_1; n) - V_1(p, t_1; n) = \alpha_1(p', t_1; n)(F_1(t_1) + p') - \alpha_1(p, t_1; n)(F_1(t_1) + p) \\
= \alpha_1(p', t_1; n)(F_1(t_1) + F(t_1, t_2)) \\
- \frac{n(p', t_1)}{n(p, t_1)} \alpha_1(p, t_1; n)(F(t_1, t_2) - p) \\
- \alpha_1(p, t_1; n)(F_1(t_1) + p)
\]

Take the change of variable \( n' = n(p', t_1) \) and define right hand side of the above equation as \( H(n'; t_2) \). Let \( n_0 = n(p, t_1) \). Note that \( H(n_0; t_2) = 0 \) for all \( t_2 \) and our goal is to find a \( n' \) (or \( p' \)) such that \( H(n'; t_2) > 0 \). For this purpose, we need to look at the derivative of \( H \) at \( n_0 \):

\[
H'(n_0; t_2) = \alpha'(n_0) (F_1(t_1) + F(t_1, t_2)) - \frac{\alpha(n_0)}{n_0} (F(t_1, t_2) - p) \\
= \left( \alpha'(n_0) - \frac{\alpha(n_0)}{n_0} \right) F(t_1, t_2) + \alpha'(n_0) F_1(t_1) + \frac{\alpha(n_0)}{n_0} p
\]

Till this point, we haven’t take a stance on which \( t_2 \in \{ t_{2,\max}, t_{2,\min} \} \) to plug into the above equation. Now I will discuss the cases. First suppose \( H'(n_0; t_{2,\max}) < 0 \). Then we can pick any \( p' > p \). From Step 2 above we see that the only type of
buyer in the submarket \((p', t_1)\) is \(t_{2,\text{max}}\). The corresponding \(n' = n(p', t_1)\) in this submarket is strictly smaller than \(n_0\). By the supposition that \(H'(n_0; t_{2,\text{max}}) < 0\) we have \(H(n'; t_{2,\text{max}}) > 0\) and thus \(V_1(p', t_1; n) > V_1(p, t_1; n)\). On the other hand, suppose \(H'(n_0; t_{2,\text{max}}) \geq 0\). Since by definition \(t_{2,\text{min}} < t_{2,\text{max}}\) and \(\alpha'(n_0) - \frac{\alpha(n_0)}{n_0} < 0\) we have \(H'(n_0; t_{2,\text{min}}) > 0\). Then we can pick any \(p' < p\). From Step 2 above we see that the only type of buyer in the submarket \((p', t_1)\) is \(t_{2,\text{min}}\). The corresponding \(n' = n(p', t_1)\) in this submarket is strictly greater than \(n_0\). By the condition that \(H'(n_0; t_{2,\text{min}}) > 0\) we have \(H(n'; t_{2,\text{min}}) > 0\) and thus \(V_1(p', t_1; n) > V_1(p, t_1; n)\). Thus the claim is proved.

**B.2 Proof of Proposition 11**

There are five type of agents in this system \(\Omega = \{K, S, B, M, N\}\), among which \(\{K, S, M\}\) are the potential sellers and \(\{B, N\}\) are the potential buyers. However, \(K\) can only sell to \(N\) and \(M\) can only sell to \(B\). Therefore there are at most four type of tradings in the system: \(S \rightarrow B\), \(S \rightarrow N\), \(M \rightarrow B\), and \(K \rightarrow N\). However, not all four types of tradings will exist in the equilibrium, depending on the given parameters: relative measure of different types and their continuation values.

The approach to solve the equilibrium is by trial and error. First suppose that all four submarkets exist in equilibrium (call it Type 0), solve for it, and see what conditions need to be satisfied to support this equilibrium. Then we can discuss about which submarket will possibly breakdown under the parameter ranges where Type 0 equilibrium is not supported. Now given the continuation values \(\vec{V}' = \{V'_i\}_{i \in \Omega}\) and

\[\int_0^n (\alpha''(x) x + \alpha'(x)) dx < \int_0^n \alpha'(x) dx = \alpha(n)\]

Therefore, we have \(\alpha'(n) - \frac{\alpha(n)}{n} < 0\) for all \(n\).

---

1To see this, note that \(\alpha''(x)x + \alpha'(x) < \alpha'(x)\) for all \(x\) since \(\alpha''(\cdot) < 0\). Take the integral on both sides we get

\[\alpha'(n)n = \int_0^n (\alpha''(x)x + \alpha'(x)) dx < \int_0^n \alpha'(x) dx = \alpha(n)\]

Therefore, we have \(\alpha'(n) - \frac{\alpha(n)}{n} < 0\) for all \(n\).
the measures \( \vec{L} = \{l_i\}_{i \in \Omega} \), let \( \Delta = V'_M - V'_N \).

**Step 1** List the four single-submarkets maximization conditions

In the submarket \( S \rightarrow B \) we have

\[
V_B = \alpha'(n_1)u_B \\
V_S = (\alpha(n_1) - n_1\alpha'(n_1))u_B + rV'_S \\
n_1 = l_{BS}/l_{SB}
\]

In the submarket \( S \rightarrow N \) we have

\[
V_N = \alpha'(n_2) \cdot r\Delta + rV'_N \\
V_S = (\alpha(n_2) - n_2\alpha'(n_2)) \cdot r\Delta + rV'_S \\
n_2 = l_{NS}/l_{SN}
\]

In the submarket \( M \rightarrow B \) we have

\[
V_B = \alpha'(n_3) \cdot (u_B - r\Delta) \\
V_M = (\alpha(n_3) - n_3\alpha'(n_3)) \cdot (u_B - r\Delta) + rV'_M \\
n_3 = l_{BM}/l_{M}
\]

In the submarket \( K \rightarrow N \) we have

\[
V_N = \alpha'(n_4) \cdot r\Delta + rV'_N \\
V_K = (\alpha(n_4) - n_4\alpha'(n_4)) \cdot r\Delta + rV'_S \\
n_4 = l_{NK}/l_{K}
\]
Step 2 Solve for the market tightness $n_i$ with $i = 1, 2, 3, 4$

Note that for $S$ to participate in submarkets $S \to B$ and $S \to N$. Similarly, for $B$ and $N$. Therefore, we have three equations:

\[
(\alpha(n_1) - n_1\alpha'(n_1))u_B + rV'_S = (\alpha(n_2) - n_2\alpha'(n_2)) \cdot r\Delta + rV'_S
\]

\[
\alpha'(n_1)u_B = \alpha'(n_3) \cdot (u_B - r\Delta)
\]

\[
\alpha'(n_2) \cdot r\Delta + rV'_N = \alpha'(n_4) \cdot r\Delta + rV'_N
\]

Since both $f(n) = \alpha(n) - n\alpha(n)$ and $f(n) = \alpha'(n)$ are strictly increasing functions, $n_2$ and $n_3$ can be uniquely solved in terms of $n_1$ from the above equations. Thus, we get $n_2 = n_4$ and implicit functions $n_2 = \phi_2(n_1)$ and $n_3 = \phi_3(n_1)$. Note that $\phi_2$ and $\phi_3$ are also strictly increasing functions. Besides, we have three market clearing conditions:

\[
l_S = l_{SB} + l_{SN}
\]

\[
l_B = l_{BS} + l_{BM}
\]

\[
l_N = l_{NK} + l_{NS}
\]

Combining the above seven equations together with the definitions for market tightness we get

\[
\frac{n_1}{l_{SB}} = \frac{l_B - l_{BM}}{l_S - l_{SN}} = \frac{l_B - l_M n_3}{l_S - l_{NS}/n_2} = \frac{l_B - l_M n_3}{l_S - (l_N - l_{NK})/n_2} = \frac{l_B - l_M n_3}{l_S + l_K - l_N/n_2}
\]

\[
= \frac{l_B - l_M \phi_3(n_1)}{l_S + l_K - l_N/\phi_2(n_1)}
\]
Since both $\phi_2$ and $\phi_3$ are strictly increasing functions, a unique solution for $n_1 = n_1^*$ can be solved from the above equation, which further gives us $n_2 = n_4 = \phi_2(n_1^*)$ and $n_3 = \phi_3(n_1^*)$. To get an analytical solution, let $\alpha(n) = \min(an^b, 1)$ with $a$ large enough such that for all the equilibrium market tightness $n$ we have $\alpha(n) < 1$. Substitute $\alpha$ into the above equations and we get

$$n_2 = \phi_2(n_1) = \left(\frac{r\Delta}{u}\right)^{-\frac{1}{b}} n_1 = c_2 n_1$$

$$n_3 = \phi_3(n_1) = \left(1 - \frac{r\Delta}{u}\right)^{1-b} n_1 = c_3 n_1$$

$$n_1^* = \frac{l_B + l_n/c_2}{l_S + l_K + c_3 l_M}$$

**Step 3** Solve for the $l_{ij}$ and check for feasibility conditions

Specifically, we need to make sure that $0 \leq l_{NK} \leq l_N$, $0 \leq l_{BM} \leq l_B$, and $0 \leq l_{SB} \leq l_S$.

$$l_{NK} = n_4^* l_K = c_2 n_1^* l_K = \frac{(c_2 l_B + l_N) l_K}{l_S + l_K + c_3 l_M}$$

$$l_{BM} = n_3^* l_M = c_3 n_1^* l_M = \frac{l_B + l_n/c_2}{l_S + l_K + c_3 l_M}$$

$$l_{SB} - \frac{l_{BS}}{n_1^*} = \frac{l_B - l_{BM}}{n_1^*} = \frac{l_B}{n_1^*} - c_3 l_M = \frac{l_B(l_S + l_K + c_3 l_M)}{l_B + l_n/c_2} - c_3 l_M$$

Note that $l_{NK}$ is always greater than 0, but for $l_{NK} \leq l_N$ we need

$$l_N(l_S + c_3 l_M) \geq c_2 l_B l_K$$

Similarly, $l_{BM}$ is always greater than 0, but for $l_{BM} \leq l_B$ we need

$$c_2 l_B(l_S + l_K) \geq c_3 l_M l_N$$
For $l_{SB} \geq 0$ we need

$$c_2 l_B (l_S + l_K) \geq c_3 l_M l_N$$

and for $l_{SB} \leq l_S$ we need

$$l_N (l_S + c_3 l_M) \geq c_2 l_B l_K$$

Note that the above four inequalities are the duplicate of two inequalities. This is intuitive as for $l_{SB} > 0$ sellers $S$ are sure to sell goods to some positive measure of buyers $B$, which is equivalent to saying that the buyers $B$ don’t all get their goods from middlemen $M$, i.e. $l_{BM} < l_B$. Now denote $\bar{l}_B = c_2 l_B$ and $\bar{l}_M = c_3 l_M$. From the above two constraints we see that when

$$\frac{\bar{l}_M}{l_S + l_K} \leq \frac{\bar{l}_B}{l_N} \leq \frac{l_S + \bar{l}_M}{l_K}$$

all the equilibrium conditions are satisfied and we have solved for one equilibrium.

### B.3 Proof of Proposition 13

First consider the case when the producers know the buyers’ preference in each period. Then obviously they can allocate the good to which type in $\{A, B\}$ it is. After that the market tightness is $l_b/l_p$ and from the section of one-period analysis, we see that the total number of match between producers and the buyers in each period is $\alpha (l_b/l_p) l_p$ and upon meeting, the buyers get $b$ share of the total surplus ($u$) of the two sides. This means $TS = a (l_b/l_p)^b l_P u$

Now consider the case when the producers have no information about the buyers’ preference. Suppose they use a symmetric mixing strategy as to the allocation decision of their good, i.e. they allocate with probability $\sigma$ to $A$ while with probability $1 - \sigma$ to $B$. Let $U_A, U_B$ denote the expected value of holding a unit of good in $A$ and in $B$
after the allocation. Then from Proposition 11 we have

\[ U_A = (1 - b)\alpha \left( \frac{l_b}{\sigma l_p} \right) up \]

\[ U_B = (1 - b)\alpha \left( \frac{l_b}{(1 - \sigma)l_p} \right) u(1 - p) \]

Since the producers should be indifferent, solve the equation \( U_A = U_B \) for \( \sigma \) we get

\[ \sigma^* = \frac{1}{1 + \left( \frac{1-p}{p} \right)^b} \]

Note that \( \sigma^* \) is increasing with \( p \) which is very intuitive as the more likely the preference be in A the producers should allocate their good to A with higher probability. There are possibly two corner equilibrium where all the producers allocate their good to A (\( \sigma = 1 \)) or B (\( \sigma = 0 \)). The incentive constraint here is bit tricky as there is continuum of producers. It is a bit hard to think about one single producer deviating and facing a continuum of buyers, as in this case there is no competition among the sellers as usual and the direct search framework doesn’t come along. The way to think about the incentive constraint here is for an \( \epsilon \) number of producers to deviate and then let \( \epsilon \) go to zero. First consider the incentive constraint for case \( \sigma = 1 \). In this case,

\[ U_A = (1 - b)\alpha \left( \frac{l_b}{l_p} \right) up \]

\[ U_B = (1 - b)\alpha (\infty) u(1 - p) = (1 - b)u(1 - p) \]
For $U_A > U_B$ we need $p > \frac{1}{1+\alpha(l_b/l_p)}$. Similarly, when $\sigma = 0$, we have

$$U_A = (1 - b)\alpha(\infty)up = (1 - b)up$$
$$U_B = (1 - b)\alpha\left(\frac{l_b}{l_p}\right)u(1 - p)$$

For $U_B > U_A$ we need $p < \frac{1}{1+\alpha(l_b/l_p)}$.

### B.4 Proof of Proposition 14

I will take a guess and verify approach to prove the existence and uniqueness of the stationary equilibrium. In this case, Definition 4 collapses to have only one state and the transition probability matrix being going from that one state to itself with probability one. Since the realization of buyers’ preference should not matter, it is necessary to have $\sigma = 0.5$ in equilibrium, for otherwise the quantities of $l_S$ and $l_K$ depend on the realization. Now given the measures $\{l_P, l_B, l\}$, want to find the constant $l^*_M$ and $\{V^*_P, V^*_M, V^*_N\}$ such that the equilibrium conditions of Definition 4 are satisfied.

**Case 1** First suppose the stationary equilibrium is of type 1. Let $n_{KN}$ and $n_{MB}$ be the submarket tightness measures as defined in Proposition 11

$$n_{KN} = \frac{l - l_M}{\frac{l_P}{2}} \text{ and } n_{MB} = \frac{l_B}{\frac{l_P}{2c3} + l_M}$$

123
Let $\Delta = V_M - V_N$. Then the equilibrium conditions are

\begin{align*}
V_M &= U_M = (1 - b)\alpha (n_{MB}) (u - r\Delta) + rV_M \\
V_N &= U_N = \alpha' (n_{KN}) r\Delta + rV_N \\
l_M &= l_M + \alpha (n_{KN}) \frac{l_P}{2} - \alpha (n_{MB}) l_M
\end{align*}

Note here I ignore the condition of $V_P$, this is because with $p = \sigma = 0.5$, this condition is always satisfied. Now take the difference of above first two equations we get an equation in $\Delta$ and combine it with the last equation above, we get a system of two equations with two variables $l_M$ and $\Delta$:

\begin{align*}
0 &= \alpha \left( l - l_M \right) \frac{l_P}{2} - \alpha \left( \frac{l_B}{l_P c_3 + l_M} \right) l_M \\
(1 - r)\Delta &= (1 - b)\alpha \left( \frac{l_B}{l_P c_3 + l_M} \right) (u - r\Delta) - \alpha' \left( \frac{l - l_M}{l_P} \right) r\Delta
\end{align*}

where $c_3 = (1 - \frac{r\Delta}{u})^{\frac{1}{1-b}}$ above is defined as in Proposition 11. It then suffices to show that there exists a unique solution $l_M^*$ and $\Delta^*$ to the above pair of equations. I will first show the existence by the fact that the implicit functions of $\Delta$ defined by the above equations are continuous in $l_M$. The uniqueness is proved by showing that $\Delta$ is increasing in $l_M$ in the first equation and decreasing in $l_M$ in the second equation.

From equation (2), solve for $\Delta$ we get

$$
\Delta = \frac{(1 - b)u}{(1 - b)r + \frac{br}{l_M - 1} + \frac{(1 - r)l_M}{l_P \alpha \left( \frac{l - l_M}{l_P} \right)}}
$$

It is obvious that $\Delta$ is decreasing in $l_M$. Moreover, the range of $\Delta$ goes from $\frac{u}{r}$ down
to 0 as \( l_M \) increases from 0 to \( l \). Now from equation (1) solve for \( \frac{l_P}{2c_3} \) we get

\[
\frac{l_P}{2c_3} = \frac{l_B}{(\frac{l_P}{2})^{\frac{1}{b} - 1}} l - l_M
\]

Let \( R(l_M) \) denote the right hand side of the above equation and let \( z \) be the constant term in it. Take the first and second derivative we get

\[
\frac{\partial R}{\partial l_M} = z\frac{\frac{1}{b}(l_M)^{\frac{1}{b} - 1}(l - l_M) + (l_M)^{\frac{1}{b}}}{(l - l_M)^2} - 1
\]

\[
\frac{\partial R^2}{\partial^2 l_M} = z\frac{\frac{1}{b^2}(l_M)^{\frac{1}{b} - 2}(l - l_M)^3 + \frac{2}{b}(l_M)^{\frac{1}{b} - 1}(l - l_M)^2 + 2(l_M)^{\frac{1}{b}}(l - l_M)}{(l - l_M)^4} > 0
\]

Let \( \tilde{l}_M \) be defined as the solution to the above equation when \( \Delta = 0 \), i.e.

\[
\frac{l_P}{2} = z\frac{\frac{1}{b} l_M}{l - \tilde{l}_M} - \tilde{l}_M
\]

Solve for \( z \) in terms of \( \tilde{l}_M \) and plug into the first derivative we get

\[
\frac{\partial R}{\partial l_M}(\tilde{l}_M) = z\frac{\frac{1}{b}(\tilde{l}_M)^{\frac{1}{b} - 1}(l - \tilde{l}_M) + (\tilde{l}_M)^{\frac{1}{b}}}{(l - \tilde{l}_M)^2} - 1
\]

\[
= \frac{(l - \tilde{l}_M)(\frac{l_P}{2} + \tilde{l}_M)}{(\tilde{l}_M)^{\frac{1}{b}}} \frac{\frac{1}{b}(l_M)^{\frac{1}{b} - 1}(l - \tilde{l}_M) + (\tilde{l}_M)^{\frac{1}{b}}}{(l - \tilde{l}_M)^2} - 1
\]

\[
= \left(\frac{l_P}{2} + \tilde{l}_M\right) \left(\frac{1}{b\tilde{l}_M} + \frac{1}{l - l_M}\right) - 1
\]

\[
> \frac{1}{b} - 1 > 0
\]

Combing the facts that \( \frac{\partial R^2}{\partial^2 l_M} > 0 \) and \( \frac{\partial R}{\partial l_M}(\tilde{l}_M) > 0 \) we see from equation (1) that \( \Delta \) increases from 0 to \( \frac{z}{\tilde{l}_M} \) as \( l_M \) increases from \( \tilde{l}_M \) to \( l \). Remember I have shown above that from equation (2) \( \Delta \) goes from \( \frac{z}{\tilde{l}_M} \) down to 0 as \( l_M \) increases from 0 to \( l \). Since these functions of \( \Delta \) are continuous, they have to intersect one and only one time on \((\tilde{l}_M, l)\).
Let \((l^*_M, \Delta^*)\) be the intersection, it is also a solution to the pair of equations (1) and (2). To show the uniqueness, I will show that the two implicit \(\Delta\) functions do not intersect on \([0, \tilde{l}_M]\). To see this, again by the facts \(\frac{\partial R^2}{\partial l^2_M} > 0\) and \(\frac{\partial R}{\partial l^M} (\tilde{l}_M) > 0\) we see that \(R(x) \leq \max(R(0), R(\tilde{l}_M)) = \frac{l_P}{2}\) on \([0, \tilde{l}_M]\). However, the left hand side \(\frac{l_P}{2\gamma} (\Delta)\) is monotonically increasing in \(\Delta\) and thus \(\geq \frac{l_P}{2}\). In other words, the implicit \(\Delta\) function derived from equation (1) is not defined on \([0, \tilde{l}_M]\). It is also obvious that the two implicit functions cannot intersect at point \(\tilde{l}_M\). Thus, the existence and uniqueness is proved. Finally, we need to check that the type 1 condition in Proposition 11 is satisfied with the unique solution \((l^*_M, \Delta^*)\).

**Case 2** Suppose the stationary equilibrium is of type 2. The pair of equations in \(l_M\) and \(\Delta\) is

\[
0 = \alpha \left( \frac{l - l^*_M}{l_P} \right) l_P - \alpha \left( \frac{l_B}{l_M} \right) l_M \tag{B.3}
\]

\[
(1 - r)\Delta = (1 - b)\alpha \left( \frac{l_B}{l_M} \right) (u - r\Delta) - \alpha' \left( \frac{l - l^*_M}{l_P} \right) r\Delta \tag{B.4}
\]

This is even simpler than the previous case as we can solve for \(l^*_M\) from equation (3) and then it into equation (4) to get \(\Delta^*\). From equation (3) we have

\[
\frac{l_B}{(l_P)^{\frac{1}{\beta} - 1}} = \frac{l - l^*_M}{(l_M)^{\frac{1}{\beta} - 1}}
\]

Since the right-hand side of above equation is monotone in \(l_M\) and decreases from \(+\infty\) down to 0 as \(l_M\) increases from 0 to \(l\), there exists a unique solution \(l^*_M\). Plug this \(l^*_M\) into equation (4) we get

\[
\Delta^* = \frac{u}{r + \frac{b}{1-b} \frac{r}{l^*_M} + \frac{1-r}{1-b} \frac{l^*_M}{l_P} \frac{1}{\alpha (\frac{l^*_M}{l_P})}}
\]
As in the last case, we still need to check that the type 2 condition in Proposition 11 is satisfied.

**Case 3** Suppose the stationary equilibrium is of type 0. The system of equations cannot be disentangled in the two variables in anyway and thus hard to prove. Consider proving by contraction mapping in the future.
Bibliography


Biography

Chao Yang attended Fudan University in Shanghai majoring international economics and trade. After that he got a master’s degree in economics from Duke University. He will graduate from Duke University’s Fuqua School of Business with a PhD in Business Administration.