Understanding Supply Chain Disruptions – Empirical Analysis of Supply Chain Structures

by

Yixin Wang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in the University of Michigan
2018

Doctoral Committee:

Professor Ravi Anupindi, Co-Chair
Assistant Professor Jun Li, Co-Chair
Professor Mark S. Daskin
Associate Professor Xun Wu

Yixin (Iris) Wang

iriswang@umich.edu

ORCID iD: 0000-0003-3152-4539

Acknowledgments

I would like to give warm thanks to Jun Li, my co-advisor, who spent a boundless amount of time training me to become a researcher, for her generous help on everything—it is my fortune to be your first student and friend; to Ravi Anupindi, my co-advisor, for his all-around advice and for calming my anxiety with encouragement during job search; to Brian Wu and Mark Daskin for serving on my thesis committee and providing valuable insights into this document; to Roman Kapuscinski, whose seminar course is my favorite, for showing me great teaching practices and enthusiasm for research; to Brian Jones and other staff in Ross Doctoral Office for making everything as smooth as possible; to Juan Serpa, who introduced me to people at POMS when I felt lost as a new comer; to Anyan Qi, Yao Cui, and Ruth Beer for useful advice when I joined the program; to Evgeny Kagan and Yanzhe (Murray) Lei for being the best cohort; to Lai Wei, Zhaohui (Zoey) Jiang, and Aravind Govindarajan with whom I shared joys and hardships of being a PhD student; to Qi Chen and Xu Zhang for being long-time friends and my matchmakers; to my parents and my grandparents for their love and constant support in my academic career; and especially to Nan, my dearest husband, for his continued love, support and understanding—asking you out on a date was the best decision I have ever made in my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Firms have encountered an ever-increasing number of supply chain disruptions in the past decade, triggered by a wide range of natural and man-made causes. Supply chain risk management is thus an active research area, while the concerning topics mainly focused on the management at the immediate supplier level. In contrast, anecdotes reveal that shortages can oftentimes be traced back to the problems at sub-tier suppliers, i.e., tier-2 or more upstream suppliers. Further, the structure of a supply chain is not exogenously generated. For example, lack of incentives may discourage manufacturers from entering a market and create a highly concentrated industry that could be vulnerable to supply shocks and price manipulation. These two topics, the sub-tier supplier structure and impacts and the entry decisions of manufacturers are the two main themes of this thesis.

More specifically, this thesis presents empirical results that improve our understanding of 1) risk propagation from sub-tier suppliers to the connected focal firms and 2) barriers to entry for manufacturers. The first part of the thesis considers the sub-tier suppliers and the network structure that connects the supply chain partners. It demonstrates the financial performance link between firms and their tier-2 suppliers respectively. It also establishes the intermediary effect of network concentration: when a firm's tier-1 suppliers share tier-2 suppliers. The second part of the thesis focuses on the generic pharmaceutical industry plagued by the high concentration of firms in markets with expired patents. This chapter studies the key determinants of market entry decisions by generic firms and confirms the role of manufacturing process and regulatory environment. Policy simulation result shows the non-monotone relationship between the speed of the government review process and market concentration level.

CHAPTER 1

Introduction

In the recent years, firms in multiple industries have experienced a surge of supply chain discontinuities. As a consequence, many consumer goods have been reported in short supply, including on-demand electronics and automobiles, as well as medications for treating life-threatening diseases (Kyodo 2011, Fink Jan 29, 2016).

The supply chain risk management literature has investigated many important factors that help mitigate supply disruptions. One stream of research analyzes the role of inventory in protecting firms against supply chain disruptions (see Parlar and Berkin 1991, Parlar 1997, Qi et al. 2009, for example). Others explore the use of supplier diversification as a viable strategy to mitigate supply chain risks (see Li et al. 2010, Babich et al. 2007, for example). Additionally, scholars have also looked into the benefit of flexibility (Tang and Tomlin 2008, Huchzermeier and Cohen 1996), vertical integration (Braunscheidel and Suresh 2009), and the level of trust between supply chain partners (Bode et al. 2011) on firms' supply chain agility. Sodhi and Chopra (2004) broadly categorize seven types of supply chain risks and discuss the drivers of each risk category and their mitigation strategies.

The literature is rich in theoretical analyses, whereas empirical work is relatively sparse. The limited data availability and the restricted data access make it particularly challenging to conduct empirical studies in this field. Unlike retailers who actively collect information about their customers, firms rarely collect performance information of suppliers. Even if firms collect such supplier information, they are not willing to share the propriety data with outsiders due to, for example, competitive reasons (Ang et al. 2016). Data for this thesis is acquired from various data sources, including publicly available databases, Freedom of Information Act data inquiries, and web crawling.

Besides the research methodology, topic-wise, the supply chain risk management literature mainly targets on the risk management at manufacturers, at immediate suppliers, or at the interactions between the two supply chain partners. On the contrary, the 2011 Japan earthquake and tsunami and Thailand flood revealed substantial supply disruption risks originating from unknown sub-tier suppliers, especially in automotive and high-tech industry

(Masui and Nishi 2012). In addition, public concerns on the recent unstable drug supply and high drug prices revealed the market concentration problem in the regulated pharmaceutical industry, i.e., the insufficient number of generic manufacturers in selected drug products (GAO 2016b). To address the two practical concerns respectively, this thesis presents empirical efforts towards providing firms and policy makers with new insights on how to maintain a continuous supply of goods.

Below I give an overview of the thesis and summarize the contributions. Chapter 2 is a descriptive analysis about the sub-tier supply chain structure. Chapter 3 and 4 contain findings from two completed research papers.

Tier-2 Sharing in Multi-tier Supply Chains (Chapter 2)

After the Japanese earthquake and the Thailand flood in 2011, firms started to realize that what they used to think as pyramid shaped supply chain is actually diamond shaped. That is, even if firms use multi-sourcing strategy to mitigate supplier risks, their suppliers may choose to source from the same tier-2 suppliers. Those shared tier-2 suppliers then become the pinch points in the supply chain. In this chapter, using panel data on firm-level supply chain relationship, we descriptively document the prevalence of tier-2 sharing across industries. We find that the tier-2 sharing has widespread presence in firms' multi-tier supply chains, and the high-tech sector has the highest degree of tier-2 sharing among the manufacturing industry.

Risky Suppliers or Risky Supply Chains (Chapter 3)

This chapter focuses on the financial risk transmission from sub-tier suppliers to customers in the high-tech industry. Motivated by industry findings (Masui and Nishi 2012, Japan METI 2011), this chapter studies the risks originated from tier-2 suppliers and looks into the impact of a specific supply network structure: the sharing of a firm's tier-2 suppliers by its tier-1 suppliers. We show the causal link between the stock market performances of remotely connected firms, and we also find that a firm experiences a more negative market reaction if its disrupted tier-2 supplier is shared by a higher number of tier-1 suppliers. Surprisingly, the magnitude of impact at the focal firm is similar to that at the directly disrupted tier-2 suppliers. This result underscores the need for firms to monitor sub-tier suppliers, and implies the potential for firms to prioritize their efforts when managing sub-tier supplier risks. This chapter is based on a joint work with Jun Li and Ravi Anupindi (Wang et al. 2017).

Manufacturing and Regulatory Barriers to Generic Drug Competition (Chapter 4)

Generic medications typically attract a small number of manufacturers, with more than half of the generic drugs produced by at most three firms. This insufficient market entry by generic manufacturers makes the pharmaceutical supply chain vulnerable to supply shock and price manipulation. This chapter investigates the determinants of a firm's entry decision into generic drug markets. Unlike entry game literature for other industries (e.g., Berry 1992, Aguirregabiria and Ho 2012), this chapter mainly focuses on the impact of regulatory environment because this is the leverage that the government can use to encourage entries from generic manufacturers. The counterfactual analysis shows a perhaps surprising result. Note that government approval is required before generic manufacturers can bring drug product to the market. We find that a shorter time to approval, which implies a reduced opportunity cost, does not always attract more manufacturers. An explanation for this phenomenon is provided based on the competition theory: the perceived crowded market deters players from entering. This result suggests that the government should be more cautious in aggressively reducing the time to approval, since it does not necessarily translate into the desired more competitive market. This chapter is based on a joint work with Jun Li and Ravi Anupindi entitled "Manufacturing and Regulatory Barriers to Generic Drug Competition: A Structural Model Approach".

Finally, we conclude in Chapter 5 with a brief summary and outline some future research directions.

CHAPTER 2

Tier-2 Sharing in Multi-tier Supply Chains

In the past decade firms have encountered an ever-increasing number of supply chain disruptions, triggered by a wide range of natural and man-made causes such as earthquake, flood, fire, labor protest, financial crisis and political unrest. These events have led to substantial short-term losses (e.g., production delay, increased labor, and supply costs) as well as long-term losses (e.g., market share erosion and bankruptcy). Among those events, supply chain disruptions originating from sub-tier suppliers have increasingly caught the attention of academia and industry, both of which used to focus on the risk management of immediate tier-1 suppliers. In fact, the annual Supply Chain Resilience Surveys has consistently found that almost half of the surveyed firms have experienced supply chain disruptions originated from the sub-tier suppliers, especially from the tier-2 suppliers.

The supply chain disruptions from sub-tier suppliers are problematic, and yet the multi-tier supply chain structure may amplify the impact of the disruptions and make the situations even worse. Monthly after the 2011 Japanese tsunami, automakers started to realize the concentration level in their upper tier supply chain. For example, Renesas Electronics, a tier-2 semiconductor chip supplier of Toyota, provided customized chips to tens of Toyota's tier-1 suppliers (Pollack and Lohr Apr 27, 2011). The overlap in the upper-tier suppliers is considered one of the causes that lead to the long lasting impact after the disaster, as firms have no direct control on the higher tiers.

Despite the anecdotes, it remains unclear 1) whether the tier-2 sharing is widely observed in firms' supply chains; and 2) whether such a phenomenon is concentrated in selected industries. To look into these issues, we collect panel data on supplier-customer relationships.

The most commonly used *firm-level* supplier-customer relationship data source in literature is Compustat. However, due to the reporting rule, Compustat only identifies those immediate customers who contribute more than 10 percent of revenues to the focal firm, while systematically under-sampling major suppliers, especially those small to medium customer firms. For this reason, we acquire the supplier-customer relationships from a different data source, Factset. Similar to Compustat, Factset is also a longitudinal dataset. It contains all relationships documented in Compustat and complement the latter with information from other sources, so the resulting dataset is more comprehensive than Compustat.

Using Factset, we build a longitudinal supply network containing 36,975 firms and 372,392 directional links. From the perspective of a tier-0 firm, a shared tier-2 supplier gives rise to a network structure that takes the shape of a diamond (Japan METI 2011). To characterize the diamond shape in supply networks, we define a metric, *degree of commonality*, which is the number of tier-1 suppliers that source from the same tier-2 supplier. Our results show that on average, each tier-2 supplier provides inputs to 1.3 of the focal firm's tier-1 suppliers, and 7.1 percent of tier-2 suppliers are shared by five or more tier-1 suppliers. The extent of tier-2 sharing also varies across sectors. Within the manufacturing industry, firms in the high-tech sector has the highest degree of tier-2 sharing.

2.1 Supplier-Customer Relationship Data

We collect relationship information from Factset, a financial information and software company for investment professionals.[1] Factset supplements Compustat, the commonly used dataset for supply chain management research, with information based on regulatory filings and company websites. Factset currently comprises supply chain relationships of 23,400 global companies, and reports the start date and the end date for each of these identified relationships. For each relationship, it also categorizes the relationship type in the following four buckets: supplier, customer, partner (marketing, licensing, etc.), and competitor. When available, Factset also specifies percentage of the supplier's revenue that the customer firm makes up (% Revenue).

We focus on relationships with type 'supplier' or 'customer' from 2004 to mid-2015, and we followed the steps below to clean the relationship data: 1) we removed those relationships with non-identifiable firm names ; and 2) we removed those relationships that involve government procurement to focus on business between firms. After data cleaning, we ended up with about 174,333 supply chain relationships across all industries, spanning thirteen years.

We now evaluate the quality of Factset relative to Compustat using valid relationships in 2012 as an example.[2] Specifically, we compare supplier coverage of these two data sources with regard to the number of suppliers identified. In 2012, Factset identified 50,385 supplier-customer relationships (9,469 firms with supplier information), and Compustat identified

[1] FactSet Revere dataset, which contains corporate relationships, supply chains, and geographic risk exposure information, is available on the Wharton Research Data Services (WRDS) platform.

[2] Evaluations using relationships in other years generate consistent conclusions.

6,351 relationships (4,115 firms with supplier information). Factset both covers more firms and identifies about *four times more* suppliers per firm than Compustat. This comparison demonstrates that Factset has a broader coverage on supply chain relationships compared to Compustat.

Note that all identified suppliers in Compustat are publicly traded firms[3] while Factset also contains privately held suppliers. The inclusion of the private firms allows us to construct the network metric and study the tier-2 sharing pattern based on the comprehensive multi-tier supply network. That said, to study the performance dependency between firms linked through supply chain relationships, we need companies' financial and operational characteristics, which are unavailable for privately held firms. If we restrict our attention to publicly traded firms, there exist other supply chain relationship databasets that are more comprehensive than Factset, such as the Bloomberg dataset described in Section 3.3.1.

The advantage of the Factset database is that it allows for a panel view of supply chain including privately held firms with a reasonable amount of data collection effort. Based on the multi-tier supply chain constructed from Factset, we examine the supply chain structure over time and construct the network metric that quantifies the degree of tier-2 sharing. In the next chapter, we use Bloomberg data to study the performance correlation in linked firms, because Bloomberg provides a broader coverage of supply chain relationships involving only public traded firms. The collection of Bloomberg data is labor intensive and is performed completely manually. As a result, we focus on the multi-tier supply chain of one particular sector instead of the entire economy. We are going to use the results from this chapter to motivate the choice of sector that we decide to focus on in the further performance analyses.

2.2 Empirical Findings

With the Factset longitudinal supply chain relationships data, we study 1) the heterogeneity of the extent of tier-2 sharing across different sectors, and 2) the stability of supply chain with regards to the tier-2 sharing pattern. To characterize the degree of tier-2 sharing, we first propose a network metric that quantifies the overlaps in a firm's tier-2 suppliers.

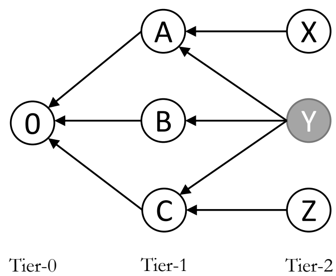
2.2.1 Network Metric

We measure the extent of tier-2 supplier sharing as follows. Consider the simple two-tier supply network depicted in Figure 2.1, in which Firm 0 has three tier-1 suppliers, labeled A,

[3] Compustat data is based on the reporting of major customers of U.S. listed firms. Therefore, all the suppliers inferred from the customer reporting are publicly traded firms in the United States.

B, and C, and three tier-2 suppliers, labeled X, Y, and Z. Note that supplier Y is shared by all tier-1 suppliers, whereas suppliers X and Z are not shared. The degree of sharing of a tier-2 supplier is reflected by the number of paths that connect the tier-0 firm to the tier-2 supplier. In this figure, three paths connect the tier-0 firm to the tier-2 firm Y, yet only one path (each) connects the tier-0 firm to X and Z. A tier-2 supplier is considered shared if more than one path links a tier-0 firm to that tier-2 supplier.

Figure 2.1: Illustration of common tier-2 supplier with degree of tier-2 commonality $k = 3$.



We now define the pair-wise degree of commonality as the number of paths that link every pair of tier-0 firm and tier-2 supplier. To obtain an aggregate measure of degree of commonality for a tier-0 firm, we take the average of the pair-wise degree of commonality of all pairs of the target tier-0 and its tier-2s. Let matrix A denote the binary customer-supplier relationship where A_{ji} indicates whether firm j supplies to firm i. The degree of commonality of firm i can be represented as $DC_i = \sum_j [A^2]_{ji} / \sum_j \mathbb{1}_{[A^2]_{ji} > 0}$. In Figure 2.1, the first-order and second-order adjacency matrices are

$$A = \begin{array}{cc} (j,i) & \begin{array}{ccccccc} 0 & A & B & C & X & Y & Z \end{array} \\ \begin{array}{c} 0 \\ A \\ B \\ C \\ X \\ Y \\ Z \end{array} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array}, \text{ and } A^2 = \begin{array}{cc} (j,i) & \begin{array}{ccccccc} 0 & A & B & C & X & Y & Z \end{array} \\ \begin{array}{c} 0 \\ A \\ B \\ C \\ X \\ Y \\ Z \end{array} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{2.1}$$

respectively. The aggregate degree of commonality of the tier-0 firm can thus be computed as $(1 + 3 + 1)/3 = 1.66$, which indicates that an average tier-2 supplier of the focal tier-0 firm is shared by 1.66 tier-1 suppliers. Thus, a degree of commonality equal to 1 indicates no sharing through tier-1 suppliers, whereas a value greater than 1 indicates the existence of sharing through the corresponding number of tier-1 suppliers.

2.2.2 Degree of Tier-2 Sharing by Sectors

We first investigate the extent of tier-2 sharing, measured as the degree of commonality, for firms in different sectors. Due to the position of industry in the whole economy, the tier-2 sharing level of firms can be inherently higher for certain sectors. We focus on the manufacturing industry in this analysis, and within the industry, we compare the degree of commonality for firms across more refined sector groups. We use the Standard Industrial Classification (SIC) to identify firms in the manufacturing industry. SIC is a commonly used classification system that uses four-digit code to represent a firm's major businesses. We categorize all firms with SIC Code 2000-3999 as manufacturing firms. The degree of commonality for all firms in the manufacturing industry is 1.38.

Table 2.1: Tier-2 Sharing Pattern across Manufacturing Sectors in 2012

Sector	Degree of Commonality	Pct. of Firms with Tier-2 Sharing$\geqslant$5	Pct. of Firms with Tier-2 Sharing$\geqslant$10
Food and Tobacco (SIC 2000 - 2199)	1.29	2.01%	0.36%
Textiles and Lumber (SIC 2200 - 2599)	1.27	1.94%	0%
Paper and Printing (SIC 2600 - 2799)	1.23	3.41%	0%
Chemicals and Petroleum (SIC 2800 - 3099)	1.38	5.20%	1.55%
Stone and Leather (SIC 3100 - 3299)	1.21	1.23%	0%
Primary and Fab. Metals (SIC 3300 - 3499)	1.23	2.72%	0.19%
Industrial Machinery (SIC 3500 - 3599)	1.55	7.65%	2.29%
Electronics (SIC 3600 - 3699)	1.46	6.26%	2.47%
Transportation Equipment (SIC 3700 - 3799)	1.44	8.80%	3.94%
Instruments (SIC 3800 - 3899)	1.34	2.35%	0.39%
Miscellaneous Mfg. (SIC 3900 - 3999)	1.34	3.67%	0.92%

Table 2.1 demonstrates the heterogeneity of tier-2 sharing across different sectors. In the table, we present the average degree of commonality, the percentage of firms with Tier-2 sharing $\geqslant 5$, and the percentage of firms with Tier-2 sharing $\geqslant 10$. We find that firms in the industrial machinery sector and the electronics sector have a higher degree of commonality compared to other manufacturing sectors. On the contrary, firms in the stone and leather sector, the primary and fabricated metal sector and the paper and printing sector have a lower degree of commonality. One potential explanation is that the three sectors are relatively upstream sectors. Tier-2 suppliers of firms in the upstream sectors are more likely to be commodity raw material providers. With the high interchangeability of commodity products, it is less likely for tier-1 suppliers to source from the same tier-2 suppliers, and thus the degree of commonality for firms in those sectors is lower.

Schmidt and Raman (2015) collect the number of announced disruptions, compiled from the press releases distributed via the PRNewswire and Business Wire from January 1, 1998 until December 31, 2011 (cited in Table 2.2). In general, we find that the sectors with a

Table 2.2: Disruption Announcements

Sector	Num. of Disruption Announcements (Schmidt and Raman 2015)
Food and Tobacco (SIC 2000 - 2199)	*15*
Textiles and Lumber (SIC 2200 - 2599)	23
Paper and Printing (SIC 2600 - 2799)	*7*
Chemicals and Petroleum (SIC 2800 - 3099)	*91*
Stone and Leather (SIC 3100 - 3299)	*3*
Primary and Fab. Metals (SIC 3300 - 3499)	*39*
Industrial Machinery (SIC 3500 - 3599)	*27*
Electronics (SIC 3600 - 3699)	*66*
Transportation Equipment (SIC 3700 - 3799)	*30*
Instruments (SIC 3800 - 3899)	*43*
Miscellaneous Mfg. (SIC 3900 - 3999)	*6*

higher degree of commonality (i.e., larger than 1.35) or a higher percentage of firms with tier-2 sharing (i.e., 5% of the firms have tier-2 suppliers with sharing $\geqslant 5$) also happened to report a greater number of disruption announcements.

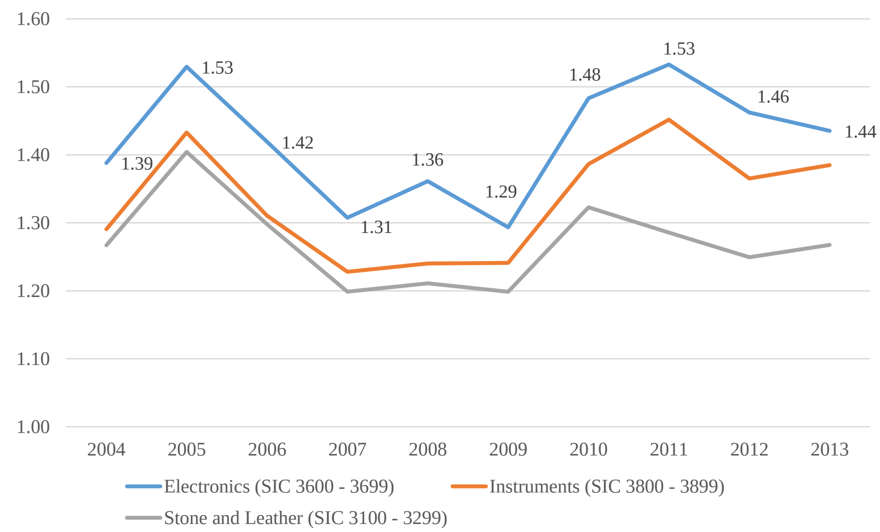
The correlation we observe here suggests that the extent of tier-2 sharing may influence the disruptions experienced at the focal tier-0 firm. We acknowledge that there are factors other than the tier-2 sharing pattern that can influence the supply disruptions, such as the geographical location of the suppliers, the inventory level at facilities, etc. Based on the empirical findings, we further validate the conjecture that the extent of tier-2 sharing influence the tier-0 firm risk in the next Chapter.

2.2.3 Stability of Supply Chain Structure

With the rapid change in the global economy, it is plausible that not all supplier-customer relationships persist for long periods of time. We use the Factset longitudinal relationship data to assess the stability of supplier-customer relationships over time. In particular, we investigate both 1) the length of each relationship and 2) the value of the sub-tier network metric, degree of commonality.

We find that the average length of a supply chain relationship in Factset is 1.5 years. This estimate is smaller than the duration of supplier relationships documented in the industry reports, potentially due to the following two reasons. Firstly, some supply chain relationships formed before 2003 and others continue beyond 2015. The length of the relationships thus got truncated because of the time frame of the dataset. In addition, if firms stop the voluntary disclosure of some of their sourcing activities, Factset would list the relationships as ended, since the data center cannot validate the continuation of relationships.

Figure 2.2: Degree of Commonality for Manufacturing Firms Across Years



We now examine the stability of the tier-2 sharing pattern. Figure 2.2 shows that the average degree of commonality for firms in the electronics sector, the instruments sector and the stone and leather sector from 2004 to 2014.[4] Taking the electronics sector as an example, the value of the degree of commonality ranges from 1.29 to 1.53 across the 10 years and starts to stay at a relatively stable level since 2010. In the 2010s, the degree of commonality for electronics firms is around 1.48, implying that an average tier-2 supplier of an electronics firm provides material inputs to 1.48 tier-1 suppliers of the focal tier-0 manufacturer. The trend in the degree of commonality over time also holds if we choose to focus on the firms in other manufacturing sectors. Generally, the extent of tier-2 sharing is high in 2005, and it decreases to a relatively low level as in 2007-2009. The extent of tier-2 sharing then bounces back starting from 2010 and stays at this relatively high level in the following years.

2.3 Conclusion and Discussion

Using the longitudinal supplier-customer relationship dataset from Factset, we construct the multi-tier supply chain, which allows us to examine the extent of tier-2 supplier sharing across industry sectors. Based on the constructed supply network, we find that tier-2 sharing is prevalent in the manufacturing industry. In particular, the supply chain for firms in the industrial machinery sector and the electronics sector contains more tier-2 sharing compared

[4] We select the three sectors based on their ranking of the degree of commonality in 2012. The electronics sector is one of the sectors that have the highest degree of commonality, i.e., 1.46 in 2012; the stone and leather sector is one of the sectors that have the lowest degree of commonality, i.e., 1.21 in 2012; the instrument sectors has the median level of degree of commonality, i.e., 1.34 in 2012.

to firms in other manufacturing sectors. On the sector level, we also find that the extent of tier-2 sharing is positively correlated with the number of supply disruption announcements.

In the next chapter, we test the performance correlation between firms linked through supplier-customer relationships and study the role of tier-2 sharing in the firms' performance dependency. As discussed in Section 2.1, we collect supply chain relationships data for one sector from Bloomberg due to the data collection limitation and the broader coverage of relationships between public traded firms.

We choose to focus on the high-tech sector, the electronics sector based on the SIC system. Based on our empirical findings using Factset supply chain relationship data, the high-tech sector is among the top few sectors that have a high degree of tier-2 sharing. According to Schmidt and Raman (2015), the high-tech sector also has the second largest number of disruption announcements from 1998 to 2011 among all manufacturing sectors. The large sample of impactful supply incidents allows us to statistically identify the impact of tier-2 sharing on the firm's financial performance. In addition, the hyper-competitive nature and rapid growth rate of the sector make firms more vulnerable to supply risks (Taylor 2002). The now famous Nokia Ericsson case (Eglin 2003) illustrates how a fire at a supplier's plant reshaped the European mobile phone market. Ericsson's slower response to the disruption, compared to Nokia, resulted in a loss in sales of approximately $400 million within a quarter after the disruption. Ericsson lost three percent market share to Nokia only six months after the incident. Short product life cycle, high demand variability, and aggressive competition make the high-tech sector a particularly interesting test-bed for the study of supply chain risks.

CHAPTER 3

Risky Suppliers or Risky Supply Chains? An Empirical Analysis of Sub-tier Supply Network Structure on Firm Risk in the High-Tech Sector

Although past research on supply chain risk management has focused on immediate supply chain connections, propagation of risks can extend beyond a firm's direct linkages. The structure of sub-tier supply network may also aid or prevent such risk propagation. In this chapter we focus on a specific aspect of sub-tier network structure, the sharing of tier-2 suppliers, and empirically study its prevalence and quantify its impact. Using firm-level supplier-customer relationship data in the high-tech industry, we find on average 20 percent of tier-2 suppliers are shared by tier-1 suppliers. We also find tier-0 firm risk is positively associated with common tier-2 supplier risk. The association is stronger with a higher degree of commonality. To disentangle the effect of risky supply network structure from risky tier-2 suppliers, we define two network metrics, viz., diamond ratio and cosine commonality score. Both metrics evaluate the extent of tier-2 supplier sharing within a firm's sub-tier supply network. The diamond ratio is constructed based on the binary supplier-customer relationships; the cosine commonality score also takes into account the firm's relative dollar spending on each of its suppliers. We find that a 10 percent increase in either metric is associated with around 5 percent increase in tier-0 firm risk. Lastly, using a new source of risk event data, we find firms experience significantly negative abnormal returns when their tier-2 suppliers are located in the event impact area, even though they themselves are not. The magnitude of this impact is much larger when the impacted tier-2 suppliers are heavily shared, similar to the scale of directly impacted firms, though taking longer to materialize. Overall our results reveal existence of substantial supply chain risks due to sub-tier supplier overlapping and highlight the need for firms to increase visibility into their extended supply network.

3.1 Introduction

The last decade has amply demonstrated that competitiveness of firms critically depends on the supply chains they orchestrate; examples include Dell, P&G, Zara, Walmart, and Toyota (*Shining Examples, The Economist*). Clearly the numerous players in a firm's supply chain are interconnected through the flow of materials and information. This connectivity, however, also extends to financial performance. For example, it has been suggested that equity and insolvency risk of supply chain partners represent sources of external risks to the interested firm (Cohen and Frazzini 2008, Menzly and Ozbas 2010, Hertzel et al. 2008). While this stream of literature focuses primarily on quantifying the impact of *direct* supply chain relationships, it is unclear whether or not the *sub-tier* suppliers are associated with a (focal) firm's financial performance.

The potential impact of sub-tier suppliers has only recently caught the attentions of supply chain professionals; for example, the 2011 Japan earthquake and tsunami and Thailand flood revealed substantial risks originating from previously unknown sub-tier suppliers (Brennan 2011). Not only does the lack of visibility into sub-tier suppliers put firms at greater risk, but analyzing the extent and impact of such risks also poses a major challenge for scholars. In this chapter, by constructing the extended supply chain network in the high-tech industry, we examine the following questions: (1) Is the financial risk of a firm associated with that of its *sub-tier* suppliers? (2) Is the financial risk of a firm associated with its *sub-tier supply network structure*? (3) Are such associations causal?

On the one hand, sub-tier suppliers may not present a significant risk to a focal firm for at least two reasons. First, sub-tier suppliers are located farther away in the supply chain. Hence, the reach of their impact may be limited, perhaps due to risk mitigation efforts taken by firms along the extended supply chain. Second, in the presence of a large number of sub-tier suppliers (e.g., hundreds, thousands, or even more), idiosyncratic risks originating from these sub-tier suppliers may, according to the law of large numbers, cancel out as they propagate down the supply chain. On the other hand, sub-tier supplier risk could indeed represent a source of neglected or underestimated risk, for several reasons. First, the fact that firms do not have direct business relationships with sub-tier suppliers and typically have very little visibility of them limit their capability to assess the size of the risk and therefore take direct and effective actions to mitigate it. The annual Supply Chain Resilience Surveys from 2009 to 2016 consistently reveal that most firms do not have full supply chain visibility.[1] Second, even with a large number of sub-tier suppliers, the law of large numbers may not

[1] The annual Supply Chain Resilience Surveys can be retrieved from the Business Continuity Institute website. http://www.thebci.org/index.php/businesscontinuity/cat_view/24-supply-chain-continuity/33-supply-chain-continuity/140-bci-resources

hold due to interdependencies between these suppliers and, as a result, their risks may not necessarily cancel out. For example, Acemoglu et al. (2012) illustrates that idiosyncratic shocks of individual firms or sectors can aggregate to systematic fluctuations through the physical network of input-output linkage.

Not only is it important to know who the sub-tier suppliers are, the structure of the network that binds them together could also be important. For example, when many tier-1 suppliers share a common tier-2 supplier, the supply base potentially has a single-point of failure. In the real business world, the structure of the sub-tier supply base appears to be a mystery to many firms. "*We thought our supply chain was pyramid shaped, but it turned out to be barrel-shaped,*" said a Toyota Motor Corporation spokesman (Brennan 2011) after realizing the extensive sub-tier sharing in the firm's supply chain.

How sharing of sub-tier suppliers impacts a firm's risk is also unclear. Sub-tier concentration may increase or decrease the risk to which the focal firm is exposed. On one hand, common tier-2 suppliers in multiple tier-1s' subnetworks will create risk inter-dependence among the tier-1 suppliers. Such interdependence will negatively impact the effectiveness of firms' existing risk mitigation strategies. For example, Yang et al. (2012) show that co-dependence between immediate supplier risks reduces the diversification benefit of dual-sourcing. Masih-Tehrani et al. (2011) show that in a multi-source supply chain, ignoring interdependence of supplier risks will lead to buyers' underestimation of inventory cost and overestimation of fill rate. On the other hand, sub-tier concentration may reduce other types of risks. For example, being a common tier-2 supplier implies that it will likely receive high-volume orders and, more importantly, from a more diversified customer base, which helps ensure healthy cash flows and business continuity (Balakrishnan et al. 1996). Such volume of business also permits the supplier to invest in innovation and to achieve a higher level of efficiency (Galbraith 1968, Kamien and Schwartz 1975). As a result, the focal firm may benefit from reduced risks from a more concentrated sub-tier supply base.

A major challenge in studying the effect of sub-tier suppliers and that of sub-tier network structure is the lack of relevant sub-tier supplier data. Previous empirical studies in supply chain management typically use the *sector-level* US input-output table of material flows (e.g., Cachon et al. 2007, Menzly and Ozbas 2010). The most commonly used *firm-level* supply-customer relationship data source is Compustat. Though Compustat offers longitudinal data, this database systematically under-sampling major suppliers for those small customer firms. Moreover, Compustat also under-represents international suppliers, because only US listed firms are required to report their major customers. However, over the past decade, firms have increasingly relied on global suppliers to take advantage of lower input costs and geographical skill specialization (Hausman et al. 2005). Due to these reasons, we rely on a new data source, Bloomberg, whose new Supply Chain Function maps 35,000 firms with their suppliers and customers. This data is more comprehensive than Compustat. For S&P 500 high-tech firms, for instance, the total number of suppliers, international and domestic, identified by Bloomberg is on average *seven times* larger than that identified by Compustat.

Using Bloomberg, we collect both domestic and international supply chain relationships and characterize the structure of a global multi-tier supply network for the high-tech sector. Our choice of the high-tech sector was informed by the presence of a higher frequency of supply shocks with substantial financial impacts in this sector. Our supply network contains 4,874 firms, including 2,427 high-tech firms, 2,447 non-high-tech suppliers, and 14,866 directional links. Using this dataset, we study the nature, structure, and influence of tier-2 supplier commonality on the performance of focal (also known as tier-0) firms.

First, we document the prevalence of overlapping sub-tier suppliers in the high-tech sector. Specifically, we find that on average 20 percent of tier-2 suppliers are shared by two or more tier-1 suppliers and 2 percent of tier-2 suppliers are shared by at least five tier-1 suppliers.

Next, we quantify the association of the tier-2 supplier risk and the total equity risk of the tier-0 firm, which is measured as stock return volatility, as in Hendricks and Singhal (2005b). We also consider idiosyncratic risk to eliminate the effect of systematic risk due to common risk factors. First, we find a positive yet small association between an average tier-2 supplier (idiosyncratic and total) risk and tier-0 firm (idiosyncratic and total) risk. However, the magnitude and the significance of this association increases with the degree of commonality, measured as the number of tier-1 suppliers who share a tier-2 supplier. In particular, when a tier-2 supplier is shared by five or more tier-1 suppliers, a 10 percent increase in tier-2 supplier total risk is associated with 1.66 percent increase in tier-0 firm total risk. Similar results are observed when we focus on idiosyncratic risk, but with a slightly smaller magnitude.

Third, we examine whether a tier-0 firm's risk comes from connectivity with risky tier-2 suppliers, or from having a risky supply network structure that embeds heavy tier-2 sharing. To characterize the degree to tier-2 commonality, we propose several metrics, viz., diamond ratio and cosine commonality score, and isolate the effect of sub-tier network structure on firm risk from the effect of risky sub-tier suppliers. We find that a 10% increase in the diamond ratio or the cosine commonality score is associated with around 5% (0.35 standard deviation) increase in a tier-0 firm's total equity risk, while controlling for average tier-1 and tier-2 supplier risks, market risk, and various firm-specific financial and operational characteristics. The effect of tier-2 commonality on a tier-0 firm's *idiosyncratic* risk remains similar, if not stronger.

Lastly, one should be cautious in interpreting the volatility co-movement as the causal impact of tier-2 suppliers' risks on tier-0 firm risk. It can be caused by supply risks propagating downstream, demand risks propagating upstream, or both. To establish the causal link of risk propagation from tier-2 suppliers to tier-0 firms, we collect *exogenous* supply shocks and study the post-event market reactions. Our approach avoids the potential disclosure bias implicit in the common approach that collects supply chain disruptions through firms' announcements (e.g., Hendricks and Singhal 2005b, Schmidt and Raman 2015), because we collect the original risk incidences, which may or may not have an impact on firms and their extended supply chains *ex-ante*. We find that firms with tier-2 suppliers located within close proximity to a event location experience significantly negative abnormal returns following the event, even though the firms themselves are not directly impacted. Moreover, the magnitude of the impact is much larger when the impacted tier-2s are heavily shared (by five or more tier-1 suppliers). This effect is on par with the magnitude when the firm is directly impacted, but it takes longer to materialize.

In summary, this study represents the first attempt to empirically study the nature and impact of sub-tier supply network structure on firm risk. Our results offer important insights for firms and their supply chain managers regarding the existence, importance, and management of sub-tier supplier risks. First, our results reveal that risks originating from sub-tier suppliers do propagate to tier-0 firms, despite being only remotely connected, highlighting the need for firms to increase visibility into their extended supply network. As reflected by the 2016 Supply Chain Resilience Survey, 66% of organizations do not have full visibility of their supply chains and 40% do not even analyze the source of supply chain disruptions, due to lack of direct business relationship with sub-tier suppliers and tier-1 suppliers' reluctance to disclose information (Grimm 2013). Our results thus reveal a potential source of unmanaged or poorly managed supply chain risk driven by sub-tier network structure. Second, even when firms complete sub-tier supply chain mapping, continuously monitoring sub-tier suppliers and updating their risk profiles still involve extensive arduous efforts from tier-0 firms and their tier-1 suppliers. Our results offer guidance on how to prioritize such efforts. Specifically, firms should identify critical sub-tier suppliers shared by multiple immediate suppliers, prioritize the monitoring of such suppliers, and manage their risks more effectively. Lastly, the sub-tier commonality metrics that we propose in this chapter, i.e., the diamond ratio and cosine commonality score, can be readily applied by firms to enhance their existing supply chain risk index, dynamically track changes in sub-tier network structure, and benchmark themselves against industry standards.

3.2 Literature Review

Despite the rich theoretical literature in supply chain risk management, empirical research in this area is relatively sparse. Early empirical research has focused on the impact of disruption announcements on a firm's performance, mostly due to lack of relevant supply chain data. Hendricks and Singhal (2003, 2005a,b) pioneered this line of research by using event studies to quantify the negative effect of supply chain disruptions on a firm's stock price. They measured change in financial performance using abnormal stock returns and return volatility around the announcement date of a supply chain disruption. Subsequent studies extend this line of work to analyze how firm characteristics and actions mitigate the impact of disruption risks. Hendricks et al. (2009) find that greater operational slack and lower geographic diversification reduces the impact of disruptions. Schmidt and Raman (2015) find that actions to improve operational efficiency have different impacts on firms facing distinct types of disruptions. Importantly, this line of research focuses on publicly announced supply chain disruptions without regard to its sources. Our contribution is to study how a firm's embedded network structure, especially its sub-tier suppliers, facilitate or prevent the propagation of risk.

More recently, researchers have found empirical support for shock transmission through inter-firm and inter-sector linkages. Menzly and Ozbas (2010) show evidence of cross-predictable returns between economically linked industries. The industry-level supply network is induced from the input-output matrix reported by U.S. Bureau of Labor Statistics. With more granular firm-level supplier-customer relationship data from Compustat, Cohen and Frazzini (2008) document the return predictability of principal customers on a supplier firm. Hertzel et al. (2008) find that a firm's bankruptcy filing makes its suppliers as well as its customer and supplier industries experience abnormal returns. Jain et al. (2013) use transaction-level Import/Export data to build a one-step relationship between US and overseas firms and study how a shift to global sourcing affects firm's inventory investment. Osadchiy et al. (2015) demonstrate that a more dispersed customer base is associated with higher systematic risk (i.e., higher correlation of sales with the state of economy) using both sector-level input-output tables as well as firm-level supply chain relationship data. Serpa and Krishnan (2017) demonstrate the productivity spillover from customer firms to supplier firms via various channels. Bray et al. (2016) work with facility-level automotive supply chain relationship data and find that larger inter-factory distances decrease the product quality. This stream of work demonstrates strong correlation between the performances of a firm and its immediate suppliers and customers, but their scope is limited to the direct business relationships. We contribute to the literature by studying firm's association with remote tier-2 suppliers.

Besides the linkage itself, the structure of the supply network also matters to the firm. The social network literature posits several measures of network structure that have been used to study risk propagation and its impacts. In particular, the most commonly used ones are degree centrality (Wu and Birge 2014), eigenvalue centrality (Acemoglu et al. 2012, Ahern 2013, Wu and Birge 2014) and information centrality (Bellamy et al. 2014). While commonly used in the social network literature, they characterize the importance of a node in the global network, but not the sub-network relevant to each specific firm. In particular, they do not capture how a pair of nodes (i.e., a tier-0 firm and any of its tier-2 suppliers) are connected locally, for example, via what paths. Therefore, they cannot be used to analyze the association of tier-0 firm risk with their sup-tier suppliers' risks. In order to do so, and informed by the recent initiations of sub-tier supply chain mapping in the industry (Sáenz and Revilla 2013), we propose new metrics that capture exactly how tier-0 firms and their sub-tier suppliers are connected. As we will show, these metrics significantly impact how risks propagate in a firm's extended supply network, and they can be easily applied by individual firms for risk monitoring and benchmarking purposes.

3.3 Data

We collected supplier-customer relationship data from a new data source, Bloomberg, a privately held financial software, data and media company. In this section, we introduce the data and the procedure we used to construct the sample for analysis. We then summarize the data sample and evaluate the coverage of supplier information from Bloomerg and compare it to Compustat, the most commonly used data source for firm-level supply chain relationships.

3.3.1 Supplier-Customer Relationship Data

We obtain information regarding global high-tech firms and their suppliers from Bloomberg, which established its new database of supplier-customer relationships using multiple sources. One source, similar to Compustat, relies on the SEC requirement that all US listed firms disclose their customers who comprise greater than 10% of annual revenues.[2] Bloomberg's database supplements the SEC dataset with information that firms disclose in a variety of

[2]Even though the 10% revenue cutoff ensures the inclusion of major customers, this does not necessarily imply the inclusion of major suppliers, because the revenue percentage cutoff is no explicitly relate to the cost percentage cutoff. As we will demonstrate later, Compustat indeed under-represents large suppliers; and given that the SEC reporting applies only to US listed firms, international suppliers are also under-represented in the Compustat database.

Table 3.1: Sample data from Bloomberg.

Name	Country	Market Cap	Sales Surprise	% Revenue	Relationship Value	Account As Type	%Cost	Source	As Of Date
SAMSUNG ELECTRON	South Korea	200.78B	-0.0123						
APPLIED MATERIAL	United States	21.74B	-0.0404	0.2	435.95M	CAPEX	0.0804	*2012A CF	11/15/2012
JEONGMOON INFO	South Korea	42.04M	N.A.	0.26	1.57M	COGS	0.0001	2013C3 CF	11/14/2012
IMARKET KOREA INC	South Korea	818.83M	-0.0151	0.3012	163.18M	SG&A	0.0142	*2013C2 CF	7/26/2013

Notes. This table displays sample of supplier data available on Bloomberg. It lists three of Samsung Electron's suppliers. Note that these suppliers' market capitalizations and relationship values are presented in dollars. For each identified relationship, Bloomberg reports the relationship value, % revenue and % cost to the supplier and the customer, the data source, and the as of date for the relationship. Specifically, % revenue lists the supplier's revenue the firm makes up and % cost lists the firm's cost that supplier represents.

media, such as annual and quarterly reports, conference call transcripts, capital markets presentations, sell-side conferences, company press releases, company websites, etc. (Davenport 2011). Using data gleaned from these sources documented in several languages, the Bloomberg database offers a comprehensive view of firm-level global supply chain relationships.

The Bloomberg dataset currently comprises supply chain relationships of 35,000 companies, and reports the source and date for each of these identified relationships. For each relationship, it also categorizes the nature of the product or service accounted for by the customer firm in the following four buckets: cost of goods sold (COGS), capital expenditure on long-term assets (CAPEX), research and development (R&D), and sales, general and administrative (SG&A). If available, Bloomberg quantifies the supplier-customer relationship value in dollars. For each *quantified* relationship, it also specifies percentage of the customer firm's cost that the supplier represents (% Cost) as well as percentage of that supplier's revenue the customer firm makes up (% Revenue). Table 3.1 lists sample data obtained from Bloomberg.

3.3.1.1 Data Summary

We retrieve supplier information for publicly traded high-tech firms from Bloomberg terminals. Recall that the database classifies the nature of the supplier relationship into four categories: COGS, CAPEX, R&D and SG&A. We first exclude relationships other than COGS, because our study focuses on risk aggregation resulting from repeated business relationships between a customer and its suppliers.[3] We then restrict our attention to publicly traded suppliers because further empirical analysis requires knowledge of financial and op-

[3]Repeated frequent interactions between customers and suppliers are necessary to allow us to observe the correlation between firms' financial risks. We thus focus our attention on supply chain relationships that are categorized as "COGS." We believe SG&A relationships are less likely to link two firms' quarterly risks. For example, stock return of a facility support supplier such as Microsoft is not likely to affect the customer's return. The other two types of supply chain relationships, CAPEX and R&D, are generally buyers' long-term investments. With the long duration of the investment, we do not expect to see the association between the stock performance of supplier and customer in a short time horizon.

erational characteristics of firms, unavailable for private firms.

We use Global Industry Classification Standard (GICS) to identify firms in the high-tech sector. GICS is an industry taxonomy developed by Morgan Stanley Capital International (MSCI) and Standard & Poor's (S&P) for the global financial community and is available on Bloomberg terminals. We categorize all firms in the information technology sector (GICS Code: 45-) as high-tech firms. The definition of GICS code is generally consistent with that of North American Industry Classification System (NAICS), the commonly used industry code in US market research.[4] We note two-thirds of suppliers of the high-tech firms also belong to the high-tech sector. In addition, many high-tech firms source from firms in three other industries, personal appliance (GICS code: 252010), electrical equipment (201040), and machinery (201060). Therefore, to ensure accuracy and representativeness of our two-tier supply network, we collect supplier information of firms in these industries as well.

The resulting supply network consists of 4,874 firms, including 2,427 high-tech firms and 2,447 non-high-tech firms who serve as suppliers of high-tech firms, and 14,866 directional links indicating supplier-customer relationships. The 2,427 high-tech firms account for about 76 percent of the total market capitalization of the sector; the remaining firms tend to be small international firms. If we focus on quantified relationships, the *quantified* supply network consists of 4,253 firms and 13,482 *quantified* directional links indicating supplier-customer relationships. We focus on the quantified relationships in the main analyses and conduct additional analyses based on both quantified and unquantified relationships in the robustness test (see Section 3.5.3 and Appendix 3.5.3.2 for details).

We now evaluate the quality of Bloomberg data relative to the commonly used supply chain relationship data source, Compustat. We compare supplier coverage of these two data sources with regard to the number of suppliers identified and purchase percentage quantified. To conduct a fair comparison, and because all Compustat relationships are quantified, we only consider quantified relationships from Bloomberg.

We find that for all S&P500 firms in the high-tech sector, Bloomberg on average identifies *four times more* US suppliers and *seven times more* global suppliers than Compustat, as illustrated in Table 3.2. On average, Bloomberg identifies 4.26 suppliers for a high-tech firm in our sample, whereas Compustat identifies only 0.24 suppliers, as shown in Table 3.3. In terms of percentage purchase, Bloomberg identifies on average 17.6% of purchase cost for a high-tech firm in our sample, whereas Compustat only identifies 0.45% of purchase. In order to calculate percentage purchase cost of each customer firm that a supplier represents,

[4] We also obtain the 8-digit GICS sub-industry code from Compustat North America and Compustat Global to cross-validate the code we obtain from Bloomberg. 98% of firms have consistent GICS codes. We only label a firm as a high-tech one if its GICS codes is associated with the information technology sector in both of the two databases.

we need to separate labor expense from COGS, following Serpa and Krishnan (2017) as described in Appendix 3.8.1.

Table 3.2: Suppliers identified by Bloomberg and by Compustat.

Name	GICS	Market Capitalization ($Mn)	GVKEY	Ticker Symbol	# of Suppliers identified in Compustat	# of US Suppliers identified in Bloomberg	# of Suppliers identified in Bloomberg
Facebook	451010	80,175	170617	FB	3	5	10
eBay Inc.	451010	55,800	114524	EBAY	0	4	5
Yahoo Inc.	451010	23,464	62634	YHOO	2	20	25
		53,146			**2**	**10**	**13**
Microsoft Corp.	451030	247,930	12141	MSFT	12	49	93
Oracle Corp.	451030	153,645	12142	ORCL	4	14	19
Salesforce.com	451030	23,036	157855	CRM	0	6	6
		141,537			**5**	**23**	**39**
Cisco Systems	452010	105,483	20779	CSCO	20	89	118
QUALCOMM Inc.	452010	102,851	24800	QCOM	3	9	17
Motorola Solutions Inc.	452010	15,248	7585	MSI	10	71	120
		74,527			**11**	**56**	**85**
Apple Inc.	452020	442,008	1690	AAPL	10	51	120
EMC Corp.	452020	52,375	12053	EMC	9	17	18
Hewlett-Packard	452020	49,967	5606	HPQ	33	94	187
		181,450			**17**	**54**	**108**
Grand Average		**67,538**			**9**	**36**	**62**

Notes. This table lists the number of suppliers reported in Compustat and Bloomberg databases. We include all suppliers reported in Compustat in 2009 and later. To simplify the table, we include the top three (highest market capitalization) SP500 firms that stay in the listed four sub-industries.

The significant differences in number of suppliers identified and percentage purchase quantified demonstrate that Bloomberg is a much more comprehensive source of supplier information than Compustat. We recognize that, even though Bloomberg is able to substantially increase supply chain visibility, it still mostly represents large and significant suppliers. Almost inevitably, due to the proprietary nature of such information, information regarding small and insignificant suppliers will be under-represented. Nevertheless, we believe that Bloomberg data provides us a unique opportunity to study sub-tier supply networks. In Appendix 3.5.3.1, we systematically investigate the existence of potential coverage bias and conduct tests to ensure that our conclusions are not driven by such biases.

3.3.2 Firm Characteristics Data

To analyze the association between a tier-0 firm's risk with its suppliers' risks, we also need to properly control for the effects of firm characteristics on its own risks. We therefore retrieve quarterly financial and operational data for high-tech firms and their suppliers. The data include market capitalization, financial leverage, return on assets, book to market ratio, days in inventory, and inventory growth. We discuss in detail why we focus on this set of control variables in Section 3.4. Summary statistics of firm characteristics are displayed in Panel B in Table 3.3. To ensure representativeness of the supply chain relationship data, we focus our study on the period from 2011 to 2013, because the supply chain relationship data

Table 3.3: Summary Statistics

		Mean	Standard Deviation			N	
Variable			Overall	Between	Within	# Firms	# Obs
Panel A: Supply Chain Relationships							
# Suppliers	*count*	4.26	17.02			2,427	
# Quantified Suppliers	*count*	4.21	13.02			2,214	
COGS Identified	*%*	8.83	15.48			2,214	
Purchases Identified	*%*	17.60	31.26			951*	
Purchases Identified (US firms)	*%*	22.00	35.32			401*	
Panel B: Firm Characteristics							
Log Market Capitalization	*$Mn*	5.61	1.99	2.00	0.24	4,089	29,614
Financial Leverage (FL)	*ratio*	0.41	0.23	0.23	0.05	4,107	29,284
Return on Asset (ROA)	*ratio*	2.25	9.63	9.46	4.37	3,809	27,613
Book to Market (BTM)	*ratio*	0.96	0.73	0.72	0.23	4,062	29,563
Days in Inventory (DII)	*days*	81.16	59.99	61.19	14.92	3,383	24,547
Inventory Growth(INGR)	*%*	9.65	36.07	25.34	28.33	3,583	25,652

Notes. The summary statistics are for the firms in the more inclusive unquantified supply network. The summary statistics for the firms in the quantified supply network are generally similar.
* Not all firms have employee data and/or compensation data available.

collected are mostly reported during fiscal year 2012 and are verified by Bloomberg as valid as of 2013 Q4. To verify the stability of supply chain relationships from 2011 to 2013, we conduct additional analyses using alternative data sources and confirm that our results are also consistent under a short 2-year study period (see Appendix 3.8.2 for details).

3.4 Empirical Model

In this section, we first propose several metrics to measure the degree of tier-2 commonality. We then discuss how we measure firm risks. Lastly, we present the empirical model used to analyze the association of a firm's risk with its tier-2 suppliers' risks as well as with the degree of tier-2 commonality.

3.4.1 Measures of Tier-2 Commonality

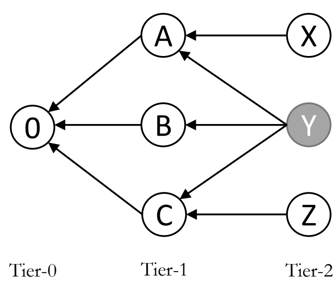
We introduce three metrics of tier-2 commonality : 1) diamond ratio and 2) cosine commonality score. The first metric is based on the binary customer-supplier relationship, while the second metric is based on customer-supplier relationship weighted by the percentage of purchase costs that a supplier represents.

3.4.1.1 Diamond Ratio:

In the previous chapter, we use the degree of commonality to characterize the degree to tier-2 commonality. While degree of commonality is an intuitive measure indicating the number of tier-1 suppliers that share a tier-2 supplier, the measure is influenced by the number of tier-1 suppliers that a firm has. That is, a tier-0 firm with more tier-1 suppliers is more likely to have a higher degree of commonality. This may introduce a bias between the degree of commonality and firm size, because a large firm with a higher number of identified immediate suppliers will tend to have a higher degree of commonality. To address this potential issue we propose another metric, the diamond ratio.

Let matrix A denote the binary customer-supplier relationship where A_{ji} indicates whether firm j supplies to firm i. The diamond ratio of firm i is defined as $DMD_i = DC_i/\sum_j \mathbb{1}_{A_{ji}>0} = \sum_j [A^2]_{ji}/(\sum_j \mathbb{1}_{A_{ji}>0} \cdot \sum_j \mathbb{1}_{[A^2]_{ji})}$, which normalizes the degree of commonality with the size of the tier-1 supply base. Specifically, the diamond ratio of each tier-0 firm is obtained by dividing the degree of commonality by the number of tier-1 suppliers.

Figure 3.1: Illustration of common tier-2 supplier with degree of commonality $k = 3$.



Note. You may refer to matrices 2.1 for the first-order and second-order adjacency matrices for this illustration example.

This metric also has an alternative intuitive interpretation. One can view the diamond ratio as the number of observed tier-0 to tier-2 paths over the number of all possible paths in a firm's supply network. Note that the number of all possible paths is precisely the product of the number of tier-1 suppliers and the number of tier-2 suppliers. For example, the diamond ratio of the tier-0 firm in the supply network depicted in Figure 3.1 equals $5/(3 \times 3) = 0.56$. By definition, the diamond ratio can only take a value between 0 and 1, and a higher value indicates the presence of more common tier-2 suppliers.

3.4.1.2 Cosine Commonality Score:

Our second measure, the cosine commonality score, considers cost-weighted supplier-customer relationships. First, we define the cost percentage matrix C where C_{ji} denotes the percentage of firm i's purchase cost attributed to supplier j. Note that the binary matrix A which indicates whether firm j supplies to firm i satisfies $A_{ji} = \mathbb{1}_{C_{ji}>0}$. The rows of C and A are indexed by supplier firms, and the columns are indexed by customer firms. If the supplier-customer relationship data are complete, the column sum of C should be 1. Using matrices C and A, we define the Cosine Commonality Score (CCS) of firm i as

$$\text{CCS}_i = \underset{j \neq m, A_{ji}=A_{mi}=1}{\text{median}} \cos(C_{\cdot,j}, C_{\cdot,m}) = \underset{j \neq m, A_{ji}=A_{mi}=1}{\text{median}} \frac{\langle C_{\cdot,j}\ , C_{\cdot,m} \rangle}{\|C_{\cdot,j}\|_2\ \|C_{\cdot,m}\|_2}$$

where $C_{\cdot,j}$ is the jth column of C, representing firm j's spending on its own suppliers. $\cos(C_{\cdot,j}, C_{\cdot,m})$ represents the pair-wise cosine similarity between the cost distributions of tier-1 suppliers j and m. Cosine similarity is a common metric used in the social network literature describing how similar two vectors are. The value of the cosine similarity between the spending vectors of any two firms ranges from 0 to 1, where 0 indicates that the two suppliers have no shared supplier (the spending vectors of the two firms are orthogonal), and 1 indicates that the two suppliers have the exact same supply base: same suppliers and same spending (the spending vectors of the two firms are identical). For example, if tier-1 supplier j single sources from firm A while tier-1 supplier m equally sources from firm B and C, the spending vectors of the two tier-1 suppliers on the union of their supply base $\{$A, B, C$\}$ are [1, 0, 0] and [0, 1/2, 1/2]. In this case, the cosine similarity between the cost distributions of tier-1 supplier j and m equals zero. If tier-1 supplier j instead sources equally from firm A and B, the spending vector of j becomes [1/2, 1/2, 0]. Now the cosine similarity between tier-1 supplier j and m equals to $\sqrt{1/2}$, suggesting the existence and the scale of tier-2 sharing.

After obtaining the cosine similarities, we aggregate the pair-wise measure for a focal tier-0 firm over all pairs of its tier-1 suppliers. We choose median (rather than mean) among all the pair-wise cosine similarities because of the high skewness of the distribution of cosine similarities.[5] Similar to the diamond ratio, a higher value of the cosine commonality score suggests the presence of more tier-2 sharing in the focal tier-0 firm's supply network.

3.4.2 Risk Measure

3.4.2.1 Firm Risk

We measure firm risks using stock return volatility, often referred to as firm total equity risk, and use it as one dependent variable. It is measured by the variance of the rate of return of the firm's equity over a certain time period. In this study, we use a three-month time window. We choose to analyze firm risk transmissions at quarterly frequencies to account for potential delays in market reactions. The total equity risk of firm i in calendar quarter t is defined as, $\text{VOL}_{it} = \text{Var}(R_{id}), d \in t$, where R_{id} represents daily return on day d. Note that the variance

[5] The results from using the average cosine similarity as the network metric are similar.

of equity return has been widely used as a measure of firm total financial risk in previous research (e.g., May 1995, Guay 1999, Hendricks and Singhal 2005b).

Asset pricing models suggest that a firm's total equity risk is affected by both systematic risks and idiosyncratic risks. Systematic risks refer to risks that cannot be eliminated by diversification. For example, all equities take on certain levels of market risk that, albeit different, are non-diversifiable. Idiosyncratic risk refers to the risk that can be avoided through a diversified portfolio. For example, the risk of a plant shut-down due to floods can be mitigated by investing in firms located in non-flood prone regions. An issue with a firm's specific supplier is typically considered as the firm's idiosyncratic risk. We therefore follow the finance literature and use the Fama-French three factor model (Fama and French 1993) to separate out idiosyncratic risk from the total risk, and use idiosyncratic risk as an alternative dependent variable in the subsequent analyses.

Factor Model To obtain the quarterly idiosyncratic risk, we first regress a firm's stock returns on the daily global factors over the three-year horizon. This is similar to the first step in FamaMacBeth regression (Fama and MacBeth 1973), which is used to determine the coefficients (βs) of systematic risks that a firm takes on. For each firm i, we regress

$$R_{id} = R_d^f + \alpha_i + \beta_i^m \cdot (\text{MKT}_d - R_d^f) + \beta_i^s \cdot \text{SMB}_d + \beta_i^v \cdot \text{HML}_d + \epsilon_{id}, \qquad (3.1)$$

where R_{id} represents the daily stock returns of firm i, R_d^f represents the daily risk-free rate, MKT_d represents the daily global market return factor, SMB_d represents the return premium of small firms over large firms in terms of market capitalization, and HML_d represents the returns premium of value stocks over growth stocks. We follow the literature and use the U.S. Treasury bill rate as the daily risk-free rate for firms in all countries (Ang et al. 2009). MKT_d, SMB_d, and HML_d are also the same for all the firms, regardless of the firm's trading market. The estimated $\beta_i = [\beta_i^m, \beta_i^s, \beta_i^v]$ gives firm i's association with the three common risk factors. The residual ϵ_{id} is firm i's daily residual return.[6] We define a firm's idiosyncratic risk, idioVOL_{it}, as the variance of the daily residual returns over a calendar quarter and use

[6]Besides the Fama-MacBeth regression, finance researchers alternatively measure the idiosyncratic risk based on a daily rolling factor model. To obtain the residual ϵ_{id} of day d, we first estimate Equation 3.1 using daily returns over the past 90 days prior to day d. At least 20 active trading days in the past quarter is required to obtain the estimated $\hat{\alpha}_i$ and $\hat{\beta}_i = [\hat{\beta_i^m}, \hat{\beta}_i^s, \hat{\beta}_i^v]$. The residual of day d is then computed as $\epsilon_{id} = R_{id} - R_d^f - \hat{\alpha}_i - \hat{\beta_i^m} \cdot (\text{MKT}_d - R_d^f) - \hat{\beta}_i^s \cdot \text{SMB}_d - \hat{\beta}_i^v \cdot \text{HML}_d$. The idea here is for investors to generate an expectation for the stock's association with the common risk factors using past data. The residual ϵ_{id} is then the excess return that cannot be predicted from prior knowledge. In this daily rolling case, the firm's idiosyncratic risk is again defined as the variance of ϵ_{id} over a calendar quarter. Note that compared to the Fama-MacBeth approach, the daily rolling approach leads to noisier measurement of the idiosyncratic risk, because each ϵ_{id} is derived after estimating Equation 3.1 over a different time period. We thus conduct the main analyses using Fama-MacBeth idiosyncratic risk. The results based on the daily rolling idiosyncratic risk are consistent, though with a smaller magnitude and significance.

calendar quarter instead of fiscal quarter to account for the potential variation of fiscal year end between businesses and countries. We retrieve daily stock prices (in local currency) of high-tech firms from the Compustat North America and Compustat Global database and compute the adjusted stock returns after controlling for dividends and splits. We obtain daily global factors from the Applied Quantitative Research (2016). The more commonly used Fama-French factors (Kenneth R. French Data Library 2016) are not available at a daily level for the global market, only available at monthly level. However, we tested the consistency of the two data sources for global factors by compounding the daily factors from the AQR Data Library into monthly factors and comparing these to Fama-French factors. We found that the two data sources are highly consistent, with a correlation of 0.998 for MKT_d, 0.856 for SMB_d, and 0.910 for HML_d.

One caveat for applying factor models in the global market: a standard approach for computing these factors has yet to be established. The currently available and most commonly used factor databases compute global factors based on developed markets only, excluding data from developing countries. We therefore conduct alternative analyses to allow for inclusion of a *country-specific* market factor when calculating the idiosyncratic risks. Please refer to discussions in Section 3.5.3 and Appendix 3.5.3.2 for details.

We end up with two risk measures, total equity risk and idiosyncratic risk, associated with each firm. A firm is included in the subsequent analyses as long as it has at least one non-missing quarterly risk for the study period.[7] We similarly construct the market risks for each trading market. All risk measures are winsorized at the 1% and the 99% levels.

We perform analyses for both the "local currency" case and the "USD" case. In the latter scenario, we convert stock prices measured in the local currency to those measured in US dollars using foreign exchange rates obtained from WRDS FX database. The "USD" case is considered a better reflection of firm financial risk because the internalized currency exchange risk is a factor that a firm should take into account when selecting suppliers (Min 1994). In Appendix 3.5.3.3, we conduct robustness checks using the risk measures in local currency to understand the influence of currency exchange risk in the risk propagation along the supply network.

3.4.2.2 Supplier Risks

To test the effect of suppliers' risks on the tier-0 firm's risk, we aggregate suppliers' risks by tier. We first follow Menzly and Ozbas (2010) to create portfolios of supplier firms and weight each supplier using a normalized cost percentage, from the perspective of a tier-0

[7] We have alternatively excluded firms that are not actively traded for the entire study period. The results from using the smaller sample of firms are consistent.

firm. For example, if a firm has only two identified tier-1 suppliers, and spends an equal amount between the two to acquire necessary inputs from them, we compute the tier-1 supplier risk as the average of the stock return volatilities of the two suppliers. For a tier-0 firm i, we let spl_VOL_{it} denote tier-1 supplier risk, and t2spl_VOL_{it} denote tier-2 supplier risk in quarter t.

$$\text{spl\_VOL}_{it} = \frac{\sum_j C_{ji} \times \text{VOL}_{jt}}{\sum_j C_{ji}}, \quad \text{and} \quad \text{t2spl\_VOL}_{it} = \frac{\sum_j [C^2]_{ji} \times \text{VOL}_{jt}}{\sum_j [C^2]_{ji}}. \tag{3.2}$$

C_{ji} represents the percentage of firm i's cost attributed to supplier j, and $\sum_j C_{ji}$ is the total percentage of purchase percentage of firm i identified. C^2 is the squared matrix of C, where $[C^2]_{ji} = \sum_k C_{jk}C_{ki}$ is the percentage of firm i's cost attributed to its tier-2 supplier j. In addition to the aggregate risk measures at each tier, we are particularly interested in risks originating from common tier-2 suppliers. Let $\text{comt2spl\_VOL}^k_{it}$ denote risks of those tier-2 suppliers that are shared by at least k tier-1 suppliers. It can be computed as the weighted risks of common tier-2 suppliers of firm i in quarter t with at least k degree of tier-2 commonality.

$$\text{comt2spl\_VOL}^k_{it} = \frac{\sum_{j:[A^2]_{ji} \geqslant k} [C^2]_{ji} \times \text{VOL}_{jt}}{\sum_{j:[A^2]_{ji} \geqslant k} [C^2]_{ji}}. \tag{3.3}$$

Recall that matrix A is the binary indicator matrix of supplier-customer relationship, namely $A_{ij} = \mathbb{1}_{C_{ij}>0}$. $[A^2]_{ji}$ is the pair-wise degree of commonality of supplier j in firm i's supply network, and $[A^2]_{ji} \geqslant k$ denotes the set of tier-2 suppliers that are shared by at least k tier-1 suppliers.

We similarly define the aggregated suppliers' idiosyncratic risks, spl_idioVOL_{it}, $\text{t2spl\_idioVOL}_{it}$, and $\text{comt2spl\_idioVOL}^k_{it}$. Note that purchase percentage may not necessarily represent the importance of the part being supplied, and thus weighting supplier risk by the normalized cost percentage may not be the correct way to aggregate the supplier risk. Alternatively, we construct un-weighted supplier risk using the arithmetic mean (instead of the percent-of-purchase weighted average) and conduct the subsequent analyses based on this un-weighted supplier risk in Appendix 3.5.3.2.[8]

[8] Although the percent of purchase cost may not signify the importance of the supplied part, we still believe that not all suppliers are equal. We therefore conduct robustness tests using alternative weighting schemes. Specifically, we compute supplier risks using firm size, measured as the log market capitalization, as weights. The results are still consistent, suggesting that the weights do not drive the final results we observe.

3.4.3 Other Independent Variables

Other factors can potentially influence firm risk. Firm size measured by the logarithm of a firm's market capitalization, and financial leverage measured by the ratio of book value of debt to the sum of the book value of debt and the market value of equity, are two important determinants of firm equity risk (Ben-Zion and Shalit 1975). Following Schmidt and Raman (2015), we also include firm profitability as a factor influencing firm risk. Although return on assets, operating margin, and gross margin are all measures of firm profitability, we only include return on assets in the regression due to collinearity. Our results are robust to alternative measures of profitability.

Firm equity risk is also affected by the amount of inventory a firm holds. The supply chain literature has long studied the role of inventory as a buffer against uncertain supply and demand (e.g., Ritchken and Tapiero 1986, Chen et al. 2007, Hopp et al. 2008). Using data from publicly listed US retailers, Alan et al. (2014) find that inventory productivity can predict future stock returns. Both inventory level (days in inventory) and inventory growth rate are included in our explanatory variables. Holding everything else constant, firms with higher inventory levels are better able to buffer supply and demand uncertainties. Following Hendricks et al. (2009), we use *industry-adjusted* days in inventory to account for different levels of normal inventory across sub-industries identified by the 8-digit GICS code. Inventory growth rate is associated with future firm performance, and hence firm risks; higher inventory growth rates may either indicate excess supply relative to realized demand or an expectation of faster growth. Apart from the inventory controls, we also include each firm's sales growth rate to account for its association with a firm's inventory decision (Gaur and Kesavan 2008). Inventory held by suppliers may actually reduce supply risks faced by their customers. We therefore also control for tier-1 suppliers' inventory level by taking the (un-)weighted average of each supplier's *industry-adjusted* days in inventory.

We control for the quarterly volatility of a firm's trading market. Separating out the market factor using the Fama-French factor model does not remove the association of a firm's return volatility and market return volatility (Herskovic et al. 2016). We obtain index performances for the ten largest trading markets of the high-tech sector. Firms in our data are traded in 60 different markets. We estimate a fixed effect using a small trading market dummy for the firms traded in the other markets — each contains less than 1% of firms in our sample. Multinational firms do not necessarily trade in countries of their main business locations, so we create dummy variables for firms' headquarter locations as well. A firm's stock performance can be influenced by the volatility intrinsic to a trading market and the level of economic uncertainties in the country of operation. The geographic distribution of firms' operating locations in our sample is as follows, United States (19%), Japan (18%), Taiwan (17%), Mainland China(14%), and South Korea (10%). We group firms operating in other origins as "other" in our regression analyses. Sub-industry dummies for every 8-digit GICS code are included as well to control for industry specific risks.

3.4.4 Model Specification

Our dependent variable is firm financial risk, measured as total equity risk, VOL_{it}, or idiosyncratic risk, idioVOL_{it}. Variables of interest are tier-2 supplier risk, common tier-2 supplier risk, and the measure of sub-tier network structure, diamond ratio (DMD) and cosine commonality score (CCS). Other independent variables include market risk (mkt_VOL_{it}), tier-1 supplier risk (spl_VOL_{it}), and financial and operational characteristics denoted by matrix X_{it}. X_{it} contains firm i's size, financial leverage, return on asset, book-to-market ratio, days in inventory, inventory growth rate, and supplier days in inventory in quarter t. Vector D_i contains the headquarter location dummies and the sub-industry dummies based on the 8-digit GICS code. We follow Engle and Ng (1993) to consider log-volatility and apply the log-log model here since both our dependent variable and the interested independent variables are volatility measures. Specifically, we use vol_{it} in lowercase to represent the natural logarithm of VOL_{it}.

$$\text{vol}_{it} = \alpha_0 + \alpha_1 \text{mkt\_vol}_{it} + \alpha_2 \text{spl\_vol}_{it} + X_{it}\gamma + D_i\phi + \epsilon_{it} \tag{3.4}$$

$$\text{vol}_{it} = \alpha_0 + \beta_1 \text{t2spl\_vol}_{it} + \alpha_1 \text{mkt\_vol}_{it} + \alpha_2 \text{spl\_vol}_{it} + X_{it}\gamma + D_i\phi + \epsilon_{it} \tag{3.5}$$

$$\text{vol}_{it} = \alpha_0 + \beta_2^k \text{comt2spl\_vol}_{it}^k + \alpha_1 \text{mkt\_vol}_{it} + \alpha_2 \text{spl\_vol}_{it} + X_{it}\gamma + D_i\phi + \epsilon_{it} \tag{3.6}$$

$$\text{vol}_{it} = \alpha_0 + \beta_3 \text{t2spl\_vol}_{it} + \beta_4 \text{DMD}_i + \alpha_1 \text{mkt\_vol}_{it} + \alpha_2 \text{spl\_vol}_{it} + X_{it}\gamma + D_i\phi + \epsilon_{it} \tag{3.7}$$

$$\text{vol}_{it} = \alpha_0 + \beta_3 \text{t2spl\_vol}_{it} + \beta_5 \text{CCS}_i + \alpha_1 \text{mkt\_vol}_{it} + \alpha_2 \text{spl\_vol}_{it} + X_{it}\gamma + D_i\phi + \epsilon_{it} \tag{3.8}$$

Equation 3.4 is the base model excluding variables of interest. Equation 3.5 adds the tier-2 supplier risk. We anticipate a positive β_1, which indicates a positive association between tier-2 supplier risk and the tier-0 firm's risk. Equation 3.6 estimates the association of common tier-2 supplier risk with tier-0 firm risk. We expect β_2^k to be positive and increase as the degree of commonality k increases. Equation 3.7 and 3.8 separate the effect of tier-2 supplier risk and sub-tier network structure on the tier-0 firm's risk. We again expect a positive coefficient for tier-2 supplier risks, β_3. We also anticipate a positive effect of the diamond ratio (DMD) and cosine commonality score (CCS). That is, we expect a supply network with a larger overlap among tier-2 suppliers to be associated with higher tier-0 firm risk. In addition to the direction, we are also interested in the scale of the coefficients.

To focus on the influence of suppliers on firms' idiosyncratic risk, we regress idioVOL$_{it}$ against the total equity risk of the tier-1 suppliers, tier-2 suppliers, and the sub-tier network measures. We also regress idioVOL$_{it}$ against the idiosyncratic risk of suppliers and the sub-tier network measures to prevent any correlation between firms' systematic risks being a factor in the regression analyses. Recall that in the Fama-French three factor model, the market factor MKT$_d$ is the same for all firms regardless of the firm's trading market. We thus include the market risk of each trading market in this second stage regression to control for systematic risks due to additional country-specific market changes that are not controlled in the three factor model. Note that the Fama-French three factor model has already accounted for the risk generated by variations in firms' market capitalization and book-to-market ratio. Controlling for these variables of firms' financial and operational characteristics in the second stage achieves the objective of isolating the effects of suppliers from those of the firms' own characteristics.

Three sets of estimates α_1, α_2, β_1, β_2^k, β_3, β_4, and β_5 are generated from the following *three* scenarios: 1) regressing total equity risk on suppliers' total equity risks; 2) regressing idiosyncratic risk on suppliers' total equity risks; and 3) regressing idiosyncratic risk on suppliers' idiosyncratic risks. Note that idiosyncratic risks are estimated rather than observed; therefore regression results involving idiosyncratic risk measures are noisier. The significance level of the latter two sets of estimates will also therefore be lower.

We estimate Equations 3.4, 3.5 and 3.6 using firm fixed effects models to account for potential unobserved firm characteristics, which can also be correlated with other covariates. For example, firms with higher supply risks are more likely to use flexible capacity, which in turn reduces risks. The use of firm fixed effect models eliminates the bias coming from time-invariant firm characteristics. Following the suggestion of Petersen (2009), we also include time fixed effects to capture the overall macroeconomic and industry trends and estimate the standard errors *clustered* on the firm dimension. To determine the coefficients of sub-tier network structure measures, namely the diamond ratio and cosine commonality score, fixed effects models are no longer valid because firm fixed effects will absorb the effect of the time-invariant sub-tier network structure. Instead, we use a random effects model with time dummies and estimate *firm-clustered* standard errors. To further control for the potential correlation in errors terms across firm dimension, we conduct a sensitivity test based on the method developed in Driscoll and Kraay (1998). The nonparametric covariance matrix estimator developed in the referenced paper adjusts for heteroscedasticity and very general forms of spatial (panel) and temporal (autocorrelation) dependence. Nevertheless, because this estimator is based on an asymptotic theory, Hoechle et al. (2007) comments that one should be cautious when applying this estimator to panels that contain a large cross-section but only a short time dimension. For this reason, we present the results with the Driscoll and Kraay standard errors in Appendix 3.5.3.2 rather than in the main text.

3.5 Empirical Results

We first demonstrate the prevalence of common tier-2 suppliers and then show how tier-2 supplier commonality is associated with tier-0 firm risk.

3.5.1 Tier-2 Commonality

Table 3.4 shows summary statistics of all three measures of tier-2 commonality, 1) degree of commonality, 2) diamond ratio, and 3) cosine commonality score. We find that the median degree of commonality of all tier-0 high-tech firms is 1.08, suggesting that more than half of the firms have common tier-2 suppliers in their supply networks.

Table 3.4: Statistics of tier-2 commonality measures.

	Degree of Commonality	**Diamond Ratio**	**Cosine Commonality Score**
Mean	1.175	0.318	0.068
Standard Deviation	0.256	0.166	0.152
.25 percentile	1.000	0.179	0.000
Median	1.083	0.323	0.001
.75 percentile	1.246	0.500	0.049
.95 percentile	1.705	0.558	0.432

Notes. Statistics are computed for high-tech firms that have at least two suppliers reported.

Using degree of commonality, we identify tier-2 firms that are heavily shared and also tier-0 firms whose supply network involves the most tier-2 sharing. Appendix Table 3.21 lists all tier-2 firms who are shared by a large number of tier-1 suppliers, i.e., greater than or equal to 20, in any tier-0 firm's network. Most of the heavily shared tier-2 suppliers are semiconductor companies. Numbers in the final column of the table represent the number of tier-0 firms that source from these tier-2 suppliers. Our data also indicates that these tier-2 suppliers are not necessarily immediate suppliers of the associated tier-0 firm, implying that without sub-tier visibility, firms may not realize that they rely on a particular set of sub-tier suppliers. For instance, 20 of Dell's tier-1 suppliers and 24 of HP's tier-1 suppliers source from Stats Chippac Ltd., which does not directly supply any of the S&P500 hardware manufacturers. Appendix Table 3.22 lists all tier-0 firms whose supply network relies on these heavily shared tier-2 firms. Most of them are in the Technology Hardware & Equipment industry group (GICS code: 4520x). The table also indicates that many tier-2 suppliers are shared in these firms' supply networks — 35% are shared by at least two tier-1 suppliers, and more than 10% are shared by at least five tier-1s.

3.5.2 The Association of Tier-2 Commonality with Firm Risk

Table 3.5 provides the estimates of the model that examines the association between the tier-0 firm risk and tier-2 supplier risk. Column (a) reports the estimates of β_1 in Equation 3.5, the association of tier-2 supplier risk with the tier-0 firm risk. The coefficients are significantly positive, yet not of large magnitude, in all three scenarios. This coefficient measures the average effect of tier-2 supplier risk and does not account for the heterogeneity among tier-2 suppliers. Because we are particularly interested in the effects of common tier-2 suppliers, we then estimate Equation 3.6 for different degrees of commonality, $k \leqslant 4$ and $k \geqslant 5$.[9] The estimates are shown in columns (b) and (c). We observe that both the magnitude and significance of the estimates are higher under higher degree of commonality. Examining the associations between 1) tier-0 firm idiosyncratic risk and tier-2 supplier total risk and 2) tier-0 firm idiosyncratic risk and tier-2 supplier idiosyncratic risk, reveals similar increasing patterns; however, the magnitudes are smaller, and significance levels are slightly lower, compared to the case with total risks. This is likely because idiosyncratic risks are noisier, as they are estimated rather than directly observed.

Table 3.5: The Association Between Tier-0 Firm Risk and Tier-2 Supplier Risk

CORRELATION BETWEEN FIRM RISK	(a) Tier-2	(b) Tier-2 shared by ≤ 4 Tier-1s	(c) Tier-2 shared by ≥ 5 Tier-1s
Firm Total Equity Risk and Supplier Total Equity Risk	0.038**	0.036*	0.166***
	(0.019)	(0.019)	(0.041)
Firm **Idiosyncratic Risk** and Supplier Total Equity Risk	0.031	0.029	0.086**
	(0.019)	(0.019)	(0.041)
Firm **Idiosyncratic Risk** and Supplier **Idiosyncratic Risk**	0.052**	0.050**	0.104**
	(0.020)	(0.020)	(0.051)

Notes. Standard errors adjusted for firm clusters are shown in parentheses. To simplify the table, we do not report on other controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

[9] We observe similar patterns in the results when using other cutoffs. However, we typically see that the coefficient of tier-2 suppliers first become statistically significant across all models at p-value$= 0.05$ when they are shared by at least 5 tier-1 suppliers. We therefore demonstrate the results using this cutoff.

To separate the effects of risky sub-tier suppliers and risky sub-tier supply network structure, we estimate Equation 3.7 using the two proposed measures of tier-2 commonality: diamond ratio and cosine similarity score. In Table 3.6, column (a) through column (c) display the estimates using the diamond ratio, while column (d) to column (f) report the estimates using the cosine commonality score. The coefficient of the diamond ratio is significant at 0.01 level, and it suggests that increasing the diamond ratio by one standard deviation results in a 6.4% or 0.45 standard deviation increase in tier-0 total equity risk. Similar results are found for the alternative measure, the cosine commonality score. The estimate is significant at 0.01 level as well, and one standard deviation increase in the cosine commonality score is associated with 4.6% or 0.32 standard deviation increase in tier-0 total equity risk. The inclusion or exclusion of tier-2 supplier risk in the model does not change the magnitude or the significance of the estimates of both sub-tier network measures, suggesting that the network measure is indeed orthogonal to tier-2 supplier risk.

We then estimate Equation 3.7 using idiosyncratic risks instead of total risks and present the results in Table 3.6. Columns (b) and (e) demonstrate the results from regressing tier-0 firm idiosyncratic risk on tier-1 and tier-2 total risks, measures of tier-2 commonality and the same control variables as before. Columns (c) and (f) demonstrate the results from regressing tier-0 firm idiosyncratic risk on tier-1 and tier-2 idiosyncratic risks for measures of tier-2 commonality and the same control variables as before. In all columns, the coefficients of diamond ratio and cosine commonality score are significant at a similar magnitude to the estimates obtained from using the total equity risk.

This set of results indicates that tier-0 firm risk is indeed positively associated with its tier-2 supplier risk, and more so when its tier-2 suppliers are heavily shared. The network structure that binds tier-2 suppliers affects tier-0 firm risk directly, regardless of the risk levels of the tier-2 suppliers themselves. Given firms' lack of visibility into sub-tier supply network, our results reveals a potential source of unmanaged or poorly managed supply chain risk driven by sub-tier network structure.[10]

3.5.3 Robustness Tests

In this section, we conduct multiple robustness analyses to ensure our results are not driven by data coverage, model or variable specifications, or alternative explanations.

[10]The propagation of supply chain risk does not necessarily require market visibility of the sub-tier supply network. Even if the market is unaware or not fully aware of the supply chain structures, the supply chain risk might not be an *ex-ante* priced risk; the market can still react to risk events *ex-post*, even without knowledge of the origin of the risk. Lack of visibility may delay market reactions, however (Hendricks and Singhal 2003, Wu 2015), which is partially why we choose to conduct the analysis at quarterly level.

Table 3.6: Regression of Firm Risk on Supplier Risk and Sub-tier Supply Network Structure

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio			Tier-2 & Cosine Commonality Score		
	(a)	(b)	(c)	(d)	(e)	(f)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	0.036**	0.028	0.049***	0.036**	0.028	0.049**
	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)
Diamond Ratio	0.635***	0.752***	0.758***			
	(0.157)	(0.164)	(0.164)			
Cosine Commonality Score				0.456**	0.475**	0.486**
				(0.202)	(0.215)	(0.215)
Market Risk	0.402***	0.323***	0.323***	0.402***	0.323***	0.323***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Tier-1 Supplier Risk	0.062***	0.063***	0.072***	0.061***	0.062***	0.071***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Financial and Operational controls	Yes	Yes	Yes	Yes	Yes	Yes
Country, Sub-industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,593	24,590	24,590	24,593	24,590	24,590
Number of Firms	2,122	2,122	2,122	2,122	2,122	2,122
Overall R-squared	0.1676	0.1514	0.1511	0.1690	0.1534	0.1531

Notes. Columns (a) and (d) regress total equity risk on suppliers' total equity risks; Columns (b) and (e) regress idiosyncratic risk on suppliers' total equity risks; Columns (c) and (f) regress idiosyncratic risk on suppliers' idiosyncratic risks. Standard errors adjusted for firm clusters are shown in parentheses. Market risk, tier-1 risk, and industry-adjusted (supplier) days in inventory are included with log transformation. Country and sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

3.5.3.1 Data Coverage Bias

Understanding what types of firms are more (or less) likely to have their supplier information identified in the dataset is critical to assessing the impact of potential coverage bias on the estimation results. In this section, we conduct tests to examine the existence and magnitude of potential coverage bias by country of origin, supply chain upstream and downstream positions (as reflected by sub-industry code), and firm size.

It is reasonable to expect that each firm in the high-tech sector sources from at least one supplier. We therefore use the percentage of having at least one supplier identified to evaluate what types of firms are more (or less) likely to have biases in their supplier coverage. Appendix Table 3.10 demonstrates how this measure varies across sub-industries (8-digit GICS code) and headquarter locations. In each cell, we report the number of firms and the percentage of firms with at least one supplier identified.

We find that firms operating in Taiwan as well as firms in the sub-industries of Technology, Hardware, Storage & Peripherals and Technology Distributors are more likely to have at least one supplier identified in the data; that is, these firms are likely to have more complete supplier information. The results are consistent when we change the cutoff to at least five identified suppliers, the average number of identified suppliers per high-tech firm (see Appendix Table 3.11), except that US firms have the leading percentage of identified suppliers. These patterns provide the basis for the sub-sample robustness tests to verify that our main results are not driven by the coverage bias.

We thus test our models on the following subsamples: 1) close-to-market firms — firms in the sub-industries of Technology, Hardware, Storage & Peripherals and Technology Distributors ; and 2) Taiwan or US firms.

Appendix Table 3.12 shows the subsample estimates for close-to-market firms and Appendix Table 3.13 shows the subsample estimates for Taiwan or US firms. Regardless of the risk measure, the effect of the diamond ratio and cosine commonality score is consistently estimated across all subsamples, and the estimates are significant in most of the cases. We note that the effect may vary across different sub-samples (either due to data coverage bias or heterogeneity), and we conclude that sub-tier network measure, in particular the diamond ratio, influences tier-0 firm idiosyncratic risk.

3.5.3.2 Model and Variable Specifications

- *Measures of idiosyncratic risks*. To account for market variations in developing countries, we conduct the following alternative analysis of the CAPM model using country-specific market returns. Note that the other two factors, SMB and HML, are not available for these developing countries. Specifically, we regress

  $$R_{id} = R_d^f + \alpha_i + \beta_i^m \cdot (R_d^m - R_d^f) + \epsilon_{id}$$

  The difference here is that we use R_d^m, the country-specific market returns, which are computed from daily index prices instead of the daily global market factor. This is the only factor we control in this regression. We can similarly define a firm i's idiosyncratic risk as the variance of ϵ_{id} over a calendar quarter. Our conclusions regarding the effect of tier-2 commonality on firm risk remain unchanged from using this alternative measure of idiosyncratic risk. Appendix Table 3.14 demonstrates the detailed results.[11]

- *Measures of sub-tier commonality*. In the main analyses, we consider the quantified supply chain relationships; however, it is unclear whether the percent of purchase cost provides additional information on how a supplier influences the customer firm. On

[11] The daily index prices (in local currency) of the ten largest trading market of high-tech firms, United States (SPX Index), Japan (NKY Index), Korea (KOSPI Index), Taiwan (TWOTCI Index), Mainland China (SHASHR Index), France (CAC Index), India (SENSEX Index), Hongkong (HSI Index), United Kingdom (UKX Index) and Germany (DAX Index), are retrieved from Yahoo Finance.

the one hand, one may expect a larger supplier, who represents a significant portion of the customer's purchase cost, to likely have a larger influence on the customer's performance. Partially due to the lack of a better measure to account for component criticality or the network location of a supplier, cost percentage has been used as a proxy for supplier significance in theoretical papers that study how economic fluctuations aggregate (Acemoglu et al. 2012) and empirical papers that study relationships between supplier and customer returns (Menzly and Ozbas 2010). On the other hand, supply chain problems have routinely been reported to be caused by suppliers who account for only a small portion of purchase cost yet constitute a major threat when impacted. For example, consequent to the Japan earthquake and tsunami in 2011, shortage of components from the shared sub-tier suppliers, like Renesas (chip supplier) and Merck (paint pigment supplier), caused months-long production delays for automotive manufacturers (Kyodo 2011, Sedgwick 2014). Despite the small cost percentage of these components, supply problems at Renesas and Merck propagated to manufactures owing to the component non-substitutability. Thus, the share of overall purchase cost may or may not be a major driver of supplier risks. We therefore also examine the results for the un-quantified supply chain relationships. Consistent results are obtained as shown in Appendix Table 3.15.

- *Measures of firm characteristics*. Models in the main analyses control for firm's financial and operational characteristic metrics, such as, financial leverage, size (log market capitalization), book to market ratio, return on assets, days in inventory, and inventory growth. Because the data sample includes both US and international firms, it is possible that the different accounting standard between countries may endow different meanings to the same metric. We therefore conduct two robustness checks: 1) estimating the equations without the panel firm-level financial and operational controls; and 2) estimating the equations using only US listed and US based firms. Results presented in Appendix Tables 3.16 and 3.17 confirm that variation in the accounting standard does not drive our main results.

- *Correlated error terms*. The fact that firms are connected through supply chain relationships makes the observations non-independent. In other words, in the econometric model (Equations 3.4 to 3.7'), error terms (ϵ_{it}) are not i.i.d across firm dimension; rather they can be correlated through supply chain relationships. Typically, the variance-covariance structure of the error terms does not affect the asymptotic consistency of the estimators; however, it does affect the accuracy of the estimation of standard errors. To adjust for such possibility, we follow the method developed in Driscoll and Kraay (1998) to correct the estimates of standard errors. Driscoll and Kraay developed a nonparametric covariance matrix estimator that adjusts for heteroscedasticity and very general forms of spatial (panel) and temporal (autocorrelation) dependence. This nonparametric technique produces consistent estimates of standard errors with no restrictions on the size of the time dimension or the size of the cross-sectional dimension – even if the number of panels is much larger than the number of time periods. We present the results of using the Driscoll and Kraay standard errors in Appendix Table 3.18. In most cases, this adjustment leads to smaller standard errors and higher statistical significance levels. However, the magnitudes of the coefficients of network measures can be smaller when using pooled OLS (to apply the Driscoll and Kraay adjustments) instead of a random effects model.

3.5.3.3 Alternative Explanations

- *Currency risks*. Global sourcing has become an increasingly popular business strategy in the past decades, due to discrepancy in inputs costs and skill specialization among countries (Hausman et al. 2005). In our data, on average about 75% of a firm's suppliers are international suppliers, based on the headquarter locations of the focal firm. The prevalence of global sourcing indicates that currency exchange rate risk is likely an important risk that many firms face. Our previous analysis, which converts stock prices measured in the local currency to those measured in US dollars, internalizes such risks for firms.

  However, one may be concerned about to the extent to which the association that we observe between firms' and their suppliers' risks are driven by currency exchange rate risks, which are generic to all firms in the local market, and to what extent the association is actually driven by idiosyncratic risks that are specific to these suppliers. To tease out the currency exchange rate risk, in this robustness test, we calculate market risk, firm risk and supplier risk using stock prices measured in local currencies instead of US dollars. We obtain consistent results as shown in Appendix Table 3.19.

- *Tier-1 supplier concentration*. Even though a firm typically does not have direct business relationships with its tier-2 suppliers, its sourcing decisions may *indirectly* affect its sub-tier network structure. For instance, firms whose tier-1 suppliers are located in geographical proximity are more likely to have common tier-2 suppliers. Similarly, firms that dual source or multi source are also more likely to have common tier-2 suppliers, because these dual- or multi-sourced tier-1 suppliers conduct similar businesses and produce similar products. Since the nature of these relationships are unknown to us (e.g., what parts or components are being supplied), we cannot obtain direct measures of dual or multi-sourcing. However, it is likely that suppliers in the same sub-industry are more likely to provide substitutable inputs to the focal firm, so the focal firm is more likely to have shared tier-2 suppliers. If indeed geographical or sub-industry concentration of tier-1 suppliers is correlated with the degree of tier-2 sharing, and if that is also correlated with risk exposures of the focal firm, our estimates of the influence of tier-2 sharing could be biased. Therefore, we test whether the effect of tier-2 supplier commonality can actually be explained away by tier-1 supplier geographical or sub-industry concentration.

We measure tier-1 geographical or sub-industry concentration using the Herfindahl index, $H = \sum_{i=1}^{N} p_i^2$, where N is the total number of countries of origins (or sub-industries) that the focal firm sources from, and p_i represents the share of the firm's tier-1 suppliers in the ith country (or sub-industry). Intuitively, a smaller Herfindahl index indicates a less concentrated, or in other words, more diversified tier-1 supplier base. We again estimate our models using both total and idiosyncratic risks. Appendix Table 3.20 confirms that the measured effects of diamond ratio and cosine similarity score are both robust after controlling for tier-1 geographic and industry concentration.

3.6 Event Study

In the previous section we demonstrate that a firm's equity risk is positively associated with its sub-tier suppliers' equity risks, and in particular, equity risks of heavily shared sub-tier suppliers. Building upon these results, this section provides further evidence about the causal link behind such risk interdependency, using new sources of individual risk event data.

Readers should be cautious about attributing the volatility co-movements documented in the previous section as the tier-2 suppliers' causal impact on tier-0 firms. In addition to supply shocks that propagate downstream, risks from a given node in the network can also propagate *upstream* in the form of demand shocks (e.g., Lee et al. 2000, Cachon et al. 2007, Bray and Mendelson 2012a, Osadchiy et al. 2015). Therefore, without pinpointing the exact origins of actual risk events, our documented volatility co-movement between tier-2 and tier-0 firms can be interpreted either as tier-2 suppliers' risks propagating (via overlapping tier-1 suppliers) down to tier-0 firms, or alternatively as tier-0 customers' risks propagating (also via overlapping tier-1 suppliers) back up to tier-2 suppliers, or both.

To address this potential reverse causality issue and establish the causal link on risk propagation from tier-2 to tier-0 firms, we need to not only clearly separate supply shocks from demand shocks, but also focus on idiosyncratic shocks that are *exogenous* in nature, i.e., shocks that are not correlated with either unobserved firm-level characteristics, or macroeconomic aggregates.

We identify a collection of such idiosyncratic shock events—specifically, geo-tagged, natural disasters and localized power outages caused by meteorological events—by working closely with a risk management data solutions provider that provides global risk monitoring services to many corporate and government clients. We match these risk events to our sample firms using a geo-matching algorithm, and use event studies to examine the following issues. Do firms directly impacted by these events subsequently experience negative abnormal returns? More importantly, do firms located *outside* the events' impact areas, but whose tier-2 suppliers may be located within the impact areas, also have negative abnormal returns? If so, does the magnitude of the market reaction vary by the degree of sharedness of the impacted tier-2s? The subsequent subsections provide details on our methodology, data sources, and results.

3.6.1 Methodology

For each firm-risk event pair that we capture in our sample, we examine the difference between the firm's observed equity return and its expected equity return (according to different risk models, discussed shortly) over a specified window of time following the event, i.e., the "event window." Following standard finance literature, we compute the expected return in two steps. We first fit the specified risk model to historical return and risk factor data, described in detail in Section 3.4.2.1, over an "estimation window" *prior* to obtaining estimates of the risk factor loadings of each stock. We then compute the daily expected returns for each day in the event window as the predicted values from the risk model, i.e., expected returns in the absence of the event. In particular, to ensure robustness of the results, we use both the one-factor market return model (CAPM) and the Fama-French Three-Factor model to calculate expected returns, as we did in the previous section. The daily abnormal return is then calculated as the difference between observed returns and the expected returns during the event window.

We implement this procedure as follows: For each firm, we define the event's first announcement date as Day 0, which may or may not be a trading day. The next trading day following the event date is Day 1, the trading day preceding the event date is Day -1, and so forth. We use 200 trading days (i.e., 40 trading weeks) prior to each event as the estimation window, and, following standard literature such as Brown and Warner (1985), we exclude the week (i.e., 5 trading days) leading up to the event to avoid overlapping with the actual event period. That is, the estimation window is defined as [-204, -5]. Sample firms with at least 30 non-missing returns in this period are included in our analysis.

To analyze the abnormal returns of firms located in the impact areas, we define two alternative event windows following Hendricks and Singhal (2003): (1) the first trading day after the event date, i.e., Day 1, and (2) the event day and the first trading day after the event date, i.e., Day 0 and Day 1. The two definitions of the event window account for variations of the actual event time: An event may happen during a trading day, or after the market closes on a trading day, or during a non-trading day. Next, to analyze the abnormal returns of firms located outside the impact areas, but with sub-tier suppliers located within, we study the abnormal return over a longer event window of five trading days (one week) after the event.

The finance literature usually uses one of two methods to compound the daily abnormal returns into an event-window abnormal return: the algebraic Cumulative Abnormal Return (CAR) and the geometric Buy-and-Hold Abnormal Returns (BHAR). CAR is the sum of the average daily abnormal returns during the event window. BHAR is calculated as the compounded abnormal returns during the event window. To ensure the robustness of our results, we use both approaches. Specifically, let T denote the event window. Let R^{obs}_{ikt} denote the observed return of firm i on day t following event k, and $\hat{R}^{\text{m}}_{ikt}$ denote the corresponding predicted return under model m, where $m = \{\text{CAPM}, \text{Fama-French Three-Factor Model}\}$. Then the daily abnormal return, AR^{m}_{ikt}, and aggregated abnormal return over the event period, CAR^{m}_{ik} and $BHAR^{m}_{ik}$ can be expressed as:

$$AR^{\text{m}}_{ikt} = R^{\text{obs}}_{ikt} - \hat{R}^{\text{m}}_{ikt},\ CAR^{\text{m}}_{ik} = \sum_{t \in T} AR^{\text{m}}_{ikt},\ BHAR^{\text{m}}_{ik} = \prod_{t=0}^{T}(1 + R^{\text{obs}}_{ikt}) - \prod_{t \in T}(1 + \hat{R}^{\text{m}}_{ikt})$$

After calculating the abnormal returns, we first verify that the events indeed have significant negative impacts on the returns of the firms located directly within the impact areas. We then examine whether this impact extends downstream to customers, and whether the magnitude of the impact indeed varies with the degree of sharedness of tier-2 suppliers.

3.6.2 Risk Event Data

To acquire the appropriate event data, we work with a third party risk management consultancy, which monitors and analyzes worldwide risk events that (1) threaten critical infrastructure, (2) interrupt business continuity or (3) affect safety and security. Being one of the early providers of risk monitoring and management tools, the company supplies real time event feeds to numerous Fortune 500 companies and government agencies (e.g., Department of National Homeland Security). The firm began tracking risks in 2013, which corresponds with our sample period.

Although each client would normally receive event feeds filtered by pre-specified criteria, we are able to obtain the unfiltered data, from which we retrieve 296 severe risk events that occurred in the calendar year of 2013. For each event, we obtain the geocoded location, the time when the event started, and the time when the event was resolved, whenever applicable. These events are classified into 14 categories, including geophysical (e.g., earthquakes), transportation (e.g., road closures), infrastructure (e.g., power outage), security (e.g. protest), labor (e.g., worker strikes), fire (e.g., industrial and resident fire), etc. To ensure the exogeneity of the events, we will focus on natural disasters as well as events caused by extreme weather conditions. Man-made events can be subject to endogeneity concerns; labor strikes and factory fires, for example, which may result from poor working conditions or inferior manufacturing practices, may also correlate with firm performance. We also focus our attention on those events that lasted multiple days (i.e., three days or longer) because these events are more likely to cause disruptions to regular business activities. The resulting sample of 45 events include 12 earthquakes with magnitude over seven and 33 power outages induced by extreme weather.

To match the event locations to firm locations, we obtained the headquarter location of each firm in our sample from the Bloomberg database. Ideally, we would also like to obtain facility locations in addition to the headquarter locations. However, to the best of our knowledge, such data is not currently available, especially for international firms. Many studies therefore use headquarter locations as a proxy (e.g., Barrot and Sauvagnat 2016), while noting that the measurement error is likely to bias the estimates against finding any effect. Moreover, events in our study, such as natural disasters and weather-induced power outages, that impact headquarter locations would also pose credible threats to business continuity, even if firms have other production facilities located elsewhere. We geocode the business addresses using Google API. Among the 4,874 firms in our sample, 3,167 can be geocoded at higher than zipcode accuracy. To avoid measurement errors when matched with the event locations, we exclude those companies whose locations can only be coded with a lower accuracy level. For each event, we define the impact area as within the 10 kilometer radius (i.e., 6.21 miles) of the geocoded location of the event. The results are consistent when we vary the definition of the impact area to a 5- or 20-kilometer radius. We identify 86 firms located in the impact areas of a total of 16 events (a total of 90 firm-event combinations), and 1,226 firms with tier-2 suppliers located in the impact areas (a total of 3,814 firm-event combinations). The details of these events are included in Appendix 3.8.5.

A key difference between our approach to identify risk events and those commonly used in the literature relates to disclosure bias. By construction, events identified through supply chain disruption announcements (e.g., Hendricks and Singhal 2003, Schmidt and Raman 2015) includes only those events that resulted in actual business disruptions confirmed by the announcing firm. By contrast, our data construction procedure avoids such disclosure bias, because we collect the original risk events, which may or may not have an impact to the firms and their extended supply chains *ex-ante*.

3.6.3 Results

We first examine whether our events indeed have significant negative impacts on firms located in the impact areas. We compute the event-window CAR and BHAR with both CAPM and Fama-French Three-Factor models, and report the results in Table 3.7. Our results confirm that these risk events lead to significantly negative abnormal returns for firms located within the impact areas, but do not have any significant effect on the valuation of firms located outside the areas. On the first trading day (Day 1) following the event, the average Three-Factor CAR, for example, is -0.153% (p-value $= 0.159$) for firms not located in the impact areas, and -1.618% (p-value $= 0.047$) for firms in the areas. We obtain similar results when we extend the event window to both the event day (Day 0) and the first trading day (Day 1), when we use the CAPM model to estimate returns, and when we use BHAR to measure the aggregate abnormal returns.[12]

Table 3.7: Event Study Results: Abnormal Returns for Firms In and Out of Impacted Areas

	Cumulative Abnormal Return (CAR)				**Buy-and-Hold Abnormal Return (BHAR)**			
Event Window:	Day 1		Day 0 + Day 1		Day 1		Day 0 + Day 1	
In Impact Areas:	No	Yes	No	Yes	No	Yes	No	Yes
Three-Factor Model	-0.153%	-1.62%**	-0.186%	-2.24%**	-0.153%	-1.62%**	-0.178%	-2.30%**
	(0.159)	(0.047)	(0.160)	(0.033)	(0.159)	(0.047)	(0.160)	(0.030)
One-Factor Model	-0.096%	-2.36%***	-0.171%	-0.176%**	-0.096%	-2.36%***	-0.146%	-0.909%**
	(0.331)	(0.001)	(0.257)	(0.032)	(0.331)	(0.001)	(0.444)	(0.024)

Note: If an event happens on a non-trading day, the corresponding firm-event combinations do not have the associated abnormal returns on the event day (Day 0). P-Value of the one-sided t test is included in parentheses and is shown as: * – 0.1 level, ** – 0.05 level, *** – 0.01 level.

Next, we examine if firms with sub-tier suppliers in the impact area also experience negative abnormal returns, even though they themselves are not in the area. We report the abnormal returns of these firms in columns 1-2, and 5-6 of Table 3.8. First, note that firms whose tier-2 suppliers are *not* located in the impact area, neither do themselves nor their immediate suppliers, do not experience significant changes in returns. The average CAR for the first week following the event is -0.042% (p-value $= 0.158$). However, for firms with tier-2 suppliers located in the area, but not themselves nor their immediate suppliers,

[12]We obtain similar results under matched samples, where firms are matched based on market size, book-to-market ratio, and previous stock performance.

the average CAR is much larger in both magnitude and statistical significance, i.e., -0.692% (p-value $= 0.004$). Again, the results are also consistent under the CAPM model, and when using BHAR instead of CAR.

Finally, for firms with sub-tier suppliers in the impact area, we examine whether the magnitude of the impact varies with the degree of sharedness of sub-tier suppliers in the area. We separately report the abnormal returns for firms with low (shared by four tier-1 suppliers or less) and high degrees of sharedness (shared by five tier-1 suppliers or more) in columns 3-4, and 7-8 of Table 3.8. The results indicate that the magnitude of the impact is even larger when the impacted tier-2 suppliers are heavily shared, with the average CAR of -2.140% (p-value $= 0.039$), as opposed to -0.660% (p-value $= 0.007$) when they are shared by fewer tier-1 suppliers (four or less).

This set of results highlights that *exogenous* supply shocks originated from sub-tier suppliers do propagate to tier-0 firms. More importantly, the magnitude of the impact on tier-0 firms is much greater when the impacted sub-tier suppliers are heavily shared, on par with that when the firm itself is directly impacted, though the effect takes longer to materialize.

Table 3.8: Event Study Results: Abnormal Returns w/ and w/o Tier-2 Suppliers in Impact Areas

	Cumulative Abnormal Return (CAR)				**Buy-and-Hold Abnormal Return (BHAR)**			
Tier-2 In Impact Areas:	**No**		**Yes**		**No**		**Yes**	
Tier-2 Sharedness:		All	Low	High		All	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Three-Factor Model	-0.042%	-0.692%***	-0.660%***	-2.14%**	0.005%	-0.690%***	-0.679%***	-1.15%**
	(0.158)	(0.004)	(0.007)	(0.039)	(0.159)	(0.001)	(0.003)	(0.048)
One-Factor Model	-0.186%	-0.967%***	-0.925%***	-2.92%**	0.077%	-1.12%***	-1.10%***	-1.81%**
	(0.262)	(0.001)	(0.001)	(0.011)	(0.293)	(0.000)	(0.001)	(0.013)

Note: Firms are excluded when they themselves or their immediate suppliers are located in the impact areas. The effect window is one trading week (5 days) following each event. Low sharedness refers to situations when the impacted tier-2 suppliers are shared by at most four tier-1 suppliers. High sharedness refers to situations when the impacted tier-2 suppliers are shared by five or more tier-1 suppliers. P-Value of the one-sided t test is included in parentheses and is shown as: * – 0.1 level, ** – 0.05 level, *** – 0.01 level.

3.7 Conclusion

Our results provide the first empirical evidence of how the sub-tier network structure is associated with total and idiosyncratic equity risk of a firm. Using a new global supplier-customer relationship dataset from Bloomberg, we are able to construct the supply network of the high-tech sector with substantively expanded coverage of suppliers for both domestic and international firms. This gives us greater visibility into the multi-tiered supply network and hence allows us to test the effect of sub-tier network structure on firm risk.

Based on the constructed supply network, we find that tier-2 supplier commonality is prevalent in the high-tech sector. On average, 20 percent of tier-2 suppliers are shared by two or more tier-1 suppliers and 2 percent of tier-2 suppliers are shared by five or more tier-1 suppliers. Such a network feature has an important implication for risk aggregation in supply networks. We find a strong positive association between the total equity risk of shared tier-2 suppliers and that of the tier-0 firm. The association increases when tier-2 suppliers are shared by more tier-1 suppliers. We observe similar results when we focus on the idiosyncratic risk.

We propose two network measures of tier-2 commonality, the diamond ratio and cosine commonality score, to isolate the influence of a risky supply network from that of risky suppliers. We find that 10% increase in the diamond ratio or the cosine commonality score is associated with 5% or 0.35 standard deviation increase in a tier-0 firm's total equity risk.

Furthermore, we provide evidence on the causal relationship of tier-2 suppliers' risks on the tier-0 firm's risk using new sources of *exogenous* risk event data. Following the event, firms with tier-2 suppliers in the impact area experience significantly negative abnormal returns, even though they themselves are not located in the area. The magnitude of this impact is more substantial when the impacted tier-2 suppliers are heavily shared.

Our study highlights the need for firms to increase visibility into their sub-tier supply network because of the significant supply chain risks they could impose. Furthermore, our results offer guidance on how to prioritize the efforts of sub-tier supplier risk management. Given knowledge of sub-tier supply network, firms should identify critical sub-tier suppliers shared by multiple immediate suppliers, prioritize the monitoring of such suppliers, and proactively manage the associated risks. Lastly, the sub-tier commonality metrics that we propose can be readily applied by firms to enhance their existing supply chain risk index, dynamically track changes in sub-tier network structure, and benchmark themselves against industry standards.

3.8 Complementary Material

3.8.1 Bloomberg Data Coverage: Percent of Purchases

Note that COGS includes not only the material purchases but also the direct labor expenses relating to the production process. Therefore, instead of using the percentage COGS, we provide a new metric to capture the coverage of Bloomberg data that takes care of the potential bias resulting from the included labor expense. We follow Serpa and Krishnan (2017) and approximate the labor expense by taking the product of the total number of employees and the sector-average labor cost. We obtain the total number of employees from Compustat and annual U.S. employee hourly compensation data by sector for U.S. firms in our sample from the Bureau of Labor Statistics (BLS). For international firms, such data is not available. However, the BLS reports hourly compensation of manufacturing employees for 34 countries through the International Labor Comparisons (ILC) program in 2011. Given the proximity of the program year and the study period of our data, we use this data source to approximate labor costs for the international firms in our sample. We assume 2,087 annual working hours amending 5 U.S. Code 5504(b).[13]

Using these data sources, we obtain estimates for labor costs at the firm level. Labor represents 17% to 70% of COGS, consistent with typical numbers identified in the literature. We then obtain an estimate of the firm's material purchases by subtracting out the labor expense from a firm's COGS in the same year. The percentage of purchases by suppliers is updated as the ratio of the supply chain relationship value over the estimated purchases of the target firm, whereas the original percent COGS is computed as the ratio of the supply chain relationship value over the target firm's total COGS.

The adjusted metric suggests that Bloomberg explains on average 17.6% of a firm's material purchases, about twice of the average percent COGS covered. For U.S. firms, it explains on average 22% of purchases. For larger firms (measured by market capitalization), the covered percentage is even higher. For example, our data covers 51.2% of Microsoft's purchases, 70.4% of IBM's purchases, and 75.2% of Qualcomm's purchases.

We also observe the following regarding approximation of labor costs. First, BLS only reports international hourly compensation for 34 countries, which covers 80% of the firms in our data sample. Second, the reported hourly compensation is at a sector-country level and does not account for the heterogeneity across firms. Third, Compustat only reports the total number of employees, which include both SG&A and manufacturing employees, whereas the hourly wage we obtained from BLS represents the average of only manufacturing work-

[13] 5 U.S.C. 5504(b) requires the hourly rates of pay for most Federal civilian employees to be computed using the 2,087-hour divisor from 1985.

ers. Finally, for international firms, wage data is only available for an average manufacturing worker, not necessarily in the high tech sector. According to the U.S. hourly compensation by sectors, a high-tech worker earns about 50% higher than an average manufacturing worker. Using manufacturing wage, it is likely that labor costs are under-estimated in our data. Therefore, the percent of purchases captured could be even higher for those firms.

3.8.2 Selection of Study Period

Bloomberg supply chain relationship data provide a snapshot of existing supplier-customer relationships. The supply chain relationships are mostly reported for the 2012 fiscal year and have been continuously updated afterwards based on new sources of information, indicating newly formed relationships or terminations of prior relationships. Specifically, each supply chain relationship listed on Bloomberg terminal is appended with a status variable "As of date." The "As of date" variable represents the most updated disclosure date of the relationship. Bloomberg only includes a relationship if the relationship is available in the current year as indicated by "As of date." Relationships are typically updated four times a year based on newly filed financial reports and news sources. The "As of date" of the retrieved relationship data concentrates in the last quarter of 2012 and the first quarter of 2013, as many of them are extracted from the companies' 2012 annual reports. Bloomberg continuously updates the relationship status based on new sources of information such as quarterly reports following 2012 annual reports, press releases, news items, etc. The relationship data thus present those that existed in 2012 and likely continued to exist through 2013 Q4 (the time of the authors' data collection) based on most updated information.

We use the Compustat longitudinal data to further assess the stability of supply relationships. Specifically, we investigate how likely is a relationship observed in 2012 to have existed in 2011, and how likely to have continued to exist in 2013. As described earlier, the SEC requires a U.S. listed firm to disclose its major customers that comprise more than 10% of the firm's revenue. Thirty years of time-series records are available through Compustat; we focus on the records of recent years because the level of stability of supply chain relationships are likely different now versus in earlier years (e.g., 1980s and 1990s). We thus retrieve Compustat segment data from 2009 to 2014. We followed the steps below to clean the relationship data: 1) we removed those relationships with non-identifiable customer names, e.g. "1 customer," "others"; 2) we unified customer names for firms with multiple common appellations, e.g. IBM and International Business Machines Corporation; and 3) we standardized company names, separated name suffixes such as Inc., Corp., and Ltd., and made the customer name case-insensitive.

After data cleaning, we end up with about 15,000 supply chain relationships across all industries, spanning five years. We find the average length of a relationship is 2.8 years, and around 65% of the relationships last for at least two years. This is an under-estimate of the average length of these supply chain relationships, because some relationships were formed before 2009 and others continue beyond 2014. This average length covers relationships across all industries; to focus on the supply chain relationships in the high-tech sector, we match customer names to firm identifiers (i.e., stock ticker) to identify the sector characterization under which a firm falls. We then focus on those supply chain relationships that include a customer in the high-tech sector (a comparable dataset to the Bloomberg data we acquired) and find that 77.0% of the high-tech supply chain relationships continue from 2012 to 2013 while 79.7% of the high-tech relationships in 2012 are inherited from 2011. For this reason, we consider the supply chain relationships observed in 2012 likely exist in adjacent years (one prior and one after). The three-year study period also limits the effect of common movements in firm volatility due to changes in firm size concentration (Kelly et al. 2013). To be conservative, we also conduct analyses on a two-year study period (2011Q3 to 2013Q2). All results are consistent as shown in Table 3.9.

Table 3.9: Alternative Study Period

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio			Tier-2 & Cosine Commonality Score		
	(a)	(b)	(c)	(d)	(e)	(f)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	0.017	0.004	0.030	0.017	0.004	0.030
	(0.020)	(0.020)	(0.022)	(0.020)	(0.020)	(0.022)
Diamond Ratio	0.443***	0.555***	0.560***			
	(0.159)	(0.167)	(0.167)			
Cosine Commonality Score				0.378**	0.393**	0.409**
				(0.186)	(0.201)	(0.201)
Market Risk	0.368***	0.298***	0.298***	0.368***	0.298***	0.298***
	(0.014)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)
Tier-1 Supplier Risk	0.073***	0.076***	0.082***	0.072***	0.075***	0.081***
	(0.014)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)
Financial and Operational controls	Yes	Yes	Yes	Yes	Yes	Yes
Sub-industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,434	16,434	16,432	16,432	16,432	16,432
Number of Firms	2,115	2,115	2,114	2,114	2,114	2,114
Overall R-squared	0.1969	0.1765	0.1760	0.1978	0.1778	0.1774

Notes. Columns (a) and (d) regress total equity risk on suppliers' total equity risks; Columns (b) and (e) regress idiosyncratic risk on suppliers' total equity risks; Columns (c) and (f) regress idiosyncratic risk on suppliers' idiosyncratic risks. Standard errors adjusted for firm clusters are shown in parentheses. Market risk, tier-1 risk, and industry-adjusted (supplier) days in inventory are included with log transformation. Sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

We acknowledge that no good data source is yet available to directly verify the length of each supply chain relationship. The stability analysis discussed above is generated using a different data set, that of the Compustat SEC filings, which captures only a subset of relationships found in Bloomberg. As we show in the Data Section, large and significant suppliers are likely under-represented in the Compustat database compared to the Bloomberg database. For this reason, the likelihood that a relationship identified in the Bloomberg database has existed in the adjacent years could be even higher than that estimated using the Compustat database.

Unstable supply chain relationships may create an attenuation bias. That is, if some relationships that we identify did not actually last for the three full years, there will be quarters in which the relationship no longer existed, making it more difficult to identify the association of risks between a focal firm and its immediate and sub-tier suppliers. In other words, the fact that we still find a statistically significant association indicates that the underlying association could be even stronger.

3.8.3 Result Tables for Robustest Tests

Table 3.10: Percentage of Firms with At Least One Supplier Identified by Headquarter Location and Sub-industry Code

	45102010	45201020	45202030	45203020	45203030	45301020		
Country\GICS	IT Consulting & Other Services	Communications Equipment	Technology Hardware, Storage & Peripherals	Electronic Manufacturing Services	Technology Distributors	Semiconductors	Other High-Tech Firms	All High-Tech
US	25	60	36	23	19	72	203	438
	(0.84)	(0.78)	(0.81)	(0.74)	(0.95)	(0.86)	(0.61)	(0.72)
Japan	56	14	26	4	51	14	216	381
	(0.82)	(0.93)	(0.85)	(0.50)	(0.98)	(0.93)	(0.78)	(0.82)
Korea	13	17	11	2	4	41	124	212
	(0.85)	(0.65)	(0.82)	(0.00)	(1.00)	(0.78)	(0.56)	(0.65)
China	22	39	19	3	4	35	111	233
	(0.77)	(0.74)	(0.95)	(1.00)	(1.00)	(0.89)	(0.66)	(0.75)
Taiwan	9	34	89	12	44	169	206	563
	(0.78)	(0.97)	(0.88)	(0.92)	(0.98)	(0.89)	(0.79)	**(0.86)**
Other	100	56	37	19	64	81	243	600
	(0.70)	(0.68)	(0.81)	(0.47)	(0.91)	(0.73)	(0.60)	(0.68)
All Locations	225	220	218	63	186	412	1103	2,427
	(0.76)	(0.78)	**(0.85)**	(0.67)	**(0.95)**	(0.84)	(0.67)	(0.76)

Notes. The number of firms in the related region and sub-industry is reported in each cell. Underneath, we report the percentage of firms that have at least one supplier identified in Bloomberg dataset.

Table 3.11: Percentage of Firms with At Least Five Suppliers Identified by Headquarter Location and Sub-industry Code

Country\GICS	45102010 IT Consult-ing & Other Services	45201020 Communi-cations Equipment	45202030 Technology Hardware, Storage & Peripherals	45203020 Electronic Manufacturing Services	45203030 Technology Distributors	45301020 Semicon-ductors	Other High-Tech Firms	All High-Tech
US	25	60	36	23	19	72	203	438
	(0.20)	(0.23)	(0.36)	(0.22)	(0.68)	(0.46)	(0.14)	**(0.26)**
Japan	56	14	26	4	51	14	216	381
	(0.13)	(0.14)	(0.27)	(0.00)	(0.43)	(0.43)	(0.11)	(0.18)
Korea	13	17	11	2	4	41	124	212
	(0.08)	(0.06)	(0.09)	(0.00)	(0.00)	(0.10)	(0.06)	(0.07)
China	22	39	19	3	4	35	111	233
	(0.09)	(0.10)	(0.26)	(0.33)	(0.25)	(0.14)	(0.05)	(0.10)
Taiwan	9	34	89	12	44	169	206	563
	(0.22)	(0.15)	(0.36)	(0.50)	(0.41)	(0.19)	(0.10)	**(0.21)**
Other	100	56	37	19	64	81	243	600
	(0.13)	(0.14)	(0.22)	(0.16)	(0.34)	(0.19)	(0.03)	(0.13)
All Locations	225	220	218	63	186	412	1103	2,427
	(0.13)	(0.15)	**(0.30)**	(0.24)	**(0.41)**	(0.23)	(0.09)	(0.17)

Notes. The number of firms in the related region and sub-industry is reported in each cell. Underneath, we report the percentage of firms that have at least five supplier identified in Bloomberg dataset.

Table 3.12: Subsample tests: Close-to-Market Firms

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio			Tier-2 & Cosine Commonality Score		
	(a)	(b)	(c)	(d)	(e)	(f)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	0.024	0.021	0.027	0.027	0.024	0.030
	(0.049)	(0.048)	(0.049)	(0.049)	(0.048)	(0.048)
Diamond Ratio	0.738**	0.838***	0.840***			
	(0.288)	(0.292)	(0.292)			
Cosine Commonality Score				0.588	0.566	0.578
				(0.387)	(0.379)	(0.380)
Market Risk	0.396***	0.328***	0.332***	0.397***	0.329***	0.333***
	(0.030)	(0.029)	(0.029)	(0.030)	(0.029)	(0.029)
Tier-1 Supplier Risk	0.071**	0.070**	0.067**	0.068**	0.067**	0.065**
	(0.031)	(0.030)	(0.028)	(0.031)	(0.030)	(0.029)
Financial and Operational controls	Yes	Yes	Yes	Yes	Yes	Yes
Country, Sub-industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,171	4,169	4,169	4,171	4,169	4,169
Number of Firms	360	360	360	360	360	360
Overall R-squared	0.2219	0.2063	0.2058	0.2252	0.2084	0.2080

Notes. Columns (a) and (d) regress total equity risk on suppliers' total equity risks; Columns (b) and (e) regress idiosyncratic risk on suppliers' total equity risks; Columns (c) and (f) regress idiosyncratic risk on suppliers' idiosyncratic risks. Standard errors adjusted for firm clusters are shown in parentheses. Market risk, tier-1 risk, and industry-adjusted (supplier) days in inventory are included with log transformation. Country and sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

Table 3.13: Subsample tests: Taiwan or US Firms

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio			Tier-2 & Cosine Commonality Score		
	(a)	(b)	(c)	(d)	(e)	(f)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	0.012	0.007	0.023	0.012	0.006	0.022
	(0.028)	(0.027)	(0.030)	(0.028)	(0.028)	(0.030)
Diamond Ratio	0.632**	0.722***	0.730***			
	(0.247)	(0.260)	(0.260)			
Cosine Commonality Score				0.558**	0.571**	0.581**
				(0.235)	(0.257)	(0.257)
Market Risk	0.239***	0.019	0.022	0.239***	0.019	0.022
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Tier-1 Supplier Risk	0.034*	0.041**	0.052***	0.032*	0.040**	0.051***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Financial and Operational controls	Yes	Yes	Yes	Yes	Yes	Yes
Country, Sub-industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,828	10,828	10,828	10,828	10,828	10,828
Number of Firms	929	929	929	929	929	929
Overall R-squared	0.2116	0.1821	0.1817	0.2137	0.1860	0.1856

Notes. Columns (a) and (d) regress total equity risk on suppliers' total equity risks; Columns (b) and (e) regress idiosyncratic risk on suppliers' total equity risks; Columns (c) and (f) regress idiosyncratic risk on suppliers' idiosyncratic risks. Standard errors adjusted for firm clusters are shown in parentheses. Market risk, tier-1 risk, and industry-adjusted (supplier) days in inventory are included with log transformation. Country and sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

Table 3.14: Robustness Test: CAPM model using country-specific market returns

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio			Tier-2 & Cosine Commonality Score		
	(a)	(b)	(c)	(d)	(e)	(f)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	0.036**	0.035*	0.055***	0.036**	0.035*	0.055***
	(0.018)	(0.019)	(0.020)	(0.018)	(0.019)	(0.020)
Diamond Ratio	0.635***	0.892***	0.898***			
	(0.157)	(0.176)	(0.176)			
Cosine Commonality Score				0.456**	0.526**	0.528**
				(0.202)	(0.241)	(0.241)
Market Risk	0.402***	0.304***	0.307***	0.402***	0.304***	0.307***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Tier-1 Supplier Risk	0.062***	0.052***	0.053***	0.061***	0.050***	0.051***
	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.011)
Financial and Operational controls	Yes	Yes	Yes	Yes	Yes	Yes
Country, Sub-industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,593	22,750	22,750	24,593	22,750	22,750
Number of Firms	2,122	1,960	1,960	2,122	1,960	1,960
Overall R-squared	0.1676	0.1700	0.1700	0.1690	0.1730	0.1730

Notes. Columns (a) and (d) regress total equity risk on suppliers' total equity risks; Columns (b) and (e) regress idiosyncratic risk on suppliers' total equity risks; Columns (c) and (f) regress idiosyncratic risk on suppliers' idiosyncratic risks. Standard errors adjusted for firm clusters are shown in parentheses. Note that we only collect market returns for firms traded in the ten largest market of the high-tech sector. Therefore, the number of firms with this alternative idiosyncratic risk measure is 1,960 instead of 2,122. Market risk, tier-1 risk, and industry-adjusted (supplier) days in inventory are included with log transformation. Country and sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

Table 3.15: Robustness Test: Unquantified Case

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio		
	(a)	(b)	(c)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	0.032*	0.028	0.049**
	(0.020)	(0.019)	(0.021)
Diamond Ratio	0.598***	0.712***	0.731***
	(0.156)	(0.162)	(0.163)
Market Risk	0.390***	0.317***	0.317***
	(0.012)	(0.012)	(0.012)
Tier-1 Supplier Risk	0.071***	0.067***	0.077***
	(0.011)	(0.011)	(0.011)
Financial and Operational controls	Yes	Yes	Yes
Country, Sub-industry controls	Yes	Yes	Yes
Random Effects	Yes	Yes	Yes
Observations	26,841	26,838	26,838
Number of Firms	2,321	2,321	2,321
Overall R-squared	0.1736	0.1610	0.1606

Notes. Columns (a) regresses total equity risk on suppliers' total equity risks; Columns (b) regresses idiosyncratic risk on suppliers' total equity risks; Columns (c) regresses idiosyncratic risk on suppliers' idiosyncratic risks. Standard errors adjusted for firm clusters are shown in parentheses. Market risk, tier-1 risk, and industry-adjusted (supplier) days in inventory are included with log transformation. Country and sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

Table 3.16: Robustness Test: Without Panel Financial and Operational Controls

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio			Tier-2 & Cosine Commonality Score		
	(a)	(b)	(c)	(d)	(e)	(f)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	0.034*	0.026	0.049***	0.034*	0.026	0.048***
	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)
Diamond Ratio	0.847***	1.016***	1.020***			
	(0.156)	(0.164)	(0.164)			
Cosine Commonality Score				0.549**	0.587**	0.598***
				(0.213)	(0.232)	(0.232)
Market Risk	0.404***	0.330***	0.329***	0.405***	0.330***	0.330***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Tier-1 Supplier Risk	0.062***	0.064***	0.074***	0.061***	0.062***	0.073***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Financial and Operational controls	No	No	No	No	No	No
Country, Sub-industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,593	24,590	24,590	24,593	24,590	24,590
Number of Firms	2,122	2,122	2,122	2,122	2,122	2,122
Overall R-squared	0.1171	0.0850	0.0852	0.1117	0.0765	0.0766

Notes. Columns (a) and (d) regress total equity risk on suppliers' total equity risks; Columns (b) and (e) regress idiosyncratic risk on suppliers' total equity risks; Columns (c) and (f) regress idiosyncratic risk on suppliers' idiosyncratic risks. Standard errors adjusted for firm clusters are shown in parentheses. Market risk and tier-1 riskare included with log transformation. Country and sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

Table 3.17: Robustness Test: US Listed and US Based Firms Only

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio			Tier-2 & Cosine Commonality Score		
	(a)	(b)	(c)	(d)	(e)	(f)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	-0.027	-0.066	-0.041	-0.026	-0.066	-0.041
	(0.044)	(0.044)	(0.051)	(0.044)	(0.044)	(0.051)
Diamond Ratio	1.513***	1.589***	1.593***			
	(0.451)	(0.478)	(0.477)			
Cosine Commonality Score				0.806*	0.779*	0.780*
				(0.420)	(0.448)	(0.447)
Market Risk	0.686***	0.721***	0.705***	0.635***	0.667***	0.652***
	(0.050)	(0.052)	(0.058)	(0.047)	(0.048)	(0.054)
Tier-1 Supplier Risk	0.031	0.035	0.033	0.027	0.031	0.028
	(0.030)	(0.030)	(0.031)	(0.030)	(0.030)	(0.030)
Financial and Operational controls	Yes	Yes	Yes	Yes	Yes	Yes
Sub-industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,706	4,706	4,706	4,706	4,706	4,706
Number of Firms	405	405	405	405	405	405
Overall R-squared	0.3359	0.3451	0.3450	0.3381	0.3532	0.3530

Notes. Columns (a) and (d) regress total equity risk on suppliers' total equity risks; Columns (b) and (e) regress idiosyncratic risk on suppliers' total equity risks; Columns (c) and (f) regress idiosyncratic risk on suppliers' idiosyncratic risks. Standard errors adjusted for firm clusters are shown in parentheses. Market risk, tier-1 risk, and industry-adjusted (supplier) days in inventory are included with log transformation. Sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

Table 3.18: Robustness Test: Driscoll and Kraay Standard Errors

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio			Tier-2 & Cosine Commonality Score		
	(a)	(b)	(c)	(d)	(e)	(f)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	0.005	-0.003	0.012	0.007	-0.002	0.012
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.009)
Diamond Ratio	0.102**	0.127***	0.127***			
	(0.034)	(0.039)	(0.039)			
Cosine Commonality Score				0.194***	0.186***	0.191***
				(0.036)	(0.039)	(0.038)
Market Risk	0.312***	0.242***	0.243***	0.312***	0.242***	0.243***
	(0.058)	(0.048)	(0.047)	(0.058)	(0.048)	(0.047)
Tier-1 Supplier Risk	0.045***	0.045***	0.043***	0.045***	0.044***	0.043***
	(0.008)	(0.007)	(0.006)	(0.008)	(0.007)	(0.006)
Financial and Operational controls	Yes	Yes	Yes	Yes	Yes	Yes
Country, Sub-industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Pooled OLS	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,593	24,590	24,590	24,593	24,590	24,590
Number of Firms	2,122	2,122	2,122	2,122	2,122	2,122
Overall R-squared	0.2145	0.2046	0.2047	0.2150	0.2051	0.2052

Notes. Columns (a) and (d) regress total equity risk on suppliers' total equity risks; Columns (b) and (e) regress idiosyncratic risk on suppliers' total equity risks; Columns (c) and (f) regress idiosyncratic risk on suppliers' idiosyncratic risks. Driscoll and Kraay (1998) standard errors are shown in parentheses. Market risk, tier-1 risk, and industry-adjusted (supplier) days in inventory are included with log transformation. Country and sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

Table 3.19: Robustness Test: Local Currency

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio			Tier-2 & Cosine Commonality Score		
	(a)	(b)	(c)	(d)	(e)	(f)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	0.043**	0.036**	0.052***	0.043**	0.035**	0.051***
	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)
Diamond Ratio	0.666***	0.767***	0.773***			
	(0.159)	(0.166)	(0.166)			
Cosine Commonality Score				0.456**	0.470**	0.481**
				(0.203)	(0.217)	(0.217)
Market Risk	0.392***	0.314***	0.315***	0.392***	0.314***	0.315***
	(0.012)	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)
Tier-1 Supplier Risk	0.056***	0.057***	0.066***	0.054***	0.055***	0.064***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Financial and Operational controls	Yes	Yes	Yes	Yes	Yes	Yes
Country, Sub-industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,593	24,590	24,590	24,593	24,590	24,590
Number of Firms	2,122	2,122	2,122	2,122	2,122	2,122
Overall R-squared	0.1654	0.1520	0.1516	0.1668	0.1542	0.1538

Notes. Columns (a) and (d) regress total equity risk on suppliers' total equity risks; Columns (b) and (e) regress idiosyncratic risk on suppliers' total equity risks; Columns (c) and (f) regress idiosyncratic risk on suppliers' idiosyncratic risks. Standard errors adjusted for firm clusters are shown in parentheses. Market risk, tier-1 risk, and industry-adjusted (supplier) days in inventory are included with log transformation. Country and sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

Table 3.20: Robustness Test: Tier-1 supplier concentration

DEPENDENT VARIABLE:	Tier-2 & Diamond Ratio			Tier-2 & Cosine Commonality Score		
	(a)	(b)	(c)	(d)	(e)	(f)
	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk	Total Risk on Total Risk	Idio Risk on Total Risk	Idio Risk on Idio Risk
Tier-2 Supplier Risk	0.035*	0.028	0.049**	0.035**	0.028	0.049**
	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)
Diamond Ratio	0.612**	0.717***	0.725***			
	(0.244)	(0.255)	(0.255)			
Cosine Commonality Score				0.378*	0.376*	0.387*
				(0.202)	(0.216)	(0.216)
Sub-industry Concentration	-0.001	-0.004	-0.011	0.284**	0.336**	0.334**
	(0.192)	(0.201)	(0.202)	(0.141)	(0.147)	(0.148)
Geographic Concentration	0.031	0.049	0.053	0.079	0.113	0.118
	(0.112)	(0.115)	(0.115)	(0.110)	(0.114)	(0.114)
Financial and Operational controls	Yes	Yes	Yes	Yes	Yes	Yes
Country, Sub-industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,593	24,590	24,590	24,593	24,590	24,590
Number of Firms	2,122	2,122	2,122	2,122	2,122	2,122
Overall R-squared	0.1687	0.1522	0.1519	0.1696	0.1532	0.1529

Notes. Columns (a) and (d) regress total equity risk on suppliers' total equity risks; Columns (b) and (e) regress idiosyncratic risk on suppliers' total equity risks; Columns (c) and (f) regress idiosyncratic risk on suppliers' idiosyncratic risks. Standard errors adjusted for firm clusters are shown in parentheses. Market risk, tier-1 risk, and industry-adjusted (supplier) days in inventory are included with log transformation. Country and sub-industry dummies are included. To simplify the table, we do not report on controls. *** --- 0.01 level, **--- 0.05 level and *--- 0.1 level.

3.8.4 Tier-2 Commonality Examples

Table 3.21 lists the tier-2 firms with the highest degree of commonality; these tier-2 firms are shared by greater than or equal to 20 tier-1 suppliers in a tier-0 firm's network.

Table 3.21: Tier-2 Suppliers with a High Degree of Tier-2 Commonality

Tier-2 Supplier Name	GICS	Headquarter Location	# of tier-0 firms that source from the listed tier-2 supplier who is shared by at least 20 of their tier-1 suppliers
Flextronics International Ltd.	45203020	USA	2
Avago Technologies	45301020	Singapore	2
Microsoft Corporation	45103020	USA	2
Nitto Denko Corporation	15101050	Japan	1
Atmel Corporation	45301020	USA	1
Intel Corporation	45301020	USA	3
LSI Corporation	45301020	USA	2
Qualcomm Inc.	45301020	USA	2
Texas Instruments Inc.	45301020	USA	6
United Microelectronics Corporation	45301020	Taiwan	4
ON Semiconductor	45301020	USA	2
Advanced Semiconductor Engineering Inc.	45301020	Taiwan	8
Taiwan Semiconductor Ltd.	45301020	Taiwan	13
Amkor Technology Inc.	45301010	USA	11
ARM Holdings plc	45301020	Great Britain	11
Broadcom Corporation	45301020	USA	2
STATS ChipPAC Ltd.	45301020	Singapore	5
NXP Semiconductors	45301020	Netherlands	1

Notes. The results are generated from the quantified supply network. Numbers in the final column can only be higher in the un-quantified supply network.

Table 3.22 displays a list of the corresponding tier-0 firms that source from the above heavily shared tier-2 suppliers. The table also demonstrates that many tier-2 suppliers in these firms' supply networks are shared.

Table 3.22: Tier-0 Firms with Common Tier-2 Suppliers

Name	GICS	Headquarter Location	Market Capitalization ($Mn)	# of tier-2s identified	# of tier-2s shared by at least two tier-1s	# of tier-2s shared by at least five tier-1s	# of tier-2s shared by at least twenty tier-1s
Cisco Systems, Inc.	45201020	USA	105,483	523	195	53	3
Dell Inc.	45202030	USA	25,465	847	340	114	6
Hewlett-Packard Company	45202030	USA	49,967	865	357	123	8
LG Electronics Inc.	25201010	South Korea	10,776	799	321	83	4
Apple Inc.	45202030	USA	442,008	889	333	110	3
Samsung Electronics Co., Ltd.	45202030	South Korea	143,504	956	346	105	6
Sony Corporation	25201010	Japan	19,422	816	282	73	2
IBM Corporation	45102010	USA	214,975	853	331	106	1
Lenovo Group Ltd.	45202030	China	8,708	838	333	91	1
WPG Holdings Ltd.	45203030	Taiwan	2,141	652	225	49	3
Avnet, Inc.	45203030	USA	4,799	720	286	88	6
Arrow Electronics Inc.	45203030	USA	4,274	926	366	122	6
Ingram Micro	45203030	USA	2,890	1100	481	198	12
Nokia Corporation	45201020	Finland	18,864	675	225	53	4
Alcatel-Lucent S.A.	45201020	France	6,584	418	150	34	1
Ericsson	45201020	Sweden	36,616	638	188	44	1
Motorola Solutions, Inc.	45201020	USA	15,248	619	190	32	2
Tech Data Corporation	45203030	USA	2,008	1099	435	173	9

Notes. The results are generated from the quantified supply network. The numbers from the un-quantified supply network are similar.

3.8.5 Event Details

Event ID 1: Energex reports indicated lost power for the majority of customers in the Brisbane City Council Area. The power outages were caused by impacts of severe weather from former Tropical Storm Oswald.

Event ID 2: NStar and National Grid reported lost power to most customers across Massachusetts. Widespread outages resulting from Winter Storm Nemo lasted for several days as repairs to infrastructure were made.

Event ID 3: National Grid reported lost power to customers throughout Rhode Island. Small pockets of outages may have persisted after the power was restored to majority customers.

Event ID 4: Hydro One reported lost power to most customers throughout Ontario following an ice and wind storm.

Event ID 5: Dominion Virginia Power reported lost power to the majority of its customers in the Richmond Metro and Tri-Cities. Earlier in the week over 90,000 customers were without power after severe weather passed through the region.

Event ID 6: Xcel Energy reports power was lost to the majority of customers in Minnesota following severe weather in the previous week. At the height of the outage, approximately 610,000 customers were without power statewide.

Table 3.23: Event Summary

ID	Lat.	Long.	Date Time	Duration	Impact	Source	Cause
1	-27.47	153.02	1/27/2013 6:08	3 days	Area-wide	Multiple	Tropical Storm Oswald
2	42.35	-71.05	2/8/2013 11:51	6 days	State-wide	Multiple	Winter Storm Nemo
3	41.82	-71.41	2/9/2013 2:54	3 days	Statewide	Multiple	Blizzard
4	43.76	-79.41	4/12/2013 11:05	4 days	Province-wide	Multiple	Ice and wind storm
5	37.54	-77.43	6/13/2013 9:49	3 days	Area-wide	Other	Severe weather
6	44.95	-93.2	6/21/2013 12:53	5 days	Statewide	Multiple	Severe weather
7	39.96	-83.00	7/11/2013 1:09	3 days	Statewide	Multiple	Severe thunderstorms
8	42.33	-83.04	7/20/2013 1:57	3 days	Area-wide	Other	Strong storms
9	45.51	-73.55	7/20/2013 3:04	5 days	Area-wide	Multiple	Severe weather
10	36.14	-95.99	7/24/2013 9:51	3 days	County-wide	Multiple	Storms
11	51.51	-0.11	10/28/2013 10:17	4 days	Region-wide	Multiple	St Jude Storm
12	45.50	-73.55	11/1/2013 6:39	3 days	Southern	Other	Strong winds
13	42.33	-83.04	11/18/2013 3:53	4 days	Region-wide	Other	Severe thunderstorms and heavy winds
14	52.23	21.01	12/6/2013 2:58	3 days	Nationwide	Media	Storm Xavier, strong winds & snow
15	32.77	-96.79	12/6/2013 10:38	4 days	County-wide	Multiple	Inclement weather
16	43.65	-79.40	12/22/2013 1:11	8 days	City-wide	Multiple	Freezing rain storm

Note: No high-tech firms in our data sample locates in the impact area of the identified 12 earthquake events. All of the 12 earthquake events struck far from land.

Event ID 7: Sources reported lost power to the majority of customers in Ohio following severe thunderstorms that occurred earlier in the week. Sandusky and Tuscarawas counties were most affected by the outages.

Event ID 8: DTE Energy reported lost power to most customers throughout Detroit's metropolitan area after strong storms swept through the area.

Event ID 9: Hydro-Quabec reported lost power to the majority of customers throughout the province of Quabec following outages caused by severe weather that hit the area the previous week. At the height of the storm, an estimated 560,000 customers reportedly lost electricity.

Event ID 10: AEP reported that approximately 5,164 customers were without power in Tulsa County following storms that moved through the area.

Event ID 11: UK Power Networks reported lost power to most customers across south-eastern England following the St Jude Storm. Residual outages may persist in some areas.

Event ID 12: Reports indicated lost power to the majority of customers following service disruptions caused by strong winds.

Event ID 13: Reports indicated that approximately 275,000 DTE customers were temporarily without power in Southeastern Michigan due to severe thunderstorms and heavy winds that swept through the area.

Event ID 14: Poland-media reports indicated widespread outages due to strong winds and snow from winter storm Xavier. At the height of the outages, approximately 300,000 people were without electricity service.

Event ID 15: Oncor reported lost power to the majority of customers throughout Dallas area, including Tarrant, Collin, Ellis, Kaufman and Hunt counties as a result of inclement weather. At least 215,000 customers had been without power for several hours in this region at the height of the outage.

Event ID 16: Toronto Hydro indicated lost power to the majority of its customers across Toronto following a freezing rain storm the previous week. Outages peaked at more than 522,426 customers across southern Ontario. Residual outages may have persisted.

CHAPTER 4

Manufacturing and Regulatory Barriers to Generic Drug Competition: A Structural Model Approach

Understanding the drivers of market concentration in the generic pharmaceutical industry is essential to guaranteeing the availability of low-cost generics. In this paper, we develop a structural model to capture the multiple determinants governing manufacturers' entry decisions; in particular, we focus on how manufacturing complexity and the regulatory environment affect concentration in generic drug markets. We estimate the model using data collated from six disparate sources. We find that manufacturing complexity, as reflected in the drug administration route, for example, significantly reduces the likelihood of generics entry. Moreover, the speed at which generic drug applications are processed by the FDA significantly affects the number of firms entering a market. Our policy simulations suggest that to encourage competition in the generic drug markets, the FDA needs to maintain a moderate review speed; being either too fast or too slow can be problematic and associated with more concentrated markets.

4.1 Introduction

The increasing use of low-cost generic drugs offers relief from rising health care costs. The U.S. Food and Drug Administration (FDA) estimated that generic drugs saved the health care system about $1.67 trillion over the last decade alone (Gottlieb June 21, 2017). According to the Association for Accessible Medicines, in 2016 generic drugs accounted for 89% of prescriptions dispensed and 26% of total prescription expenditures (Association for Accessible Medicines 2017).

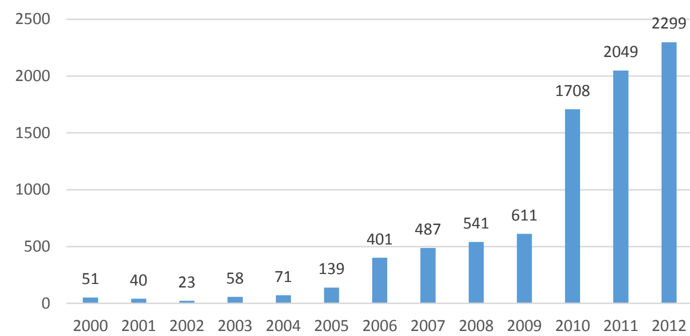
Competition among generic drug manufacturers significantly reduces the price of generics. According to an FDA report on generics competition and drug prices (FDA 2005), a generic drug could cost as much as its branded version in the presence of only one generics manufacturer in the market. A second entrant brings the generic drug's price down to 52% of the branded version. The price will further fall to 39% and 21% of the branded version with four or eight competitors, respectively. However, generic drug markets do not always attract a large number of manufacturers. Presently, about 10 percent of branded drugs with expired patents have no generics competition (Department of Health and Human Services 2017). Further, a quarter of the markets have only one generic-version manufacturer, and about a half of them have at most three generic-version manufacturers.

Multiple factors may drive generics manufacturers' market entry decisions. One of these factors is manufacturing complexity. Although brand-drug manufacturers have established reliable manufacturing processes for their product, won through years of experience, a generics manufacturer who is producing a drug for the first time will need to develop a safe and efficient production process. For example, all penicillin manufacturers need to establish a comprehensive quality control strategy to prevent cross-contamination, and such requirements include separate production facilities and equipments, separate air handling systems and testing for traces of penicillin where possible exposure exists. Such requirements can often be more demanding for generics manufacturers because they often run production lines of different chemical substances in parallel to keep the production costs low, but which also increases the risk of cross-contamination.

Besides manufacturing complexity, the FDA also plays an important role in determining the generics market structure. To obtain an approval to produce generic drugs, manufacturers are required to submit Abbreviated New Drug Applications (ANDAs) and demonstrate to the FDA that the generic products are safe and effective. As more branded drugs have become off-patent over the past decade, the number of generics applications submitted to the FDA has risen substantially. With limited funding and staff, the FDA struggled to review the influx of ANDAs in a timely manner. As of October 1, 2012, nearly 2,300 generic drug applications were in the queue awaiting an FDA decision.

To address this problem, FDA initiated the Generic Drug User Fee Amendments of 2012 (GDUFA), a five-year act aimed to speed up the ANDA review process. The act enabled the FDA to recruit more staff and keep up with its workload. As a result, the accumulated ANDA review backlog was mostly eliminated in 2016 (Woodcock 2016). The act has also been extended for another five years to continue speeding up the ANDA review process (Brennan Sept 26, 2016). The briefer turnaround time enables manufacturers to enter a new generics market earlier. It also enables manufacturers to generate revenue from the investment quicker, which effectively increases the manufacturer's payoff. However, from the perspective of a single firm, a faster FDA also increases the likelihood of entry by its competitors.

Figure 4.1: Number of Pending Applications Over 180 Days



The firm may be deterred from entering if it anticipates aggressive price competition. The two countervailing effects of a speedy review process imply that the relationship between ANDA review time and entry is not obvious.

In this paper, we examine the following questions: (1) What are the key determinants of a generics manufacturer's entry decision? and (2) Would a shorter review time necessarily lead to more competitive generics markets?

To answer these questions, we develop a structural model that explicitly captures a firm's decision in a simultaneous entry game. Like all other entry models, the existence of multiple equilibria is the key challenge to identification. To simplify the model and enable identification, many studies assume firms are symmetric, which allows them to link firms' equilibrium entry decisions to the number of entrants. However, such simplification fails to accommodate heterogeneous competitive effects, which is often the case in industries where firms vary significantly in size, specialty, and experience. In generic drug markets, in particular, manufacturers vary by size (e.g., the three largest manufacturers each captures 10% market share while the other manufacturers capture 4% of the market share at most, Statista 2015) and specialization (e.g., Fresenius Kabi focuses on providing medications dispensed via injection while Perrigo is a leading producer of extended tropical drugs). To study heterogeneity across manufacturers without making equilibria selection assumptions, we adopt the bound approach proposed in Ciliberto and Tamer (2009). This approach allows us to capture a general form of heterogeneity, albeit at the price of identification complexity: our structural model is not point-identified. In addition, applying this partial identification bound approach requires a non-parametric estimation of the empirical probabilities, which can be challenging when there are a large number of characteristics under consideration. We adopt the random forest approach to estimate such probabilities with sparse models (Breiman 2001).

Another major challenge in studying entry decisions in the generic pharmaceutical industry is the lack of a unified database. To understand the determinants of a firm's decision to enter a particular market, one needs to gather data on market conditions, drug forms, product costs, firms' manufacturing capability and specialization. For lack of comprehensive data, few comprehensive studies analyze the entry decisions of generic drug makers. Existing studies either use much older data or analyze only a handful of markets. For example, Morton (1999) focuses on the entry decisions of generic drug manufacturers in 1984 to 1994, Olson and Wendling (2013) studies the oral solid drug medications only, and Kong (2016) considers the top 100 drug markets by sales with at least one generics entrant.

In this paper, we overcome the data challenge by gleaning information from six disparate sources to collate a dataset for generic drug entries from 2002 to 2014. In particular, we acquired data from the Annual Editions of the Orange Book, the Clinformatics Data Mart, the National Drug Code (NDC) Directory, the DrugBank database, the FDA inspection database, and the FDA report on Implementation of the Generic Drug User Fee Amendments of 2012 (Woodcock 2016). The data period is selected to avoid industry-wide merger waves and to correspond with the period of drastic changes in the ANDA review time. Our dataset characterizes all 802 generic small-molecule markets during this time frame.[1]

To interpret the results from the simultaneous entry model, we evaluate the magnitude of each factor's impact based on its relative effect compared to that of the market size, measured as the log-scaled total prescription charge of the reference branded drug in the year before patent expiration. We find that those drugs administered parenterally (e.g., drugs administered via injection) require a double-sized market to attract the same number of manufacturers, because of the more stringent sterile requirement for manufacturing. Besides, ANDA backlogs significantly affect market outcome. The benefit of entering a double-sized market can be offset by an increase in the number of ANDA backlogs at the FDA by 500 applications.

Based on our structural model, we conduct a policy experiment to estimate the effect of regulatory barriers on the level of competition in generic drug markets. Using two performance metrics – the average number of generic entrants and the percentage of markets with zero or one entrant – we find that increase in ANDA backlogs significantly reduces the level of market competition. The average number of generics entrants drops by around 5% when the ANDA backlog increases from 50 cases (the actual queue length in the early 2000s) to 2,300 cases (the actual queue length in 2012). Interestingly, we find a non-monotone relationship between the percentage of markets with at most one entrant and the length of ANDA approval queue. The fraction of limited-entry markets first decreases and then increases with

[1]There are altogether 12 generic biological markets whose patents expired between 2002 and 2014. We exclude them from the main analysis due to the small number of markets and the different manufacturing requirements involved.

the size of the ANDA review backlog. This is because, on one hand, when the FDA only have a few ANDAs to review, medium-sized firms may choose not to enter a market when they perceive a high probability of entry from larger manufacturers, resulting in more monopolistic generics markets. On the other, when the backlog continues to accumulate, all firms will find entering a market less attractive, regardless of the competitors' decisions.

To sum up, our research demonstrates that manufacturing complexity and the length of the ANDA approval queue significantly affect the level of competition in generic drug markets. The policy simulation results reveal that the FDA should continuously monitor its ANDA review process and control the review time at a moderate speed to ensure sufficient competition in the generic markets. While a long review time reduces the number of approved manufacturers in the market, too short a review time may discourage entries from medium-sized manufacturers for fear of intense post-entry competition with market leaders.

4.2 Literature Review

Previous research regarding the generic pharmaceutical industry has looked at the efficiency of generic entry. Early research focused on the markets with branded drugs that expired post the Hatch-Waxman Act and found that a large number of generic entrants results in competitive generic prices (see Grabowski and Vernon 1992 and Frank and Salkever 1997, for example). More recently, Olson and Wendling (2013) estimate the causal effect of the second and third generic competitors on reducing the generic prices.[2] Generic entries reduce prices; however, factors that affect the manufacturers' generic entry decisions and the role of these factors in their entry decisions are not well known.

A stream of literature in the generic pharmaceutical industry has studied the impact of several supply and demand characteristics on a manufacturer's entry decision. Using a probit model, Morton (1999) shows that larger branded markets, markets to address chronic diseases, and markets in which firms have prior experience attract more generic entries. Two subsequent studies incorporate competitor effects into the entry decision and formulate an oligopoly game amongst manufacturers using dynamic structural models. Gallant et al. (2010) focus on the oral solid drug markets and document the spillover effect of a firm's past entry decisions on its future entry cost. Unlike Gallant et al. (2010) who take the generic revenue as exogenously given, Ching (2010) explicitly models the price evolution of branded markets and their generic counterparts to study the strategic interaction between these two

[2] Generic manufacturers are allowed to challenge the patent(s) of a innovator drug before the patent expiration date. If the generic manufacturer(s) managed to repeal the patent(s), a 180-day patent challenge exclusivity will be granted to the generic manufacturer(s).

types of markets. A recent paper by Kong (2016) examines the effect of competitor entry using a static discrete game and finds that each additional competitor on average reduces a firm's entry probability by 10 percent. Importantly, mainly due to identification issues, these studies that consider competitor effects treat firms as homogeneous entities. We contribute to the literature by modeling market and firm heterogeneity (i.e., market-specific profitability and firm-specific ANDA preparation cost and production cost) as well as the heterogeneous competitor effects. Our work is also more comprehensive in terms of the number of markets studied and the inclusion of manufacturing complexity metrics.

To the best of our knowledge, Ching (2010) is the only other study that looks at the role of the application review process in generic entry decisions. However, the author does not include any metric for the review time in the model; instead, the author conducts the policy experiment by directly increasing the entry probabilities of firms. In our study, we consider many more markets than Ching (2010). We also retrieve the number of ANDA backlogs from a FDA report and use it as a proxy of the speed of the ANDA review process. This metric allows us to link the application review time with a firm's entry decision and enables us to assess the magnitude of the impact of this important policy factor.[3]

The effect of waiting on the decision to join a queue has been studied in the operations literature. Deacon and Sonstelie (1985) and Png and Reitman (1994) both empirically study the effect of waiting time on the demand at gas stations. Using a structural model, Allon et al. (2011) estimate the cost that customers attribute to their waiting time and show that a fast-food chain can significantly increase its market share by reducing the customer waiting time. When the queue length is visible to the customers, Lu et al. (2013) find that the abandonment decision of customers mostly depends on the number of people waiting in the queue. When the queue length is not visible, delay announcements help customers form the estimated waiting time. Yu et al. (2016) demonstrate how delay announcements at the call center affect customers' perceived per-unit waiting cost. Recent papers in the literature have also documented the heterogeneity in customer's waiting sensitivity (see Akşin et al. 2013 and Lu et al. 2013, for example). We contribute to this line of work by relating the effect of queue length in a competitive setting, where manufacturers compete after waiting in a queue.

Methodologically, our paper builds on the literature on structural estimation of entry games. In an entry game model, multiple firms make entry decisions simultaneously, and the profit each firm obtains depends on other players' decisions. This type of model has been widely adopted in economic literature (Reiss and Spiller 1989, Bresnahan and Reiss 1990, Bresnahan and Reiss 1991, Berry 1992, Berry et al. 2006, Ciliberto and Tamer 2009,

[3]If the capacity did not change significantly, the ANDA review time is proportional to the number of backlogs.

and Aguirregabiria and Ho 2012). To account for heterogeneous competitive effects and to accommodate the existence of multiple equilibria, we adopt a similar estimation strategy as in Ciliberto and Tamer (2009), where moment inequalities are used for identification, as opposed to using moment equations. Our approach, however, diverges from theirs in two aspects. First, instead of contracting medium-sized firms' decisions into a binary decision, our model is more flexible and explicitly derives the equilibrium number of medium-sized entrants. Second, we apply the random forest method to estimate the empirical conditional choice probabilities (Breiman 2001). This method allows us to better estimate probabilities with non-parametric sparse models.

Our work also contributes to the growing literature on the use of structural estimation in operations management literature to identify underlying drivers of operational efficiency in various business settings and conduct policy simulations to improve the system. This literature studies pricing strategy under strategic consumer behavior (Li et al. 2014, Moon et al. 2017), effect of service quality on customer demand (Guajardo et al. 2015, Mani et al. 2015), geographic expansion strategy of retail stores (Zheng 2016), bidder behavior in auctions and contests (Olivares et al. 2012, Kim et al. 2014, Jiang et al. 2016), supply chain bullwhip effects (Bray and Mendelson 2012b, Bray and Mendelson 2015), and scheduling decisions in the operating room (Olivares et al. 2008).

4.3 Background of the Generic Drug Industry

This section provides background information of the generic pharmaceutical industry. We first present an overview of the industry and discuss its competitive landscape. We then discuss the key issues in generic drug manufacturing, and the process and requirements for manufacturers to obtain an ANDA approval.

4.3.1 Overview

Prescription medications comprise over 10% ($324.6 billion) of the United States' health care spending in 2015 (CMS 2015). Over the past decade, there has been an overall declining trend in the price of generic drugs and a corresponding, steady increase in their use (GAO 2016a). Since 2003, prescription drug spending growth has considerably slowed down, thanks largely to the increasing use of the relatively low-cost generic medications (Liberman and Roebuck 2010). The savings from prescribing generics has increased from $85 billion in 2007 to $253 billion in 2016 (Association for Accessible Medicines 2017). The Association for Accessible Medicines report also documented that nearly half of the cost savings from prescribing generics goes directly to patients.

In the United States today there are approximately two thousand drugs with generic versions. The generics industry started its boom period after the Drug Price Competition and Patent Term Restoration Act of 1984, often referred to as the Waxman-Hatch Act, under which generic pharmaceutical manufacturers were no longer required to repeat the costly clinical studies the innovators had already conducted. Manufacturers could submit Abbreviated New Drug Applications (ANDAs) to the FDA for marketing generic drugs in the United States. Apart from quality and safety requirements, as long as a manufacturer scientifically proves to the FDA that the generic version is *bioequivalent* to the reference brand-name drug (FDA 2017a), it would obtain an approval for production. Once the FDA approved the ANDA and the patent and exclusivity on the branded version had expired, the generics manufacturer could bring the product to market.

4.3.2 Competitive Landscape

Thirty percent of the country's generic pharmaceutical market is served by three manufacturers (i.e., Teva, Mylan and Sandoz), each accounting for around 10% of the generics market share in the U.S. Following these three market leaders, there are also a handful of medium-sized generics manufacturers that capture between 1% and 4% of the total market share (Statista 2015). There also exist about one hundred small players in the industry.

Regardless of the relatively large number of potential market players, the generic drug market is fairly concentrated according to the observations in the FDA's Orange Book. Among those markets with generic versions available, a quarter lists only one generic alternative. About half of those markets have no more than three generic alternatives (i.e., by three separate manufacturers). Surprisingly, there are also hundreds of drug products with an expired patent but have zero generics alternative (FDA 2017b).

The high market concentration in the generic drug market keeps prices of generics high, obviously hurting the patients. The small number of generics manufacturers also makes the generic pharmaceutical supply chain less resilient to demand or supply shocks, which may in the future lead to the price instability or drug shortage.

In the past several years, hundreds of generic drugs experienced a price increase of at least 100 percent within a year. Some of the drug prices increased ten-fold within a year. For example, the price of Piroxicam, an anti-inflammatory drug used to treat rheumatoid arthritis, increased from \$0.09 per capsule to \$1.94 per capsule (more than a 2,000 percent increase) from 2010 to 2011 (GAO 2016a). These extraordinary price increases mostly happen in less popular drugs that are marketed by a small number of manufacturers.

Starting from the mid-2000s, the United States has also seen an increasing number of generic drug shortages, which has led to rationing in treatment (Fink Jan 29, 2016), delays in care, and an increasing number of medication errors (McLaughlin et al. 2013). The new drug shortages in 2011 alone shot up to 270, with a record high of 430 cumulative active drug shortages. Focusing on the sterile injectable anti-infective and cardiovascular drugs, the 2016 U.S. Government Accountability Office report on drug shortages found that a small number of generics manufacturers is one of the factors that led to shortages of these drugs (GAO 2016b). With one or only a few manufacturers, supply disruptions such as manufacturing quality and production issues are more likely to lead to drug shortages (Kim and Morton 2015).

4.3.3 Manufacturing of Generic Drugs

While generics manufacturers are no longer required to conduct costly clinical trials to establish efficacy and safety, they must ensure that the drug is manufactured under the same quality standards as the brand-name drug. All drug manufacturers are required to conform to the Current Good Manufacturing Practices (CGMP). The CGMP cover all aspects of pharmaceutical production and are designed to minimize risks involved in the manufacturing process. Specifically, the regulations aim to minimize risks of product contamination, incorrect labeling, and incorrect dosage of active ingredients (WHO 2002). In the United States, Section 21 of the Code of Federal Regulations (CFR) summarizes the regulations pertaining to pharmaceutical products.

For example, to minimize potential product contamination, generics manufacturers need to satisfy certain sterile requirements. These requirements can vary substantially for drugs with different active ingredients and drugs administrated through different routes. Drugs administered via parenteral routes (e.g., injection) are more difficult to produce than those administered via enteral routes (e.g., solid pills). Producing parenteral drugs requires sterile formulation and demands that the products should not introduce contaminants into the human body because injected drugs bypass some of the body's natural defenses and can therefore pose particular risks to human health (Perspective Press 2016). Generics manufacturers are also required to document manufacturing processes and ensure data integrity to pass quality inspections conducted by the FDA. All records required under the CGMP are subject to inspection (FDA 2016).

Given all these requirements, safe manufacturing of generic drugs is not an easy task, and it is not uncommon for generics manufacturers to fail the FDA quality inspection. For instance, recent quality issues have been reported in Lupin (Edney Nov 14, 2017) and Dr.

Reddy's Laboratories (Tremblay Aug 4, 2017). Manufacturers are required to take remedial efforts to address quality-control issues before they can resume production.

4.3.4 ANDAs and the Role of the FDA

While the Waxman-Hatch Act significantly lowered the entry barrier for generics manufacturers, the cost of filing an ANDA is still non-trivial. Generics manufacturers need to fulfill requirements in the ANDA's six areas: chemistry, manufacturing, testing, labeling, inspections, and bioequivalence. To show that the generic products are *bioequivalent* to the reference branded products alone, manufacturers need to acquire the reference product, produce the generic alternatives on-site, and recruit and conduct pharmacokinetic crossover studies on human subjects, all of which are costly and time-consuming. The cost of preparing an ANDA is estimated to range from 2 to 5 million dollars (Berndt and Newhouse 2012).

With more generics applications submitted to the FDA in the past decade, the administration was unable to keep up with the pace of demand. As a result, by 2010, significant delays afflicted the ANDA review process. In 2003, the average approval time for a generic drug was 20 months (Meadows 2003). The time-to-approval jumped to 31 months in 2011 and continued to rise to 42 months in 2014 (Ebert Aug 20, 2016). Such a long wait time raises the opportunity cost vis-à-vis compromised sales, creating a burden for generics manufacturers. The long backlog of unapproved ANDAs have also been blamed in part for high drug prices (Kaplan Dec 29, 2015).

In response to the long review time[4] and the resulting backlog, the FDA initiated the GDUFA in 2012 to speed up the review of ANDAs. To further ensure the timeliness of ANDA review, the administration has extended the act for another five years and also recently took steps to speed up approvals for markets with limited or no competition (Brennan June 27, 2017).

4.4 Data

We retrieve information from six disparate data sources. In this section, we first introduce these data sources. We then define the drug market and discuss why we choose to focus our study on competition among initial entrants. Lastly, we present how we construct and define the variables to be included in the empirical model.

[4]We refer to the review time as the total delay at the FDA, including both the processing time and the wait time.

4.4.1 Data Sources

To study the effect of each of the market entry determinants, we obtain data from the following sources:

(1) Annual editions of the Orange Book (also known as *Approved Drug Products with Therapeutic Equivalence Evaluations*) from 2000 to 2016;

(2) Clinformatics Data Mart;

(3) National Drug Code (NDC) Directory;

(4) DrugBank database (Version 5.0);

(5) FDA Inspections database;

(6) Implementation of the Generic Drug User Fee Amendments of 2012 (Woodcock 2016).

The Orange Book is used to identify drug markets that are subject to potential generics entries. Clinformatics Data Mart is used to construct proxies for market profitability. We link these two databases using the National Drug Code from the National Drug Code Directory. The DrugBank database is used to obtain characteristics of drug products such as molar mass and indications of the active ingredient. The FDA Inspection database and the FDA report on Implementation of the Generic Drug User Fee Amendments are used to construct proxies for manufacturing quality and the ANDA backlog, respectively.

We obtain patent data, exclusivity data, and approved drug products data from annual editions of the Orange Book. The Orange Book is published by the FDA and identifies the *complete* set of branded innovator products and generic products approved by the administration. We use patent and exclusivity expiration dates associated with the branded drug applications to identify generics entry opportunities. Each approved product in the Orange Book with application type “A” corresponds to a generics-related application. For these observations, the FDA provides the applicant name, the approval date, the reference branded product, as well as characteristics of the drug product, including number of active ingredients, route of administration, and strength. Based on its past approval history, we are also able to construct a firm’s experience with the ingredient and the firm’s experience with the route.

To measure market size, we obtain the annual prescription charge and quantity for the reference branded drugs from Clinformatics Data Mart, provided by Optum, Inc. The original data source of the Clinformatics Data Mart comes from a national US private health insurer. We cross validate the data with data from the Centers for Medicare & Medicaid

Services (CMS) and Drugs.com.[5] According to the Drug Listing Act of 1972, drug products are required to be identified and reported using a universal product identifier, the National Drug Code (NDC). At the level of the nine-digit NDC, we retrieve the annual total claim counts and total charges from Clinformatics for 2001 through 2015. The nine-digit NDC can be used to uniquely identify a drug product. We rely on the NDC Directory to merge the Clinformatics data with the Orange Book.[6]

In order to better assess market profitability and production difficulty of each drug product, we obtain additional characteristics of the drugs from the DrugBank database. This database provides extensive biochemical and pharmacological information about drugs marketed in different countries (Law et al. 2013). Ayvaz et al. (2015) map all other drug databases to the DrugBank database to study the overlaps between various data sources, due to the DrugBank database's broad inclusion. This latest version of the DrugBank database (Version 5.0) contains 2,021 FDA-approved small-molecule drugs, 233 FDA-approved biological drugs, 94 nutraceuticals and over 6,000 experimental drugs (DrugBank 2017). For each identified active ingredient, the DrugBank database provides the molecule type (small-molecule ingredient or biological protein), the molar mass, the structured indications extracted from the FDA drug labels and scientific publications, as well as the therapeutic class of the active ingredients, which is indicated using the Anatomical Therapeutic Chemical (ATC) classification code.

To obtain a proxy for manufacturing quality, we obtain facility inspection records since October 1, 1999 from the FDA Inspections database through a Freedom of Information Act (FOIA) request. We focus on inspections that received final classifications. An inspection classification reflects the compliance status of the manufacturer site at the time of the inspection. The conclusions are reported as Official Action Indicated (OAI), Voluntary Action Indicated (VAI), or No Action Indicated (NAI). FDA concludes an inspection with OAI if significant objectionable conditions or practices were found and the firm must take regulatory action to address the issues. Contrarily, a VAI classification indicates that the FDA revealed objectionable conditions, but the issues were not significant enough to warrant reg-

[5] We are unable to disclose the name or the market share of the insurer due to a non-disclosure agreement with the data provider. However, we cross validate it with CMS's Part B and Part D data as well as the top 100 drug list from Drugs.com. We do not directly use the Part B National Summary Data File from CMS because there are only a limited number of branded drug products covered under the Medicare Part B program. The majority of the prescription drugs are covered under the Part D program; whereas the longitudinal annual revenue and quantity data for the drug products covered in the Part D program is not available. The top-100 drug list (2003 – 2013 by sales amount and by sales units), available from Drugs.com, is compiled from QuintilesIMS, a company that provides proprietary data on the total sales and volumes of drug products. We find that the two data sources are consistent for those best-selling drugs.

[6] We retrieve the NDC for the branded drugs no longer marketed in the United States from past NDC data available at http://www.nber.org/data/national-drug-code-data-ndc.html.

ulatory actions; whereas an NAI classification indicates that the FDA found no objectionable conditions.[7] Besides the final classification, inspection records also provide us with the list of the Code of Federal Regulations (CFRs) a facility violates.

To study the impact of an ANDA backlog on a firm's entry decision, we obtain the annual number of pending generics applications since 2000 from the FDA report on Implementation of the Generic Drug User Fee Amendments of 2012 (GDUFA) (Woodcock and Wosinska 2013). After the implementation of the GDUFA, more detailed ANDA submission information and progress became available since 2015; however, this level of information was not documented by the FDA and thus not available via FOIA request.

4.4.2 Definition of Drug Markets

Since the Orange Book does not provide a history of patents expired before its published date, we retrieve the earlier patent data by combining the patent expiration dates documented in the annual Orange Book from 2000 to 2016. It is worth noticing that the patent expiration date for a certain drug may change over time. The date may be revised to an earlier date if generics manufacturers successfully repeal the innovator's patent; the date may also be revised to a later date if the innovator receives a patent extension. Therefore, whenever there is a conflict in the patent expiration date, we always keep the most updated one.

We define a drug market as a potential generic drug market to enter when the patent of the branded version expires. In particular, we focus on patents with expiration dates during the period from 2002 to 2014. We exclude markets with patent expiration dates before 2002 due to lack of data on market and firm characteristics. We do not include drugs with expiration dates after 2014 for the following reason. The U.S. generic drug industry experienced dramatic change with multiple large consolidations since late 2012 (Torreya Partners 2016). For example, Watson acquired Actavis in November 2012 and became the third largest generics company in the U.S. market, which was then acquired by Teva in 2015. Recall that the average review time of an ANDA application for a generics market open in 2014 was 42 months (Ebert Aug 20, 2016). In other words, for a generics manufacturer to ensure that its product enters the market as soon as a patent expires, it must submit its application at least 3 years before the expiration date. Therefore, by studying the markets with patent expiry no later than 2014, we focus on ANDAs filed before 2012 and thus limit the impact of the mergers and acquisitions on our study. This data exclusion also minimizes the influence of the implementation of GDUFA, the 2013 FDA act that reduced the average

[7] Inspections - Background, retrieved on July 10, 2017 from the FDA website, https://www.fda.gov/downloads/AboutFDA/Transparency/PublicDisclosure/GlossaryofAcronymsandAbbreviations/UCM212061.pdf.

review time of generic applications.

The most recent 37th edition of the Orange Book contains all approved ANDAs in the United States before 2017. Each observation is a unique combination of active ingredient, route of administration, strength, type, and the applicant. We filter the drug type and focus on prescription drugs. We omit over-the-counter (OTC) medications since most OTC drugs do not require FDA approvals to produce.

We define a potential generics market at the level of a combination of active ingredient and route of administration. Most of the previous studies on generic drug market define a market at the level of its active ingredient (Ching 2010, Kong 2016) but not route of administration. We choose to include the route of administration in the definition because of the variations in the production cost aforementioned in Section 4.3 and because manufacturers are required to submit an ANDA for each route of administration separately. We do not further separate the market based on the strength because most of the manufacturers enter the markets with same ingredient and route but different strengths at the same time.

4.4.3 Competition among Initial Entrants

We identify the entry time of a manufacturer by the FDA approval date, also available in the Orange Book. Even though manufacturers are allowed to submit ANDAs at any time, the applications submitted by generics manufacturers can only be approved after the patent of the branded version has expired. In other words, the ANDA approval date is the first day that a manufacturer is allowed to market its product to the public. As most of the firms start selling products immediately after their ANDAs get approved, we use the approval dates as market entry dates.

For each potential generics market, we choose to focus on the initial entries, i.e., we focus on those manufacturers with ANDA approval dates within the first two years after the patent expiry date.[8] We choose to focus on this period for the following reasons. Generics manufacturers change their competitive behavior based on the stage of the product cycle. Specifically, manufacturers behave quite differently in the initial stage when generics prices continue to fall down and in the latter stage when those prices are relatively stable at a low level. We only focus on and model the entry decisions for firms that choose to compete in the generic's early product life. The length of the generic's early product life is estimated to be around two years (Fein 2012). During this time, prices generally drop significantly as manufacturers compete for market share (Emanuel Aug 6, 2011).

[8]The main conclusions of the paper are consistent when we use either a one-year or a three-year initial period.

As a result, generics firms always prefer to enter a new market as early as they can. With an initial period of 24 months, we capture most of the market entries resulting from generics manufacturers who choose to compete in the initial stage.

During this initial stage, manufacturers' entry decisions are considered to be simultaneous. Because manufacturers do not reveal their entry plans due to strategic business considerations (Morton 1999), and because the FDA does not reveal information about the received applications, manufacturers do not observe or react to their competitors' decisions within this period of time. Therefore, their entry decisions can be considered simultaneous rather than sequential.

The total number of market entry opportunities from 2002 to 2014 is 814, including 802 chemical markets and 12 biological markets. In this study, we focus on the 802 generic drug markets with chemically manufactured active ingredients for the reasons outlined in Section 4.3.[9]

4.4.4 Variable Definitions

In the following, we discuss in detail how we construct the variables which will be used in our empirical analysis.

4.4.4.1 Measures of Market Characteristics

We include several attributes that affect the market profitability and the production cost for all manufacturers.

1) *Market size*: We follow Morton (1999) and use the total prescription charges of branded products one year before patent expiry as a proxy for the revenue of the branded drug market, i.e., the potential market size of the generic products.[10] Even though market size is an important predictor of the level of competition, there is still significant variation left to be explained. To illustrate this point, we classify the drug markets into three categories in Table 4.1 according their sizes: small (i.e., less than the 33.3 percentile), medium, and large (i.e., greater than the 66.7 percentile), and summarize the distribution of the number of entrants in each category. We see that larger markets tend to attract more entrants in general. However, a significant portion (about 9 percent) of large markets fail to attract any

[9]Our results are robust with the inclusion of the biological markets. A dummy variable is included to identify the biological markets.

[10]Besides volume, price of the branded drug may also affect manufacturers' entry decisions (Lee et al. 2016). We thus consider an alternative specification where we include both the prescription count and the average charge per prescription in the year before patent expiry. Results from using this alternative specification are consistent.

Table 4.1: Distribution of the Number of Entrants by Total Prescription Charge

	Large	**Medium**	**Small**	**Total**
Markets with 0 entrants	9.16%	23.22%	57.25%	29.55%
Markets with 1 entrant	26.37%	40.82%	23.28%	30.17%
Markets with 2 entrants	16.85%	16.10%	9.92%	14.34%
Markets with 3 entrants	11.36%	10.86%	3.82%	9.48%
Markets with 4 entrants	9.16%	3.75%	2.29%	5.74%
Markets with 5 entrants	5.86%	1.50%	1.91%	4.36%
Markets with $\geqslant$ 6 entrants	21.24%	3.75%	1.53%	6.36%

generics manufacturer and sometimes (about 5 percent) small markets can attract more than three generics manufacturers.

2) *Chronic drug*: We construct the chronic drug indicator from the structured indications of the active ingredients. In medical terminology, an indication refers to the use of a drug for treating a particular disease. For example, anti-inflammatory is an indication of ibuprofen. Based on the chronic disease list retrieved from the New York State Department of Health, we first identify the structured indications that treat chronic diseases, such as arthritis, asthma, cancer, diabetes and coronary artery disease. We then label a drug market as chronic if the active ingredient used in the drug is indicated to treat chronic diseases.

3) *Substitutability*: We consider a drug to be a substitute for another drug if the two medications have the same structured indication. Most of the drug products are listed with multiple indications. We calculate the number of substitutes for each of the drug's listed indications and take the minimum as the substitutability of the drug. We consider the minimum because the exclusivity of a drug, even for just one indication, demonstrates that the drug is hardly substitutable.[11] For example, naloxone is used to treat types of severe pains, opioid dependence and opioid overdose. There are several other painkillers on the market; however, naloxone is the only drug to treat opioid overdose, and we thus consider naloxone as non-substitutable (or have very low substitutability compared to other drugs).

4) *Route*: According to the Pharmacy Technician Perspective Press (2016), we categorize the route of administration into several groups: (1) topical route; (2) enteral route (including pill, oral liquid, extended-release pill, and delayed-release pill); and (3) parenteral route (including injection, ophthalmic solution, and otolaryngology solution). We specifically include a dummy indicator for drug products administered parenterally to account for the heterogeneity in the sterile requirements as discussed earlier in Section 4.3.

5) Finally, we also consider *the number of active ingredients* and *the molar mass of the*

[11]Our estimation results are also robust from using the average number of substitutes of the drug's listed indications.

ingredient in the analyses, as they affect the manufacturing complexity of a generic drug.

4.4.4.2 Measures of Firm Characteristics

We now consider firm-drug specific attributes that affect entry or production costs of a firm in a drug market.

1) *Firm i's experience in the therapeutic class:* A manufacturer's prior production experience may affect which markets it is likely to enter. For example, a firm's prior experience in a therapeutic class increases its likelihood of entering a drug market within that same class again (Morton 1999, Lee et al. 2016). To construct the firm's experience in a therapeutic class, we assign the four-digit Anatomical Therapeutic Chemical (ATC) code to each active ingredient.[12] For a manufacturer i and a drug market m, we count the number of drug products in the same therapeutic class as drug m that have been approved for production by the manufacturer i in the past ten years prior to drug m's patent expiry date. If drug m is associated with one therapeutic class, we use this count as firm i's experience in the therapeutic class. If the drug is associated with multiple therapeutic classes, we take the average count over all related therapeutic classes.

2) *Firm i's Herfindahl index over therapeutic classes:* While some firms specialize in certain therapeutic classes, others have broad portfolios. We measure a firm's therapeutic concentration using the Herfindahl index (Morton 1999), $H_i = \sum_{c=1}^{C_i} p_{ic}^2$, where p_{ic} represents the fraction of drug products produced by firm i that are in therapeutic class c, and C_i denotes the set of therapeutic classes that firm i produces. To construct this measure, we again consider all drugs approved for production in the past ten years prior to the patent expiry date of the drug market under consideration. Intuitively, a higher Herfindahl index indicates a more concentrated therapeutic base.

We also follow Morton (1999) to construct two additional count variables indicating the following:

3) *Firm i's experience in the same active ingredient*, and

4) *Firm i's experience in the same route of administration.*[13]

[12] The most detailed seven-digit ATC code uniquely identifies the active ingredient. We choose the four-digit ATC code to balance the accuracy and the inclusion for our definition of the therapeutic class. A four-digit ATC code provides three levels of classifications. At the first level, the ATC code identifies the anatomical group, i.e., the system the drug targets (e.g., cardiovascular system, nervous system); at the second level, the ATC code identifies the therapeutic group, i.e., the disease the drug treats (e.g., cardiac therapy, vasoprotective drugs); at the third level, the ATC code identifies the pharmacological group, i.e., the medical effect the drug has (e.g., antiarrhythmic drugs, cardiac stimulants).

[13] Besides counting the number of drug products administered via the same route, we also construct an alternative zero-one indicator of whether firms have prior experience in the route. Our estimation results are robust from using this alternative measure.

Firms with prior experience in the same active ingredient typically incur lower costs of finding certified active ingredient suppliers. Firms with prior experience in the same route of administration are already familiar with the corresponding sterile requirements and the use of related equipment, and therefore incur lower fixed and variable production costs.

Besides a firm's prior experience, a firm's manufacturing quality and prior approval history may also affect the firm's likelihood of entering a market.

5) *Firm i's manufacturing quality:* We use the outcomes of the FDA inspections during the ANDA review period as a proxy for firm i's manufacturing quality. Specifically, we consider inspections conducted within two years before the patent expiry date of the generic market m. We compute the total number of citations a firm received during the two-year period as the indicator for the firm's manufacturing quality.[14] We obtain a reasonable proxy for manufacturing quality since the pre-approval inspection usually happens during the later stage of the ANDA review process. A higher citation count implies a lower quality at the firm's manufacturing facilities.

6) Lastly, we control for *firm i's recent ANDA approvals* in the two years prior to the patent expiry date of the market under consideration to account for potential serial correlations in firms' manufacturing and financial conditions.

4.4.4.3 Regulatory Environment Measure

We use the annual count of ANDAs that were pending for over 180 days as a proxy of the ANDA approval delays at FDA. The backlog of pending ANDAs creates delays in market entry, which leads to slower return on investment and compromised sales opportunities. In particular, we assign the annual backlog count two years before the patent expiry date to each drug market. Note that the number of pending applications might be an endogenous variable, as manufacturers' entry decisions may inversely affect the queue length. First, we would like to point out that we use the ANDA queue length two years prior to the expiration date, that is, the queue length around the time of ANDA submission, which is only affected by previous entry decisions. This alleviates the endogeneity concern. However, since submission time could vary by firms and markets, the queue length may not be an accurate measure of what a firm actually faces when it submits an ANDA. Moreover, serial correlations in firm-level profitability may also introduce a correlation between firms' past and future entry decisions.

[14]We compute the number of Official Action Indicated (OAI) classifications a firm received after inspections during the two-year period and use it as a alternative quality measure. The result from using OAI counts generates the same insight. We present the result with the citation count in the main analyses because it offers more variation in the quality metric. We also weight the total number of citations by the number of inspections as another alternative quality measure. The estimation result from using the weighted citation counts produces similar implication.

To ensure the robustness of our results and conclusions, we further adopt an instrumental variable strategy when estimating the effect of ANDA queue length. Specifically, with a relatively stable FDA staff level during our study period (Woodcock 2016), the number of backlogged applications is primarily driven by an increasing number of expiring patents over the years. Since these patents were typically established twenty years ago, the number of expiring patents in a year can be considered exogenous. Consequently, we use the total number of patent expiries in the adjacent three-year time window, year t to $t + 2$, as an instrument for the ANDA queue length in year t.

4.5 Model and Estimation

In this section, we first describe the model framework. We then discuss the equilibrium strategies of the entry game and introduce the identification and estimation methods. Lastly, we address the computational problem by reducing the dimensionality of manufacturers' decision space.

4.5.1 Framework

Let N denote the number of pharmaceutical manufacturers and M denote the number of generic drug markets. As discussed earlier, a drug market is considered a potential generics market to enter when the patent of the brand-name drug expires. We focus on the initial entries, i.e., for each generics market, we focus on the manufacturers with ANDA approval dates within the first two years after the branded version went off-patent.

In each potential generics market, manufacturers play an entry game of complete information. All N manufacturers simultaneously decide whether to enter a market. Post-entry payoffs are determined by characteristics of the market (e.g., market size, chronic vs. acute disease, drug complexity) and characteristics of firms (e.g., specialty, prior experience, manufacturing quality). We index manufacturers by i and drug markets by m, and let y_{im} be the binary indicator of "manufacturer i enters the generics market m". Let $y_m = \{y_{im}\}_{N\times 1}$, a vector of zeros and ones, denote the decision vector of the game. A manufacturer who chooses to enter i.e., $y_{im} = 1$, receives a payoff of value $\pi_{im}(y_m)$. The payoff of not entering is normalized to zero.

Note that the payoff of entering a market depends not only on a firm's profitability while operating in the market, but also the cost the firm incurs in getting approved by the FDA. Since the FDA does not make information available on firms that have submitted ANDAs but were not approved, we are unable to model the application stage and approval stage separately. However, our model accounts for market and firm characteristics that affect both production costs and application costs.

Following Ciliberto and Tamer (2009), we specify manufacturer i's payoff in market m as a linear function of market- and firm-specific characteristics as well as the regulatory environment.

$$\pi_{im}(y_m) = \alpha_0 + \alpha' \cdot D_m + \beta' \cdot X_{im} + \gamma \cdot P_t + \sum_{j \neq i} \delta_j \cdot y_{jm} + s_m + s_{im}, \qquad (4.1)$$

where matrix D_m denotes market characteristics, and matrix X_{im} denotes firm i's characteristics in market m. Note that firm characteristics are market-specific, as a firm's experience varies when we consider markets of different therapeutic classes, active ingredients and administrative routes. Vector P_t is the regulatory environment in the year t when market m opens to generics manufacturers. Lastly, s_m and s_{im} are the unobserved shocks to payoff. Market-specific shock s_m represents unobserved market heterogeneity. Firm-market-specific shock s_{im} represents the idiosyncratic shock to the payoff for firm i in market m. In particular, we assume $s_{im} \sim \mathcal{N}(0, \rho^2)$ and normalize the variance of the market-specific shock such that $s_m \sim \mathcal{N}(0, 1)$.

The parameter set $\theta = (\alpha_0,\ \alpha',\ \beta',\ \gamma,\ \rho,\ \{\delta_j : j = 1, \cdots, N\})$ is to be estimated: α_0 is the constant term in the firm's payoff, α' is a vector of parameters that measures how market characteristics affect payoff, β' is a vector of parameters that measures how a firm 's own characteristics affects its payoff, γ measures the impact of the regulatory environment on firm's entry decision, and ρ is the standard deviation of the firm-market-specific shock mentioned above. $\{\delta_j : j = 1, \cdots, N\}$ measures the effect of competitors' entries on a firm's payoff.

The inclusion of asymmetric competitive effort enables us to capture heterogeneity across firms. This is particularly important in the pharmaceutical industry because even the top ten generics manufacturers are fairly different in terms of market share and distribution privileges (Statista 2015). Most of the entry game models in the literature consider homogeneous competitors, because allowing asymmetric competitive efforts leads to multiple equilibria and cause identification problems in the empirical inference stage. We discuss estimation and identification in detail in Section 4.5.3.

4.5.2 Equilibrium Strategies

Without loss of generality, we normalize the value from the outside option to be zero. Consequently, in equilibrium, a manufacturer will enter a market if and only if the post-entry payoff is non-negative; we can thus model the entry decisions to market m as

$$\begin{aligned} y_{im} &= 1[\pi_{im}(y_m) \geqslant 0] \\ &= 1[\alpha_0 + \alpha' \cdot D_m + \beta' \cdot X_{im} + \gamma \cdot P_t + \sum_{j \neq i} \delta_j \cdot y_{jm} + s_m + s_{im} \geqslant 0], \text{ for } i = 1, \cdots, N. \end{aligned} \tag{4.2}$$

Any vector y_m that satisfies the above binary simultaneous equations is a pure-strategy Nash equilibrium of the game. We rule out mixed strategy equilibria following the literature (Bresnahan and Reiss 1990, Berry 1992, Ciliberto and Tamer 2009). Despite this, the above game has multiple pure-strategy equilibria even with the simplest model specification. To illustrate this point, consider a model with just two firms, and their entry decisions are described by the following equations. For simplicity, we omit market- and firm-specific variables in this example.

$$\begin{aligned} y_{1m} &= 1[\pi_{1m}(y_m) \geqslant 0] = 1[1 - y_{2m} + \epsilon_{1m} \geqslant 0], \\ y_{2m} &= 1[\pi_{2m}(y_m) \geqslant 0] = 1[1 - y_{1m} + \epsilon_{2m} \geqslant 0]. \end{aligned} \tag{4.3}$$

One can easily show that $(y_{1m}, y_{2m}) = (1, 0)$ and $(0, 1)$ are both pure-strategy equilibria to this model when $\epsilon_{1m}, \epsilon_{2m} \subseteq [-1, 0)$; that is, one firm enters but the other does not. The existence of multiple equilibria results in an incomplete econometric structure. The multiplicity implies that the relationship between the entry decisions and the exogenous variables is a many-to-one mapping rather than a one-to-one mapping. This poses significant challenges to model identification and empirical inferences.

To address the challenge of multiple equilibria, three solutions have been proposed in the literature. The first solution transforms the problem of predicting the exact equilibrium, who enters and who does not, to predicting the equilibrium number of entrants instead (Bresnahan and Reiss 1991). For example, in Equation (4.3), exactly one firm will enter the market in the equilibrium when $\epsilon_{1m}, \epsilon_{2m} \subseteq [-1, 0)$, even though the entering firm could be either firm. Berry (1992) shows the uniqueness of the equilibrium number of entrants under mild conditions with regard to the payoff function. Uniqueness is an attractive property; however, in this model, competitors can only affect a firm's payoff through the equilibrium number of entrants and no heterogeneous competitive effects are allowed. Since one of the objectives of our study is to capture heterogeneous competitive effects in the generic pharmaceutical industry, this approach is not desirable.

The second solution proposed in the literature defines a selection rule that chooses a particular equilibrium when facing multiplicity (Bjorn et al. 1984, Bajari et al. 2010). However, the choice of the selection rule can be arbitrary and often imposes strong assumptions on the model.

The last solution – the solution we adopt – is proposed by Ciliberto and Tamer (2009). They implement a bound approach that partially identifies the model, i.e., the model is set-identified but not point-identified. The identified set, however, will shrink to a point estimate under certain distributional conditions of the observables. Even when the model is not point-identified, one can still make inferences based on the partially identified parameter sets. This method captures heterogeneous competition effects and does not impose arbitrary selection rule on multiple equilibria. In the next section, we discuss how this method is applied to our setting. In particular, we will discuss how we adapt the model and the estimation method such that (1) we can account for a much larger number of firms' entry strategies than Ciliberto and Tamer (2009), and (2) overcome the challenge of dimensionality when estimating the model.

4.5.3 Identification and Estimation

To address the issue of asymmetric firms and multiple equilibria, we adopt a bound approach first proposed by Ciliberto and Tamer (2009). We denote $\Pr(s|I_m)$ as the conditional choice probability (CCP) that we observe s as the equilibrium entry strategy conditional on the exogenous market- and firm-specific characteristics and the regulatory environment measure, contained in I_m.

We denote the total unobserved shock to firm i's payoff in market m as ϵ_{im}, the sum of s_m and s_{im}, and denote $\epsilon_m = \{\epsilon_{im}\}$ as the vector of unobservables in market m. Under Equation (4.2), the conditional probability of observing s as the equilibrium can be written as

$$\begin{aligned}\Pr(s|I_m) &= \int \Pr(s|\epsilon_m, I_m)dF(\epsilon_m)\\ &= \int_{R^u(\theta,I_m)} 1\, dF(\epsilon_m) + \int_{R^m(\theta,I_m)} \Pr(s|\epsilon_m, I_m)dF(\epsilon_m),\end{aligned} \tag{4.4}$$

where $R^u(\theta, I_m)$ denotes the region of the unobserved ϵ_m within which the entry game admits s as the unique equilibrium conditional on I_m at parameter θ, $R^m(\theta, I_m)$ denotes the region of ϵ_m within which s is one of the multiple equilibria conditional on I_m at parameter θ, and $\Pr(s|\epsilon_m, I_m)$ is the probability that s will be selected as the equilibrium when there are multiple equilibria.

As discussed earlier, the merit of the bound approach is that it does not impose an *a-priori* selection rule. Instead, it derives bounds of $\Pr(s|I_m)$ by taking advantage of the fact that $\Pr(s|\epsilon_m, I_m)$ is a well-defined probability function, and thus only takes a value between zero and one. In our model, the strategy vector s can take 2^N different values. To see how we derive bounds of $\Pr(s|I_m)$, let us denote $\{s^k\}$ as the set of all potential strategies, i.e., all potential values that the decision vector y_m can take. The value of each s^k is one of the 2^N permutations of a $N \times 1$ binary vector, representing all possible combinations of the N firms' entry decisions.

Let us also define

$$l(\theta, I_m) \equiv \int_{R^u(\theta,I_m)} dF \quad \text{and} \quad u(\theta, I_m) \equiv \int_{R^u(\theta,I_m)} dF + \int_{R^m(\theta,I_m)} dF. \tag{4.5}$$

We can now obtain the upper and lower bounds on the conditional choice probabilities in the following vectorized format:

$$\underbrace{L(\theta, I_m) \equiv \begin{bmatrix} l^1(\theta, I_m) \\ \vdots \\ l^{2^N}(\theta, I_m) \end{bmatrix}}_{\text{lower bound on conditional probability}} \leqslant \underbrace{\begin{bmatrix} \Pr(s^1|I_m) \\ \vdots \\ \Pr(s^{2^N}|I_m) \end{bmatrix}}_{\text{empirical conditional probability}} \leqslant \underbrace{\begin{bmatrix} u^1(\theta, I_m) \\ \vdots \\ u^{2^N}(\theta, I_m) \end{bmatrix} \equiv U(\theta, I_m)}_{\text{upper bound on conditional probability}} \tag{4.6}$$

Specifically, the lower bound $l^k(\theta, I_m)$ indicates the probability that we observe the strategy vector s^k as the unique equilibrium of the game, and the upper bound $u^k(\theta, I_m)$ represents the probability of observing s^k either as the unique equilibrium or as one of the multiple equilibria. Inequality (4.6) represents the conditional moment inequity for our empirical estimation. The identification approach is to find the set of parameters such that the empirical conditional probability of observing strategy s^k in the equilibrium is admitted within in the range generated by the lower and upper bounds given the set of parameters.

We estimate the model in two steps:

Step 1. Estimate the conditional choice probabilities, $\{\Pr(s^k|I_m) : k = 1, \cdots, 2^N\}$.

These probabilities can be estimated through a multinomial logit model that regresses the observed entry decision vector s^k on the exogenous variables, i.e., the market, firm characteristics and the ANDA backlog contained in I_m in our setting. An alternative method is to partition the exogenous variables and then nonparametrically estimate the CCPs by counting the fraction of observations conditional on the given realizations of the exogenous variables. We adopt the latter approach because it does not impose a distribution. As a result, this non-parametric approach can accommodate more complex interactions between the exogenous variables.

In order to construct the nonparametric estimator, Ciliberto and Tamer (2009) group the exogenous variables into bins based on their quantiles. In our setting, because we include a relatively large number of exogenous variables in the model, if we follow their approach, we are going to create more market groups than the number of observed markets. We thus consider an alternative approach and partition the drug markets using the *random forest* method. Intuitively, the random forest algorithm bootstraps the original market data (Breiman 1996) and fits a classification tree for each bootstrapped data set using a random subset of exogenous variables (Ho 1998).

The classification tree works by recursively partitioning the sample based on the value of the exogenous variables. Specifically, in each step, the algorithm searches for a variable in I_m and a split point p to partition the markets in a group into two exclusive subgroups G_1 and G_2 such that the within-group sum-of-squared error is minimized:

$$\sum_{k=1}^{2^N}\left[\sum_{m\in G_1}(s_m^k-\bar{s^k}_{G_1})^2+\sum_{m\in G_2}(s_m^k-\bar{s^k}_{G_2})^2\right],$$

where $\bar{s^k}_{G_i}$ is the average probability of observing s^k as the equilibrium for markets in the subgroup G_i. Intuitively, the classification tree groups markets in a way such that the probability of observing a particular equilibrium is similar for markets within each subgroup.

We derive the final estimation result of the random forest method by averaging over the result of each classification tree. The merit of this approach is that it allows us to more efficiently partition the exogenous variables based on their effects on the outcome variables, i.e., manufacturers' entry decision. In Section 4.9.1, we compare the prediction performance of various methods.

Once we obtain the estimation of CCPs following the random forest method, we substitute them in the optimization problem described in the next step.

Step 2. Find the parameter set Θ that minimizes the violation of the moment inequality conditions specified in Inequality (4.6).

Because the two bounds $L(\theta, I_m)$ and $U(\theta, I_m)$ cannot be derived analytically, we calculate them numerically through simulations. Specifically, we simulate the market-specific and firm-specific unobservables R times for every market. Given a parameter vector θ and a realization of the unobservables, we compute every firm i's payoff function in market m and verify whether s^k is an equilibrium strategy according to the requirements in Inequalities (4.2). Among the R rounds of simulations for market m, if a strategy s^k is observed as the unique equilibrium for l^k rounds and observed as one of the multiple equilibria for u^k rounds, the simulated lower and upper bounds on the probability of observing s^k in market

m can be computed as

$$\hat{l}^k(\theta, I_m) = \frac{l^k}{R} \quad \text{and} \quad \hat{u}^k(\theta, I_m) = \frac{(l^k + u^k)}{R}, \quad k = 1, \cdots, 2^N. \tag{4.7}$$

We can now estimate the parameter Θ by minimizing the sum of the squares of violations in Inequality (4.6). In other words, we minimize the incidences that the empirical CCPs derived in Step 1 fall outside the simulated bounds. Let us denote the estimated empirical CCPs as $\hat{p}(s^k|I_m)$. The objective function of the optimization problem is

$$\min_\theta V(\theta) = \sum_m \sum_{k=1}^{2^N} \Big[\big(\hat{p}(s^k|I_m) - \hat{l}^k(\theta, I_m)\big)_-^{\,2} + \big(\hat{u}^k(\theta, I_m) - \hat{p}(s^k|I_m)\big)_-^{\,2} \Big], \tag{4.8}$$

where $(a)_- = a \cdot 1_{[a<0]}$. The objective function calculates the total violations over all potential strategies and all markets. Due to the complexity of the objective function, the minimization problem is usually not point-identified. In general, the model gives an identified set Θ such that $\{\theta \in \Theta \mid V(\theta) = V(\arg\min_\theta V(\theta))\}$.

Even though the model is not point-identified, our model is likely to generate a small identified set of parameters. Tamer (2003) derives the sufficient conditions for point identifying such models. Specifically, it requires wide supports of some firm-specific characteristics X_{im}. We illustrate the identification idea in the following example. Consider the binary simultaneous equations:

$$\begin{aligned} y_{1m} &= 1[\pi_{1m}(y_m) \geqslant 0] = 1[\alpha_0 + \alpha \cdot D_m + \beta \cdot X_{1m} + \delta_2 \cdot y_{2m} + \epsilon_{1m} \geqslant 0], \\ y_{2m} &= 1[\pi_{2m}(y_m) \geqslant 0] = 1[\alpha_0 + \alpha \cdot D_m + \beta \cdot X_{2m} + \delta_1 \cdot y_{1m} + \epsilon_{2m} \geqslant 0]. \end{aligned} \tag{4.9}$$

Denote x_{1m} as one of Firm 1's characteristics. Without loss of generality, we consider the scenario where the corresponding parameter of x_{1m}, β_1, is positive. Consider the case when x_{1m} has a wide support on $\mathbb{R}$. When we drive x_{1m} to $-\infty$, Firm 1 will never enter the market m regardless of Firm 2's entry decision. Mathematically, the empirical probability of observing no firms entering the market equals to just the probability Firm 2 does not enter:

$$\hat{p}((0,0)|I_m) \xrightarrow{x_{1m} \to -\infty} \Pr(y_{2m} = 0|I_m) = \Pr(\alpha_0 + \alpha \cdot D_m + \beta \cdot X_{2m} + \epsilon_{2m} \geqslant 0).$$

This equation allows us to estimate the parameters related to Firm 2 from the reduced form regression model and guarantees point identification. Intuitively, the existence of variables with wide support helps isolate a manufacturer's entry decision from its competitors' and thus solves the identification problem resulting from interdependent firm decisions. It is unfortunate that one can rarely find variables with such wide support; however, the idea behind this sufficient condition sheds lights on how the variations in the firm-specific characteristics reduce the size of the identified parameter sets. As we will show later, in our setting, there exist multiple firm characteristics that retain the desired variation property.

To conduct inference of this partially identified moment inequality model, we use the subsampling method discussed in Chernozhukov et al. (2007) to construct the 95% confidence regions. The constructed contour region contains the true parameter set Θ with 95% probability. As a result, all estimation results are reported as intervals instead of point estimates.

4.5.4 Reducing the Dimensionality of the Decision Space

There are more than hundred generics manufacturers in the United States. From a computational point of view, it is extremely challenging to estimate a model that explicitly characterizes every manufacturer's entry decision. The size of the potential strategy set $\{s^k\}$ equals 2^N and grows exponentially in the number of firms N. We cannot obtain a reliable estimate of the empirical probability of observing each strategy vector with 802 generic drug markets.

To reduce the size of the decision space, we need to make simplifying assumptions. A typical solution to this challenge is to consolidate multiple firms' decisions into one decision. For example, when studying airlines' entry decisions, Ciliberto and Tamer (2009) consider all medium-sized carriers' entry decisions as one decision, and all low-cost carriers' entry decisions as one decision as well. Instead of predicting which medium-sized (or low-cost) carrier enters the market, they define entry by medium-sized (or low-cost) carriers as at least one medium-sized (or low-cost) carrier entering the market. They justify this approach by noting that in most markets, there is either zero or one medium-sized (or low-cost) carrier in the market.

We do not think this is a valid approach for the generic pharmaceutical industry. In this industry, there are multiple large generics manufacturers each capturing at least 10% of the market share, whereas there are also dozens of medium manufacturers each capturing at least 1% of the market share (Statista 2015). A drug market can still be considered a competitive market if multiple medium manufacturers are present. In our dataset, around 30% of the generic drug markets attract at least two medium entrants and around 18% of the markets attract at least three medium entrants. Therefore, we would like our model to be able to capture the number of medium firms in a market.

We now introduce the modified model that incorporates the number of medium entrants. We individually characterize the entry decision of the three largest generics manufacturers,

Table 4.2: Top Twenty Manufacturers by the Number of Approvals in 802 Generic Drug Markets

Firm Name	**No. of Generics Approvals**	**Firm Name**	**No. of Generics Approvals**
Mylan	178	Lupin	50
Teva Pharmaceutical	169	Taro Pharmaceutical	45
Sandoz	129	Zydus Pharmaceuticals	40
Apotex	112	Fresenius Kabi	39
West Ward Pharmaceuticals	97	Impax Laboratories	33
Sun Pharmaceutical	94	Perrigo	32
Dr. Reddy's Laboratories	80	Par Pharmaceutical	31
Aurobindo Pharma	74	Hospira	30
Actavis	69	Glenmark Pharmaceuticals	28
Watson Pharmaceuticals	66	Amneal Pharmaceuticals	25

i.e., Teva, Mylan and Sandoz. These are the three firms that have the highest number of generics approvals in our database. They are also the three largest firms by market share. In 2015, they together captured 40% of the U.S. market with the fourth-largest firm capturing only 4% of the market (Statista 2015).

We consider a generics manufacturer to be a medium-sized manufacturer if it has at least 25 generics approvals in the Orange Book and it captures at least 1% of the total U.S. generics market in 2015. This gives us seventeen medium-sized firms in the database. We will focus on the entry decisions by the three largest firms as well as these seventeen medium firms. These twenty firms participate in 85% of the markets where we observe generic entrants in our data. We do not lose generality of the results by excluding remaining small firms from the analysis. We obtain similar results if we include a binary decision variable capturing the decision of the remaining small firms (see Section 4.9.2).

With twenty firms at hand, there still can be 2^{20} equilibria. To further simplify the model, we assume that the medium manufacturers are symmetric and we capture the total number of medium entrants in equilibrium. This assumption largely reduces the dimension of the decision space. With this assumption, every medium manufacturer who chooses to enter the market receives the same payoff. We can now rewrite the manufacturer i's payoff in market m as

$$\pi_{im}(y_m, n_m) = \begin{cases} \alpha_{i0} + \alpha_i' \cdot D_m + \beta' \cdot X_{im} + \gamma_i \cdot P_t + \sum_{j \neq i} \delta_j \cdot y_{jm} + \delta_4 \cdot n_m + s_m + s_{im}, \\ \hfill i, j \in \{1, 2, 3\}, \\ \alpha_{i0} + \alpha_i' \cdot D_m + \beta' \cdot \bar{X}_{im} + \gamma_i \cdot P_t + \sum_{j \neq i} \delta_j \cdot y_{jm} + \delta_i \cdot (n_m - 1) + s_m \\ +s_{im}, \hfill i = 4. \end{cases} \tag{4.10}$$

Following the assumption that medium firms are symmetric, the medium firm characteristics

$\bar{X}_{im}$ denotes the average firm characteristics in market m across all medium firms. Similar to Ciliberto and Tamer (2009), we collapse the decision of the medium firms; however, instead of using a binary indicator, we allow the decision variable to take values of zero, one, two, three and above. Due to computational difficulties, we are not able to capture every possible value in terms of the number of medium firm entrants. We therefore treat three or more manufacturers present in the market as one case.[15] In other words, we predict whether zero, one, two, or more medium manufacturers are present in the market. This simplification further reduces the dimension of the decision space while preserving a moderate level of flexibility in characterizing market concentration levels.

For computational reasons, we do not allow full flexibility in the parameters. Specifically, we allow heterogeneous competitive effects $\{\delta_j\}$. We also allow the estimator of regulatory environment control (γ_i) as well as the constant term (α_{i0}) to vary between the top three large firms and the remaining medium firms. We incorporate the heterogeneous impact of ANDA queues on firms' payoff function for the following reasons. Recall that a longer ANDA review queue affects a firm's expected payoff through the increasing opportunity cost. Firstly, larger firms are likely to have more alternative investment options. For example, instead of investing in the generic application, larger firms can seek acquisition opportunities and start generating revenue as soon as they take over the target firm. In addition, because larger firms have higher production capacity, higher pricing power, and more distribution channels, delayed ANDA approval means more compromised sales opportunity and lost revenue.

In this modified model, the equilibrium vector for the game becomes (y_m, n_m), where y_m is a three-by-one vector that represents the three large manufacturers' entry decisions, and n_m is a numerical value ranging from zero to three that indicates the number of medium entrants. In total, there are altogether 32 possible values. Under this model, a firm's payoff from entry is not only affected by the presence of each of the three major firms, but also by the number of medium-sized firms, n_m.

4.6 Estimation Results

In this section, we first provide model-free evidence of market and firm heterogeneity. We then show the estimates of the empirical conditional choice probabilities derived from the random forest method. Lastly, we present the parameter estimates of the entry model and discuss how one should interpret the estimates.

[15] In our data, only 8.97% of generics markets attract more than three medium manufacturers in the initial stage.

4.6.1 Market-Level and Firm-Level Heterogeneity

Table 4.3 reports summary statistics of market and firm characteristics. Among the 802 generic drug markets, 191 drug products are used to treat chronic diseases and 315 drugs are dispensed parenterally. We find that the total prescription charge of the reference branded drug varies significantly across drug markets. Each drug product on average has 2.62 substitutes, and around two-thirds of the drug products are the only product in at least one of their structured indications. The majority of the drug products (87%) contain only one active ingredient, while the molecule size of the contained active ingredient also varies across drug products.

Among all firm-market combinations, there are 927 cases (5.8%) when a firm has prior experience in manufacturing the same ingredient. It is more common for a firm to have experience in producing drugs in the same therapeutic class (36%) or with the same route of administration (about 60% of the cases). In more than half of the cases, firms did not have citation records on file, whereas some firms experienced significant quality-control problems in the two-year period. We also observe that both the product diversity and the recently approved applications vary across firms.

Table 4.3: Summary Statistics

Variable	Mean	St. Dev.	Min	1st Quartile	Median	3rd Quartile	Max
Panel A: Market Characteristics							
Total Prescription Charge ($Mn)	13.98	31.9	0.001	0.34	2.47	11.49	365.61
No. Substitutes	2.62	7.14	0	0	0	2	56
No. Active Ingredients	1.17	0.56	1	1	1	1	8
Molar Mass (g/mol)	363.3	211.5	6.94	256.1	325.6	412.8	2,933.5
Panel B: Firm Characteristics							
Experience in Therapeutic Class	0.72	1.71	0	0	0	1	35
HHI over Therapeutic Classes†	654	895	0	316	478	680	10000
Experience in Route	10.12	15.12	0	0	2	15	85
Experience in Ingredient	0.04	0.24	0	0	0	0	3
No. of Citations	4.03	7.05	0	0	0	5	46
Recent ANDA Approvals	14.08	12.97	0	4	11	21	77

†*Note.* The HHI index over therapeutic classes is rescaled to a fraction in the estimation stage.

4.6.2 Parameter Estimates

Table 4.4 provides the estimates of the entry model. Column (a) reports the estimation results from using the observed ANDA queue length, while Column (b) reports the estimation results from using the instrument variable method. Recall that our model is only set-identified but not point-identified.[16] Therefore, for each parameter, we show the 95% confidence interval rather than a point estimate and its standard error. The results presented in the two columns are consistent both in terms of magnitude and statistical significance.

Table 4.4: Model Estimates

	(a) **Heterogeneous Competitive Effects**	(b) **Heterogeneous Competitive Effects with IV on Backlog**
Market Characteristics		
Total Prescription Charge (logarithm)	[0.074, 0.120]**	[0.059, 0.132]**
Chronic Diseases	[0.059, 0.176]**	[0.054, 0.225]**
No. Substitutes	[-0.011, 0.002]*	[-0.012, 0.001]*
Parenteral Route	[-0.177, 0.010]*	[-0.220, 0.031]*
No. Active Ingredients	[-0.075, -0.012]**	[-0.064, -0.011]**
Molar Mass (100g/mol)	[-0.037, 0.013]	[-0.037, 0.026]
Firm Characteristics		
Experience in Therapeutic Class	[0.003, 0.019]**	[0.002, 0.021]**
HHI over Therapeutic Classes	[-0.590, 0.036]*	[-0.733, -0.064]**
Experience in Route	[0.001, 0.006]**	[0.002, 0.008]**
Experience in Ingredient	[0.015, 0.117]**	[0.019, 0.122]**
No. Citations (tens)	[-0.005, -0.001]**	[-0.005, -0.001]**
Recent ANDA approvals	[-0.007, 0.001]*	[-0.006, 0.000]*
Regulatory Environment		
No. ANDA backlogs (thousands)		
Large Firm	[-0.214, -0.038]**	[-0.178, 0.022]*
Medium Firm	[-0.110, -0.030]**	[-0.129, -0.034]**
Competitive Effect		
Teva on Other Firms	[-0.967, -0.699]**	[-1.121, -0.904]**
Mylan on Other Firms	[-0.757, -0.458]**	[-0.733, -0.495]**
Sandoz on Other Firms	[-0.574, -0.374]**	[-0.654, -0.451]**
Medium Firm on Other Firms	[-0.391, -0.268]**	[-0.310, -0.208]**
Constant		
Large Firm	[-0.466, -0.065]**	[-0.555, 0.091]*
Medium Firm	[-0.459, -0.009]**	[-0.405, 0.440]
St. Dev. of Firm-Market Shock (ρ)	[0.013, 0.052]**	[0.013, 0.050]**
Model Performance		
Objective Function Value	69.34	69.02
Correctly Predicted Equilibrium Outcome†	0.2120	0.2057

†*Note.* If the model predicts multiple entry outcomes, we conclude that the model achieves a correct prediction as long as the observed market entry outcome is one of the predicted equilibria. ** — 0.05 level, * — 0.1 level

As shown in Table 4.4, the total prescription charge of the reference branded drug, the proxy of generic market size, is a significant indicator of how profitable the generic drug market will be. Drugs used to treat chronic diseases are more profitable than those used to treat acute diseases. This is likely because demand for chronic medications are more stable. We also find medications with fewer substitutes to be associated with a slightly higher payoff, though not significant at the 95% level.

Apart from the market profitability, our estimation results also suggest that production complexity significantly deters manufacturers from entering the market. This is because the

[16]Point-identified models generate a best point estimate of the parameters of interest. Set-identified models (models that allows for partial identification) make fewer assumptions but can only generate bounds on the parameters of interest.

complexity in the manufacturing process increases the production cost and thus results in a lower payoff. We find that drugs administered parenterally is generally less attractive to manufacturers compared to drugs administered via other routes due to the more stringent sterile requirements. The other two metrics of production complexity, the number of active ingredients contained in the drug product and the molar mass of the active ingredients, are also negatively associated with the payoff of manufacturers.

Besides the market characteristics, firm characteristics also affect a firm's payoff in a market. Results in Table 4.4 show that firms prefer to enter those generic markets where they have prior experience. A manufacturer's payoff is significantly higher if the firm has manufactured drugs with the same active ingredient ($[0.015, 0.117]$), or with the same administration route ($[0.001, 0.006]$), or in the same therapeutic class ($[0.003, 0.019]$). Comparably, the experience in the ingredient is the most valuable experience, whereas the experience in the administration route is the least valuable one. This is likely because experience in the ingredient is a more specific experience than in others. Firms in general are also rewarded for holding a diverse portfolio of therapeutic classes, as a higher Herfindahl Index (a more concentrated product portfolio) is associated with lower payoffs. This is likely because a more diversified product portfolio signals higher production capability and generates more stable incomes.

The estimation results demonstrate that a manufacturer's likelihood of entering a market is negatively associated with the number of inspection citations it receives around the ANDA review period. With a higher number of citations, a manufacturer is less likely to be approved for production by the FDA or has to incur higher costs in order to be approved. We also find that the firm's recent ANDA approvals is negatively associated with its likelihood of entering a market, likely because of capacity constraints, though this effect is small.

Lastly, we find that ANDA queue length, measured by the number of ANDAs pending for over 180 days, has a significantly negative impact on the likelihood of a manufacturer entering a market. The magnitude of the association also varies between manufacturers. Specifically, the association is twice as large for the three large firms compared to the average medium firms, i.e., $[-0.214, -0.038]$ for large firms compared to $[-0.110, -0.030]$ for medium firms. This result suggests that these industry giants are more influenced by the ANDA delays than their medium-sized competitors. This is likely because delays in ANDA approval lead to more compromised sales and lost revenues for these large firms or because they have more options elsewhere.

We also find evidence of heterogeneous competitive effects from the estimation results. The competitive effect from Teva's entry, which is estimated to be $[-0.967, -0.699]$, is significantly larger than that from an average medium manufacturer, which is estimated to be $[-0.391, -0.268]$. The differences in the competitive effects are sufficiently large to lead to multiple equilibrium numbers of firms, necessitating the incorporation of heterogeneous competitive effects.

To better understand the magnitude of the measured effects, we calculate the size of each effect relative to that of market size. In particular, medications used to treat acute diseases need a market three times as large as the market for medications used to treat chronic diseases in order to attract the same number of generic drug manufacturers. Drugs administered parenterally demand a double-sized market to secure the same level of competition as those drugs administered through other routes. If the ANDA queue length were increased by 500 cases, a market would need to increase its size by 43.6% to make it attractive for a medium-sized firm to enter, or double its size for a large firm to enter. We find that the competitive effects are of the largest scale. A thirty-fold increase in market size is needed to compensate for the entry from a medium competitor. And a nearly hundredfold increase in the market size is needed to offset the impact of the entry of a large firm like Sandoz.

4.7 Policy Simulations

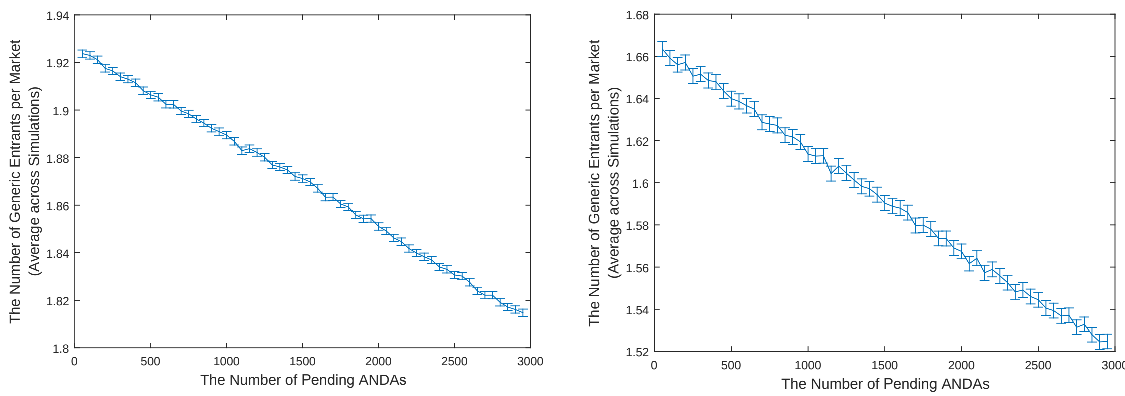
The rationalization of having generic drugs on the markets is to offer patients accessibility to low-cost medications; however, due to lack of competition, relatively high generic drug prices are observed in a significant proportion of the generic drug markets (e.g., see FDA 2005 and Dave et al. 2017). In order to achieve the goal of lowering pharmaceutical costs, the FDA has taken several actions to increase competition in the generic pharmaceutical industry, such as the GDUFA program implemented in 2012 which significantly reduced the ANDA backlog by the end of 2016.

In this analysis, in particular, we empirically test the influence of the ANDA queue on the competition level in the U.S. generics markets. More specifically, we want to understand whether the FDA's effort to shorten the wait time always increases the number of entries in generics markets. The policy experiment is performed on all 802 markets in the study period. In addition, we also look into the future and conduct another set of policy simulation on 141 future generics markets, whose patent will expire between 2018 and 2022.

We simulate the entry game with ANDA queues of different lengths and compare the market concentration level with the following two metrics: (1) the average number of generics entrants across markets, and (2) the percentage of markets with at most one generics entrant. We define a market with at most one generics manufacturer as a market with limited entry. We focus on these limited-entry markets for the following reasons. According to an FDA report on generics competition and drug prices FDA (2005), generic drug prices can be almost as high as that of branded drugs when there is only one manufacturer in the market. The price of generic drugs continues to fall when there are more manufacturers entering the market, where the largest price drop happens after the second entry into the market. It is thus important to reduce the number of markets with limited entry. Our focus is also consistent with recent FDA efforts in prioritizing reviews of drugs that have little or no competition (Brennan June 27, 2017).

Figure 4.2: The Average Numer of Generics Entrants per Market under Different Levels of ANDA Queue Length

(a) Markets in Study Period: 2002-2014 (b) Future Markets: 2018-2022



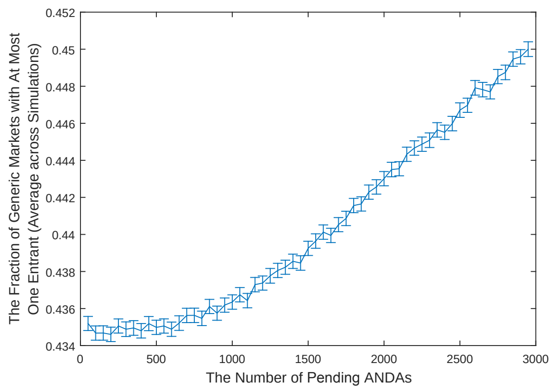
Note. The 95% confidence interval error bars are included for simulated average at each level of ANDA queue length.

At each level of ANDA queue length, we solve the entry games using numerical methods and obtain the equilibrium number of entrants. We again take the average number of entrants if multiple equilibria exist. We simulate the entry game with 5,000 draws of market-specific and firm-market-specific shocks and report the average value of the two concentration metrics across the simulations. The error bars are also generated to indicate the statistical significance.

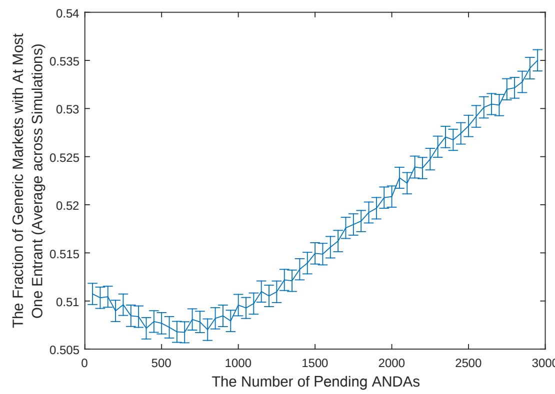
We first look at how the average number of generics entrants per market changes with the number of pending applications. By allowing firms to enter the market earlier, a shorter ANDA queue directly increases manufacturers' probability of entering. At the same time, a shorter queue also increases competitors' likelihood of entering and thus indirectly deters a firm from entering a market. We can see in Figure 4.2 that the mean entrants per market follows a monotonic decreasing trend. For the generics markets in our study period, the count of average entrant decreases by a total of 4.4% (from 1.924 entrants to 1.838 entrants) when we increase the ANDA queue length from 50 (the actual number in the early 2000s) to 2,300 (the actual number of 2012). For future generics markets, the reduction is 6.6% level (from 1.664 entrants to 1.553 entrants).

Figure 4.3: The Fraction of Generic Markets with At Most One Entrant under Different Levels of ANDA Backlogs

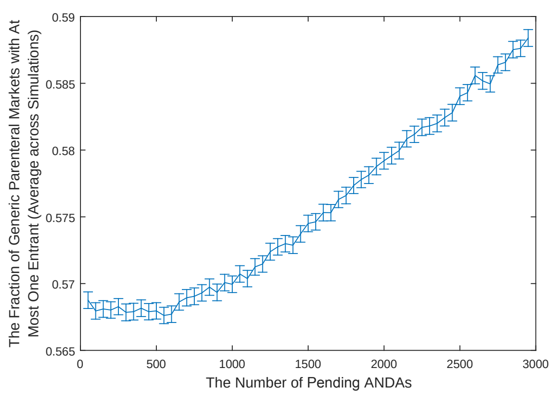
(a) All Markets in Study Period: 2002-2014



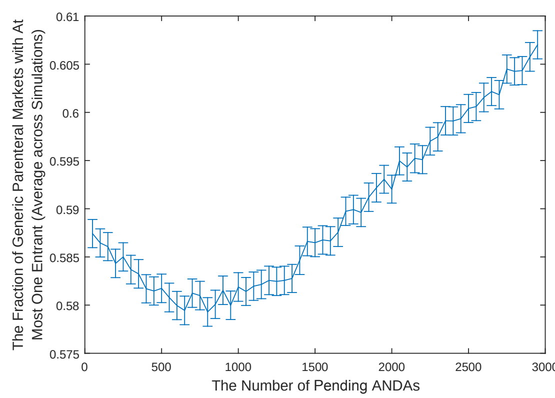
(b) All Future Markets: 2018-2022



(c) All Parenteral Markets in Study Period: 2002-2014



(d) All Future Parenteral Markets: 2018-2022



Note. The 95% confidence interval error bars are included for simulated average at each level of ANDA backlogs.

Even though the average number of entrants per market is larger with a shorter ANDA queue, the percentage of markets with limited entry may actually be higher with a shorter queue. In Figure 4.3, we observe a non-monotone relationship between the ANDA queue length and the percentage of markets with limited entry. In particular, the fraction of markets with at most one entrant is the lowest with moderate ANDA queue length at around 500. Specifically, for those generics markets with patent expires between 2002 to 2014, the percentage of limited-entry markets decreases by 3% (or 1.2 percentage points) when we reduce the ANDA queue length from 2,300 to 500. This reduction can be translated into approximately 10 more markets (1.2 percent of 802 markets) to have at least two generic entrants. Our experiments also show a discrepancy in the market outcome across different type of drug markets. Those markets with drugs dispensed parenterally are more likely to become limited-entry markets. The non-monotonic impact of the ANDA review time is also more salient on this market segment.

A longer or shorter ANDA queue will both increase the number of limited-entry markets. To understand why a short ANDA queue may actually hurt competition, recall that based on our estimates, a large manufacturer benefits from a shorter queue more than a medium manufacturer does. Therefore, a very short queue significantly increases the probability of a large manufacturer entering a market, making it unattractive for a medium firm. Likely due to the additional production cost, this competition effect appears to be stronger on those parenteral drug markets.

The results of the policy simulation suggest that ANDA backlog has a reasonably large impact on the competition level in the market for generic drugs. The FDA should avoid having a very long ANDA queue, which can significantly dampen the level of competition in terms of both the average number of entrants per market and the percentage of markets with limited entry. To do so, the FDA should carefully monitor the number of drug markets coming off-patent and plan ahead of time regarding their staff and funding level. That said, maintaining a very short ANDA queue is both costly and undesirable for the purpose of minimizing the number of markets with limited entry. Therefore, the more cost-efficient and effective approach for the administration is to maintain the queue length at a moderate level. Alternatively, the FDA may also consider establishing separate queues for generics applications with different manufacturing complexity level. There is a trade-off for prioritizing the review of a subset of ANDAs. For example, if the FDA prioritizes the process of parenterally dispensed drugs, the queue length of this market segment is reduced and thus the proportion of limited-entry markets can be brought down. However, the prioritization also means that less resources are available for other types of drug markets. The overall impact of such a strategy on market structure depends on the relative time involved in reviewing different kinds of generics applications. According to the FDA, this strategy may help balance the level of competition across markets and reduce the total number of limited-entry markets.

4.8 Conclusions

To deal with the rising health care costs and to guarantee the accessibility of low-cost generics, it is important to understand how this industry behaves. However, few studies have comprehensively analyzed the factors that lead to concentration in generic drug markets, partially due to lack of a unified database. In this paper, we collect data from six different sources and analyze a host of factors that may have impacted the level of competition in more than 800 drug markets that were opened for generics manufacturers between 2002 and 2014.

We develop a static entry game model to capture the factors governing manufacturers' entry decisions and estimate the impact of each factor using the data we have collected. Besides market size and firm's prior experience in similar drugs (drugs in the same therapeutic class, with the same active ingredient, or administered through the same route), we find that product difficulty significantly affects the level of competition in a drug market. For example, due to more stringent sterile requirements in the production process, drugs dispensed parenterally need a double-sized market in order to attract the same number of entrants. Moreover, we also find that delays in ANDA approvals can significantly dampen competition in generic drug markets. An increase in ANDA queue length by 500 applications requires a market to be twice as large to ensure the same level of competition. This effect is particularly concerning because *all* markets that are open for generics competition at the time will be impacted by delays in the ANDA approval process.

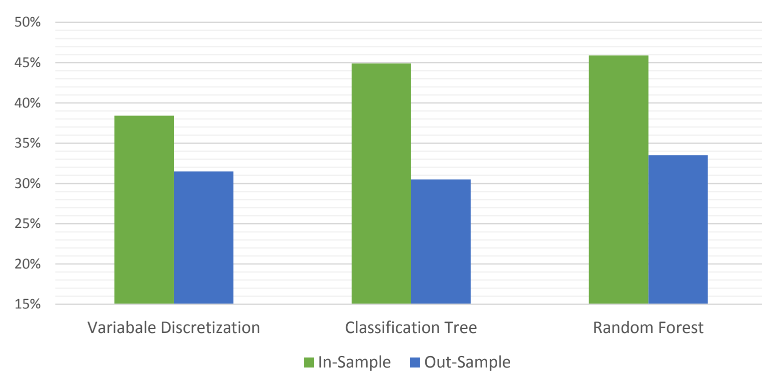
Our policy simulation results show that the excessive delays in the ANDA review process was one of the causes that led to the high concentration level in the generics markets. In order to achieve the goal of increasing generics competition and lowering pharmaceutical costs, the FDA should plan ahead based on the number of drugs coming off-patent. The administration should also maintain the time-to-approval at a moderate level to minimize the number of markets with no or limited generics entry.

4.9 Complementary Material

4.9.1 Estimation of Conditional Choice Probabilities

As described in Section 4.5.3, we calculate the empirical conditional choice probabilities in the first step using the random forest algorithm. Within each bootstrapped data, we can obtain CCP estimates of each possible market entry equilibrium for every generic market. The final estimates of the CCPs are obtained by taking the average among the 100 estimated CCPs derived from the bootstrapped data sets.

Figure 4.4: Comparison of Prediction Performance



We now compare the perdiction performance using different estimation methods. The first alternate method is the variable discretization method described in Ciliberto and Tamer (2009). This method partitions the markets based on the quantiles of exogenous variables. We partition the markets based on the market characteristics and stop further partitions when the markets within each group is smaller than 20. The second alternate method is classification tree method. Besides using the common CP criteria as the stopping rule, we also require a minimum leaf size of 20 to avoid over-fitting and make the results comparable between various methods. The final alternate method is random forest algorithm. This algorithm is an advanced version of the classification tree method. Specifically, to construct the prediction using the random forest method, we consider 100 bagged classification trees.

We now examine the in-sample and out-sample prediction accuracy, i.e., the probability that the methods correctly predict the market structure. Because the classification tree method is likely to over-fit the data sample, we anticipate that the method to have a relatively high in-sample accuracy but a low out-sample accuracy. This is indeed what we see in Figure 4.4. The two tree-based methods outperform the discretization method in the in-sample prediction performance, whereas the classification tree method perform the worst in terms of the out-sample predication performance.

We find that among the three methods, the random forest algorithm, due to its bagging procedure and random feature selection, generates the highest level of both in-sample (0.459) and out-sample (0.335) accuracy. We thus estimate the empirical CCPs using the random forest algorithm in the main analyses given its relatively better performance. Our estimation result of the structural model is also robust from using the variable discretization method in the first estimation step.

4.9.2 Robustness Tests: Model Specification

In this section, we conduct a robustness analysis to ensure that our estimation results are not driven by the model specification. We focus on the entry decision of the 20 largest generics manufacturers in the main analysis for computational reasons. We now extend the model by incorporating the entry decision of the remaining small firms; however, unlike including the number of medium firms, we only include a binary variable y_m^s that captures whether there exist other small manufacturers. In our database, 79.3% of the drug markets attract either zero or one small firms. The choice of including the binary decision variable guarantees the computational efficiency while managing to keep the model representative.

We do not construct firm characteristics for small manufacturers because there are constantly small firms entering and leaving the generic pharmaceutical industries. Instead, we focus on how the other factors, i.e., the market characteristics, the regulatory environment, and the competitive effects, determine the entries from those small players.

We now rewrite the manufacturer i's payoff in market m as

$$
\begin{aligned}
\pi_{im}(y_m, n_m, y_m^s) = \quad & \alpha_{i0} + \alpha_i' \cdot D_m + \beta' \cdot X_{im} + \gamma_i \cdot P_t + \sum_{j \neq i} \delta_j \cdot y_{jm} + \delta_4 \cdot n_m + \delta_5 \cdot y_m^s \\
& + s_m + s_{im}, \quad i, j \in \{1, 2, 3\}, \\
\pi_{im}(y_m, n_m, y_m^s) = \quad & \alpha_{i0} + \alpha_i' \cdot D_m + \beta' \cdot \bar{X}_{im} + \gamma_i \cdot P_t + \sum_{j=1}^{3} \delta_j \cdot y_{jm} + \delta_i \cdot (n_m - 1) \\
& + \delta_5 \cdot y_m^s + s_m + s_{im}, \quad i = 4, \\
\pi_{im}(y_m, n_m, y_m^s) = \quad & \alpha_{i0} + \alpha_i' \cdot D_m + \gamma_i \cdot P_t + \sum_{j=1}^{3} \delta_j \cdot y_{jm} + \delta_i \cdot n_m + s_m + s_{im}, \quad i = 5.
\end{aligned}
\tag{4.11}
$$

Besides the heterogeneous competitive effects $\{\delta_j\}$, we also allow the estimator of regulatory environment control (γ_i) as well as the constant term (α_{i0}) to vary between the top three large firms, the seventeen medium firms, and the remaining small firms. We incorporate the heterogeneous effect of ANDA queues length for the reasons discussed in Section 4.5.4.

Table 4.5: Robustness Test: Model Estimates

	(a) Main Model Eq. 4.10	(b) Extended Model Eq. 4.11
Market Characteristics		
Total Prescription Charge (logarithm)	[0.074, 0.120]**	[0.089, 0.127]**
Chronic Diseases	[0.059, 0.176]**	[0.027, 0.344]**
No. Substitutes	[-0.011, 0.002]*	[-0.006, 0.003]
Parenteral Route	[-0.177, 0.010]*	[-0.220, 0.023]*
No. Active Ingredients	[-0.075, -0.012]**	[-0.068, 0.025]
Molar Mass (100g/mol)	[-0.037, 0.013]	[-0.036, 0.003]*
Firm Characteristics		
Experience in Therapeutic Class	[0.003, 0.019]**	[-0.003, 0.034]*
HHI over Therapeutic Classes	[-0.590, 0.036]*	[-0.372, 0.037]
Experience in Route	[0.001, 0.006]**	[0.000, 0.006]**
Experience in Ingredient	[0.015, 0.117]**	[-0.049, 0.148]
No. Citations (tens)	[-0.005, -0.001]**	[-0.023, 0.005]
Recent ANDA approvals	[-0.007, 0.001]*	[-0.003, 0.002]
Regulatory Environment		
No. ANDA backlogs (thousands)		
Large Firm	[-0.214, -0.038]**	[-0.235, -0.006]**
Medium Firm	[-0.110, -0.030]**	[-0.153, -0.007]**
Small Firm		[-0.216, -0.045]**
Competitive Effect		
Teva on Other Firms	[-0.967, -0.699]**	[-1.032, -0.468]**
Mylan on Other Firms	[-0.757, -0.458]**	[-0.723, -0.381]**
Sandoz on Other Firms	[-0.574, -0.374]**	[-0.687, -0.355]**
Medium Firm on Other Firms	[-0.391, -0.268]**	[-0.431, -0.167]**
Small Firm on Other Firms		[-0.527, -0.128]**
Constant		
Large Firm	[-0.466, -0.065]**	[-0.649, -0.139]**
Medium Firm	[-0.459, -0.009]**	[-0.550, -0.052]**
Small Firm		[-0.345, 0.098]
St. Dev. of Firm-Market Shock (ρ)	[0.013, 0.052]**	[-0.037, 0.101]
Model Performance		
Objective Function Value	69.34	49.72

In this extended model, the equilibrium vector for the game becomes (y_m, n_m, y_m^s), where y_m is a three-by-one vector that represents the entry decisions of the market leadersm Teva, Mylan, and Sandoz, n_m is a numerical value ranging from zero to three that indicates the number of medium players, and y_m^s is a binary variable that stands for the entry decision of small manufacturers.

Table 4.5 provide the estimates of the entry model. For ease of comparison, we append in Column (a) the original estimation results in the main analysis. Column (b) reports the estimation results after including the entry decision of small firms. With the consistent estimates between the two columns, we conclude that our results are robust from using this extended model.

CHAPTER 5

Conclusion

We motivated this theses to better understand the role of supply chain structure on disruptions. This thesis attempts to find empirical evidences of direct and intermediary effects of supply chain structures. In particular, we have looked at two types of supply chain structures:

1. Chapters 2 and 3 investigated the commonly cited phenomena of tier-2 supplier sharing; showed that it is prevalent for firms to have common tier-2 suppliers in their multi-tier supply networks; and found that those shared tier-2 suppliers impose the focal tier-0 firm with higher financial risk. Findings in the two chapters underscore the necessity for firms to look beyond their direct tier-1 suppliers and suggests that firms prioritize their efforts, based on the level of such concentration, when managing sub-tier supplier risks.

2. Chapter 4 looked at the limited entry problem in the generic pharmaceutical industry, identified the key determinants of market entry decisions, and illustrated the non-monotonic relationship between the firm's entry decision and the backlog of generic applications at the government regulatory agency. Though overall a shorter time to approval encourages more generics entries, quicker approvals may sometimes deter manufacturers from entering if firms perceive an increasing probability of competitors' entry. We show that it is essential for the government to continuously monitor the review process and prioritize the markets with a high risk of limited market entry.

We now briefly discuss the limitations of this thesis and outline related future research. In Chapter 3, we established that the extent of tier-2 sharing influences tier-0 firm financial risk. A natural question to ask is whether we can find similar results for firm operational risk. In order to look into the operational performance, however, we need to build a facility-level supply network that documents the material flows between firms' production sites. We collaborate with a large automaker to gain access to their extended multi-tier supply relationship data. We find that tier-2 sharing phenomenon is also prevalent at the facility level, and the degree of tier-2 sharing is heterogeneous across supplier locations and the nature of the commodity sourced. Using the tier-1 supplier's delay in delivery as the key operational performance metric, we find that tier-2 sharing affects delays when disruptions occur, but it does not impact delays in routine performance.

In Chapter 4, we examined the market entry decisions of generic manufacturers. This serves as a starting point to achieve the ultimate goal: to uncover the root cause(s) of the persistent drug shortage and price hikes in the generic pharmaceutical industry. A next step is to look into other drivers of the drug supply problem, such as the manufacturing quality issue and the concentration in active pharmaceutical ingredient (API) producers. Apart from the limited market entry, manufacturing quality problem is suspected to be another main driver of the unstable drug supply. In an ongoing work, we specifically focus on the impact of quality inspections. Conditioned on the citations manufacturers received, our analysis reveals a significant correlation between the tone of government warning letters and the number of future drug shortages. We found that those drug products manufactured by firms receiving a more stringent warning letter are more likely to experience shortages. It is likely the case that manufacturers opted for taking remediation efforts and halted the production lines when they perceived more stringent quality requirements. To further elaborate on the relationship between quality problem and production decision, we plan to construct a dynamic model that identifies the factors influencing manufacturers' production decisions. Disruptions in supply of generic drugs to end consumers could also occur due to decisions made by downstream supply chain players like distributors and retailers. We investigate these issues for the pharmaceutical supply chain in China. Specifically, we have set up a field experiment that leverages the ordering behavior of retailer managers to focus on improving the supply continuity at the retailer level.

BIBLIOGRAPHY

Daron Acemoglu, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 2012.

Victor Aguirregabiria and Chun-Yu Ho. A dynamic oligopoly game of the us airline industry: Estimation and policy experiments. *Journal of Econometrics*, 168(1):156–173, 2012.

Kenneth R Ahern. Network centrality and the cross section of stock returns, 2013. Working paper, Marshall School of Business, University of Southern California, Los Angeles, CA.

Zeynep Akşin, Barış Ata, Seyed Morteza Emadi, and Che-Lin Su. Structural estimation of callers' delay sensitivity in call centers. *Management Science*, 59(12):2727–2746, 2013.

Yasin Alan, George P. Gao, and Vishal Gaur. Does inventory productivity predict future stock returns? a retailing industry perspective. *Management Science*, 60(10):2416–2434, 2014.

Gad Allon, Awi Federgruen, and Margaret Pierson. How much is a reduction of your customers' wait worth? an empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Operations Management*, 13(4):489–507, 2011.

Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. High idiosyncratic volatility and low returns: International and further US evidence. *Journal of Financial Economics*, 91(1): 1–23, 2009.

Erjie Ang, Dan A Iancu, and Robert Swinney. Disruption risk and optimal sourcing in multitier supply networks. *Management Science*, 2016.

Applied Quantitative Research. The Devil in HMLs Details: Factors, Daily. https://www.aqr.com/library/data-sets/the-devil-in-hmls-details-factors-daily, 2016.

Association for Accessible Medicines. Generic Drug Access and Savings in the U.S. (retrieved from AAM website). URL:http://accessiblemeds.org/resources/blog/2017-generic-drug-access-and-savings-us-report (accessed on July 17, 2017), 2017.

Serkan Ayvaz, John Horn, Oktie Hassanzadeh, Qian Zhu, Johann Stan, Nicholas P Tatonetti, Santiago Vilar, Mathias Brochhausen, Matthias Samwald, Majid Rastegar-Mojarad, et al. Toward a complete dataset of drug–drug interaction information from publicly available sources. *Journal of biomedical informatics*, 55:206–217, 2015.

Volodymyr Babich, Apostolos N Burnetas, and Peter H Ritchken. Competition and diversification effects in supply chains with supplier default risk. *Manufacturing & Service Operations Management*, 9(2):123–146, 2007.

Patrick Bajari, Han Hong, and Stephen P Ryan. Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5):1529–1568, 2010.

Ramji Balakrishnan, Thomas J Linsmeier, and Mohan Venkatachalam. Financial benefits from jit adoption: Effects of customer concentration and cost structure. *Accounting Review*, pages 183–205, 1996.

Jean-Noel Barrot and Julien Sauvagnat. Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, 131(3):1543, 2016.

Marcus A Bellamy, Soumen Ghosh, and Manpreet Hora. The influence of supply network structure on firm innovation. *Journal of Operations Management*, 32(6):357–373, 2014.

Uri Ben-Zion and Sol S Shalit. Size, leverage, and dividend record as determinants of equity risk. *The Journal of Finance*, 30(4):1015–1026, 1975.

Ernst R Berndt and Joseph P Newhouse. Pricing and reimbursement in us pharmaceutical markets. *The Oxford Handbook of the Economics of the Biopharmaceutical Industry*, page 201, 2012.

Steven Berry, Michael Carnall, and Pablo T Spiller. Airline hubs: costs, markups and the implications of customer heterogeneity. *Competition policy and antitrust*, 2006.

Steven T Berry. Estimation of a model of entry in the airline industry. *Econometrica: Journal of the Econometric Society*, pages 889–917, 1992.

Paul A Bjorn, Quang H Vuong, et al. Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation. Technical report, 1984.

Christoph Bode, Stephan M Wagner, Kenneth J Petersen, and Lisa M Ellram. Understanding responses to supply chain disruptions: Insights from information processing and resource dependence perspectives. *Academy of Management Journal*, 54(4):833–856, 2011.

Michael J Braunscheidel and Nallan C Suresh. The organizational antecedents of a firm's supply chain agility for risk mitigation and response. *Journal of Operations Management*, 27(2):119–140, 2009.

Robert L. Bray and Haim Mendelson. Information transmission and the bullwhip effect: An empirical investigation. *Management Science*, 58(5):860–875, 2012a.

Robert L Bray and Haim Mendelson. Information transmission and the bullwhip effect: An empirical investigation. *Management Science*, 58(5):860–875, 2012b.

Robert L Bray and Haim Mendelson. Production smoothing and the bullwhip effect. *Manufacturing & Service Operations Management*, 17(2):208–220, 2015.

Robert Louis Bray, Juan Camilo Serpa, and Ahmet Colak. Supply chain proximity and product quality, 2016. Working paper, Kellogg School of Management, Northwestern University, Evanston, IL.

Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Patrick Brennan. Lessons learned from the Japan earthquake. *Disaster Recovery Journal*, 24(3): 22–26, 2011.

Zachary Brennan. FDA to Speed Reviews of More Generic Drugs, Offers Lists of Those With No Competition. *Regulatory Affairs Professionals Society*, (accessed on Nov 5, 2017), June 27, 2017.

Zachary Brennan. GDUFA II: FDA Looks to Speed Up Generic Drug Approval Process. *Regulatory Affairs Professionals Society*, (accessed on Nov 22, 2017), Sept 26, 2016.

Timothy F Bresnahan and Peter C Reiss. Entry in monopoly market. *The Review of Economic Studies*, 57(4):531–553, 1990.

Timothy F Bresnahan and Peter C Reiss. Empirical models of discrete games. *Journal of Econometrics*, 48(1-2):57–81, 1991.

Stephen J Brown and Jerold B Warner. Using daily stock returns: The case of event studies. *Journal of financial economics*, 14(1):3–31, 1985.

Gérard P Cachon, Taylor Randall, and Glen M Schmidt. In search of the bullwhip effect. *Manufacturing & Service Operations Management*, 9(4):457–479, 2007.

Xin Chen, Melvyn Sim, David Simchi-Levi, and Peng Sun. Risk aversion in inventory management. *Operations Research*, 55(5):828–842, 2007.

Victor Chernozhukov, Han Hong, and Elie Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284, 2007.

Andrew T Ching. A dynamic oligopoly structural model for the prescription drug market after patent expiration. *International Economic Review*, 51(4):1175–1207, 2010.

Federico Ciliberto and Elie Tamer. Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828, 2009.

CMS. National Health Expenditures 2015 Highlights (retrieved from CMS website). URL:https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/highlights.pdf(accessed on Nov 18, 2017), 2015.

Lauren Cohen and Andrea Frazzini. Economic links and predictable returns. *The Journal of Finance*, 63(4):1977–2011, 2008.

Chintan V Dave, Aaron S Kesselheim, Erin R Fox, Peihua Qiu, and Abraham Hartzema. High generic drug prices and market competitiona retrospective cohort studygeneric drug prices and market competition. *Annals of Internal Medicine*, 2017.

Richard Davenport. Supply Chain on Bloomberg. Bloomberg L.P. https://www.bloomberg.com, September 2011.

Robert T Deacon and Jon Sonstelie. Rationing by waiting and the value of time: Results from a natural experiment. *Journal of Political Economy*, 93(4):627–647, 1985.

Department of Health and Human Services. Understanding Recent Trends in Generic Drug Prices. URL:https://aspe.hhs.gov/system/files/pdf/175071/GenericsDrugpaperr.pdf (accessed on July 17, 2017), 2017.

John C Driscoll and Aart C Kraay. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, 80(4):549–560, 1998.

DrugBank. The DrugBank database (Version 5.0). URL:https://www.drugbank.ca/about) (accessed on July 5, 2017), 2017.

Todd Ebert. The Generic Drug Market. *New York Times*, (accessed on July 5, 2017), Aug 20, 2016.

Anna Edney. FDA Warns India's Lupin Over Generic Drug Quality Violations. *Bloomberg*, (accessed on Nov 18, 2017), Nov 14, 2017.

R Eglin. Can suppliers bring down your firm? *Sunday Times (London)*, page 6, 2003.

Ezekiel J. Emanuel. Shortchanging Cancer Patients. *New York Times*, (accessed on July 10, 2017), Aug 6, 2011.

Robert F Engle and Victor K Ng. Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778, 1993.

Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

Eugene F Fama and James D MacBeth. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, pages 607–636, 1973.

FDA. Generic Competition and Drug Prices (retrieved from website). URL:https://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm129385.htm (accessed on July 17, 2017), 2005.

FDA. Guidance for Industry: Data Integrity and Compliance With CGMP (retrieved from FDA website). URL:https://www.fda.gov/downloads/drugs/guidances/ucm495891.pdf(accessed on Nov 17, 2017), 2016.

FDA. Abbreviated New Drug Application: Generics. URL:https://www.fda.gov/drugs/developmentapprovalprocess/howdrugsaredevelopedandapproved/approvalapplications/abbreviatednewdrugapplicationandagenerics (accessed on Sept 19, 2017), 2017a.

FDA. List of Off-Patent, Off-Exclusivity Drugs without an Approved Generic (retrieved from FDA website). URL:https://www.fda.gov/downloads/Drugs/ResourcesForYou/Consumers/BuyingUsingMedicineSafely/GenericDrugs/UCM564441.pdf(accessed on Nov 17, 2017), 2017b.

Adam J Fein. Pharmacy Profits Over the Generic Life Cycle. URL:http://www.drugchannels.net/2012/12/pharmacy-profits-over-generic-life.html (accessed on Nov 25, 2017), 2012.

Sheri Fink. Drug Shortages Forcing Hard Decisions on Rationing Treatments. *New York Times*, (accessed on July 17, 2017), Jan 29, 2016.

Richard G Frank and David S Salkever. Generic entry and the pricing of pharmaceuticals. *Journal of Economics & Management Strategy*, 6(1):75–90, 1997.

John Kenneth Galbraith. *American capitalism: The concept of countervailing power*, volume 619. Transaction Publishers, 1968.

A Ronald Gallant, Han Hong, and Ahmed Khwaja. Dynamic entry with cross product spillovers: An application to the generic drug industry. *Working Paper*, 2010.

GAO. Generic Drugs under Medicare GAO-16-706 (retrieved from GAO website). URL:https://www.gao.gov/products/GAO-16-706 (accessed on July 17, 2017), 2016a.

GAO. Drug Shortages GAO-16-595 (retrieved from GAO website). URL:https://www.gao.gov/products/GAO-16-595 (accessed on July 17, 2017), 2016b.

Vishal Gaur and Saravanan Kesavan. The effects of firm size and sales growth rate on inventory turnover performance in the US retail sector. In *Retail Supply Chain Management*, pages 25–52. Springer, 2008.

Scott Gottlieb. FDA Working to Lift Barriers to Generic Drug Competition, June 21, 2017.

Henry G Grabowski and John M Vernon. Brand loyalty, entry, and price competition in pharmaceuticals after the 1984 drug act. *The Journal of Law and Economics*, 35(2):331–350, 1992.

Jrg Hubert Grimm. *Ensuring Suppliers and Sub-Suppliers' Compliance with Corporate Sustainability Standards in Supply Chains*. PhD thesis, University of St.Gallen, Germany, 2013.

Jose A Guajardo, Morris A Cohen, and Serguei Netessine. Service competition and product quality in the us automobile industry. *Management Science*, 62(7):1860–1877, 2015.

Wayne R Guay. The sensitivity of ceo wealth to equity risk: an analysis of the magnitude and determinants. *Journal of Financial Economics*, 53(1):43–71, 1999.

Warren H Hausman, Hau L Lee, and Uma Subramanian. Global logistics indicators, supply chain metrics, and bilateral trade patterns. *World Bank Policy Research Working Paper*, (3773), 2005.

Kevin B Hendricks and Vinod R Singhal. The effect of supply chain glitches on shareholder wealth. *Journal of Operations Management*, 21(5):501–522, 2003.

Kevin B Hendricks and Vinod R Singhal. Association between supply chain glitches and operating performance. *Management Science*, 51(5):695–711, 2005a.

Kevin B Hendricks and Vinod R Singhal. An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *Production and Operations Management*, 14(1):35–52, 2005b.

Kevin B Hendricks, Vinod R Singhal, and Rongrong Zhang. The effect of operational slack, diversification, and vertical relatedness on the stock market reaction to supply chain disruptions. *Journal of Operations Management*, 27(3):233–246, 2009.

Bernard Herskovic, Bryan Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh. The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics*, 119(2):249–283, 2016.

Michael G Hertzel, Zhi Li, Micah S Officer, and Kimberly J Rodgers. Inter-firm linkages and the wealth effects of financial distress along the supply chain. *Journal of Financial Economics*, 87 (2):374–387, 2008.

Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, 20(8):832–844, 1998.

Daniel Hoechle et al. Robust standard errors for panel regressions with cross-sectional dependence. *Stata Journal*, 7(3):281, 2007.

Wallace J Hopp, Seyed MR Iravani, and Zigeng Liu. Strategic risk from supply chain disruptions. Working paper, Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI., 2008.

Arnd Huchzermeier and Morris A. Cohen. Valuing operational flexibility under exchange rate risk. *Operations Research*, 44(1):100–113, 1996.

Nitish Jain, Karan Girotra, and Serguei Netessine. Managing global sourcing: Inventory performance. *Management Science*, 60(5):1202–1222, 2013.

Japan METI. Summary of the white paper on manufacturing industries. Japan Ministry of Trade, Economy, and Industry. http://www.meti.go.jp, 2011.

Zhaohui Zoey Jiang, Yan Huang, and Damian R Beil. The role of feedback in dynamic crowdsourcing contests: A structural empirical analysis. *Working Paper*, 2016.

Morton I Kamien and Nancy L Schwartz. Market structure and innovation: A survey. *Journal of Economic Literature*, 13(1):1–37, 1975.

Sheila Kaplan. One reason for high drug prices: a huge backlog of unapproved generic drugs. *STAT*, (accessed on Nov 11, 2017), Dec 29, 2015.

Bryan Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh. Firm volatility in granular networks. Technical report, National Bureau of Economic Research, 2013.

Kenneth R. French Data Library. Developed market factors and returns. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, 2016.

Sang-Hyun Kim and Fiona M Scott Morton. A model of generic drug shortages: Supply disruptions, demand substitution, and price control. *Working Paper*, 2015.

Sang Won Kim, Marcelo Olivares, and Gabriel Y Weintraub. Measuring the performance of large-scale combinatorial auctions: A structural estimation approach. *Management Science*, 60(5): 1180–1201, 2014.

Edward Kong. Estimating competitive effects in firm entry with applications in the generic pharmaceutical industry. 2016.

Kyodo. Manufacturers expect lasting quake impact. The Japanese Times. http://www.japantimes.co.jp/news/2011/05/20/business/manufacturers-expect-lasting-quake-impact/, 2011.

Vivian Law, Craig Knox, Yannick Djoumbou, Tim Jewison, An Chi Guo, Yifeng Liu, Adam Maciejewski, David Arndt, Michael Wilson, Vanessa Neveu, et al. Drugbank 4.0: shedding new light on drug metabolism. *Nucleic acids research*, 42(D1):D1091–D1097, 2013.

Chia-Ying Lee, Xiaohan Chen, Robert J Romanelli, and Jodi B Segal. Forces influencing generic drug development in the united states: a narrative review. *Journal of Pharmaceutical Policy and Practice*, 9(1):26, 2016.

Hau L Lee, Kut C So, and Christopher S Tang. The value of information sharing in a two-level supply chain. *Management Science*, 46(5):626–643, 2000.

Jian Li, Shouyang Wang, and TC Edwin Cheng. Competition and cooperation in a single-retailer two-supplier supply chain with supply disruption. *International Journal of Production Economics*, 124(1):137–150, 2010.

Jun Li, Nelson Granados, and Serguei Netessine. Are consumers strategic? structural estimation from the air-travel industry. *Management Science*, 60(9):2114–2137, 2014.

Joshua N Liberman and M Christopher Roebuck. Prescription drug costs and the generic dispensing ratio. *Journal of Managed Care Pharmacy*, 16(7):502–506, 2010.

Yina Lu, Andrés Musalem, Marcelo Olivares, and Ariel Schilkrut. Measuring the effect of queues on customer purchases. *Management Science*, 59(8):1743–1763, 2013.

Vidya Mani, Saravanan Kesavan, and Jayashankar M Swaminathan. Estimating the impact of understaffing on sales and profitability in retail stores. *Production and Operations Management*, 24 (2):201–218, 2015.

Behdad Masih-Tehrani, Susan H Xu, Soundar Kumara, and Haijun Li. A single-period analysis of a two-echelon inventory system with dependent supply uncertainty. *Transportation Research Part B: Methodological*, 45(8):1128–1151, 2011.

Yoko Masui and Naofumi Nishi. Toyota's supply chain disruption in the great east japan earthquake. Technical report, Stanford University, Stanford, CA, 2012.

Don O May. Do managerial motives influence firm risk reduction strategies? *The Journal of Finance*, 50(4):1291–1308, 1995.

Milena McLaughlin, Despina Kotis, Kenneth Thomson, Michael Harrison, Gary Fennessy, Michael Postelnick, and Marc H Scheetz. Effects on patient care caused by drug shortages: a survey. *Journal of Managed Care Pharmacy*, 19(9):783–788, 2013.

Michelle Meadows. Greater access to generic drugs. New FDA initiatives to improve drug reviews and reduce legal loopholes. *FDA consumer*, 37(5):12–17, 2003.

Lior Menzly and Oguzhan Ozbas. Market segmentation and cross-predictability of returns. *The Journal of Finance*, 65(4):1555–1580, 2010.

Hokey Min. International supplier selection: a multi-attribute utility approach. *International Journal of Physical Distribution & Logistics Management*, 24(5):24–33, 1994.

Ken Moon, Kostas Bimpikis, and Haim Mendelson. Randomized markdowns and online monitoring. *Management Science*, 2017.

Fiona M Scott Morton. Entry decisions in the generic pharmaceutical industry. *The Rand journal of economics*, pages 421–440, 1999.

Marcelo Olivares, Christian Terwiesch, and Lydia Cassorla. Structural estimation of the newsvendor model: an application to reserving operating room time. *Management Science*, 54(1):41–55, 2008.

Marcelo Olivares, Gabriel Y Weintraub, Rafael Epstein, and Daniel Yung. Combinatorial auctions for procurement: An empirical study of the chilean school meals auction. *Management Science*, 58(8):1458–1481, 2012.

Luke Olson and Brett W Wendling. Estimating the effect of entry on generic drug prices using hatch-waxman exclusivity. 2013.

Nikolay Osadchiy, Vishal Gaur, and Sridhar Seshadri. Systematic risk in supply chain networks. *Management Science*, 62(6):1755–1777, 2015.

Mahmut Parlar. Continuous-review inventory problem with random supply interruptions. *European Journal of Operational Research*, 99(2):366–385, 1997.

Mahmut Parlar and Defne Berkin. Future supply uncertainty in eoq models. *Naval Research Logistics (NRL)*, 38(1):107–121, 1991.

Perspective Press. *The Pharmacy Technician*. Morton Publishing Company, 6 edition, 2016.

Mitchell A Petersen. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1):435–480, 2009.

Ivan PL Png and David Reitman. Service time competition. *The Rand Journal of Economics*, pages 619–634, 1994.

Andrew Pollack and Steve Lohr. A Japanese Plant Struggles to Produce a Critical Auto Part. *New York Times*, (accessed on June 4, 2018), Apr 27, 2011.

Lian Qi, Zuo-Jun Max Shen, and Lawrence V Snyder. A continuous-review inventory model with disruptions at both supplier and retailer. *Production and Operations Management*, 18(5):516–532, 2009.

Peter C Reiss and Pablo T Spiller. Competition and entry in small airline markets. *The Journal of Law and Economics*, 32(2, Part 2):S179–S202, 1989.

Peter H Ritchken and Charles S Tapiero. Contingent claims contracting for purchasing decisions in inventory management. *Operations Research*, 34(6):864–870, 1986.

María Jesús Sáenz and Elena Revilla. Cisco Systems, Inc.: Supply chain risk management. *The Supply Chain Management Casebook: Comprehensive Coverage and Best Practices in SCM*, page 80, 2013.

William Schmidt and Ananth Raman. When supply-chain disruptions matter. *Harvard Business School Working Paper*, (13-006), 2015.

David Sedgwick. Toyota asks suppliers for disaster plans. Automotive News. http://www.autonews.com/article/20140804/OEM10/308049947/toyota-asks-suppliers-for-disaster-plans, 2014.

Juan Camilo Serpa and Harish Krishnan. The impact of supply chains on firm-level productivity. *Management Science*, 2017.

ManMohan S Sodhi and Sunil Chopra. Managing risk to avoid supply-chain breakdown. *MIT Sloan Management Review*, 46(1):53–61, 2004.

Statista. Top 20 companies based on unbranded generics U.S. market share during 2014-2015. URL:https://www.statista.com/statistics/522095 (accessed on July 5, 2017), 2015.

Elie Tamer. Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, 70(1):147–165, 2003.

Christopher Tang and Brian Tomlin. The power of flexibility for mitigating supply chain risks. *International Journal of Production Economics*, 116(1):12–27, 2008.

Terry A Taylor. Supply chain coordination under channel rebates with sales effort effects. *Management Science*, 48(8):992–1007, 2002.

Torreya Partners. Generic Pharmaceutical Industry Yearbook. URL:https://www.fda.gov/ICECI/ComplianceManuals/CompliancePolicyGuidanceManual (accessed on July 5, 2017), 2016.

Jean-Franois Tremblay. Dr. Reddys struggles to meet FDA concerns. *Chemical & Engineering News*, (accessed on Nov 18, 2017), Aug 4, 2017.

Yixin Iris Wang, Jun Li, and Ravi Anupindi. Risky suppliers or risky supply chains? an empirical analysis of sub-tier supply network structure on firm risk in the high-tech sector. 2017. Working paper, Ross School of Business Paper No. 1297.

WHO. WHO Expert Committee On Specifications For Pharmaceutical Preparations (retrieved from WHO website). URL:http://apps.who.int/medicinedocs/pdf/h3009e/h3009e.pdf(accessed on Nov 17, 2017), 2002.

Janet Woodcock. Implementation of the Generic Drug User Fee Amendments of 2012 (GDUFA). URL:https://www.fda.gov/newsevents/testimony/ucm484304.htm (accessed on July 5, 2017), 2016.

Janet Woodcock and Marta Wosinska. Economic and technological drivers of generic sterile injectable drug shortages. *Clinical Pharmacology & Therapeutics*, 93(2):170–176, 2013.

Di Wu. *Shock Spillover and Financial Response in Supply Chain Networks: Evidence from Firm-Level Data*. Working paper, University of Pennsylvania, Philadelphia, PA, 2015.

Jing Wu and John R Birge. Supply chain network structure and firm returns, 2014. Working paper, University of Chicago, Chicago, IL.

Zhibin Yang, Goker Aydin, Volodymyr Babich, and Damian Beil. Using a dual-sourcing option in the presence of asymmetric information about supplier reliability: Competition vs. diversification. *Manufacturing & Service Operations Management*, 14:202–217, 2012.

Qiuping Yu, Gad Allon, and Achal Bassamboo. How do delay announcements shape customer behavior? an empirical study. *Management Science*, 63(1):1–20, 2016.

Fanyin Zheng. Spatial competition and preemptive entry in the discount retail industry. *Working Paper*, 2016.