Essays on Supply Chain Management: Risk Management and Productivity Spillovers

by

Juan Camilo Serpa

B.Sc., Trent University, 2009
M.A., The University of British Columbia, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Business Administration)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2015

Abstract

This thesis comprises three independent essays on supply chain management. In the first essay we collect data on 27,000 vertical relationships to study the importance of different channels of productivity spillovers between upstream and downstream firms. We explore the relative influence of two types of channels: endogenous and exogenous. The endogenous channel measures how a firm's productivity is affected by knowledge transfers (arising from collaboration and peer-mentoring). The exogenous channels measure the extent to which productivity is influenced by the partners' characteristics (e.g. geographic location, inventory turnover, financial leverage, etc.). We find that the endogenous channel is the primary source of spillovers. We also find that a firm's productivity is influenced more by the operational, than by the financial characteristics of its partners.

The second essay unveils a previously unexplored role of business insurance in managing supply chain risk. We show that firms may strategically buy insurance purely as a commitment mechanism to prevent excessive free-riding by other firms. Specifically, we show that contractual incentives alone leave wealth-constrained firms with low incentives to prevent operational accidents, and firms with sufficient wealth with excessive incentives. Insurance allows the latter firms to credibly commit to lower effort, thereby mitigating the incentives of the wealth-constrained firms to free-ride.

The third essay explores the interplay between public policy and risk management, when governments must strike a balance between safety and industry welfare. We focus on industries where operational accidents can be destructive and, as a result, where the cost of third-party liability is significant. Firms in these industries may be discouraged from entering the market as a result of these costs. If entry is inefficiently low, a social planner can incentivize firms through *ex ante subsidies*, which defray the costs associated with making operations safer, or *ex post subsidies*, which mitigate the financial damages caused by the accident. We demonstrate that when the social planner values reliability over market competition, it is optimal to offer ex ante subsidies alone. Conversely, when competition outweighs the benefits of reliability, a combination of ex ante and ex post subsidies is optimal.

Preface

Modified versions of Chapter 2, 3 and 4 have been submitted for publication. All essays are co-authored with Professor Harish Krishnan.

Professor Krishnan was involved in the early stage of the problem formulation. He also provided supervision and feedback in the design of the methodology, corrected technical and editorial mistakes, and suggested improvements in the exposition of the results. I was responsible for developing and writing most of the work found in this thesis and I take full responsibility for editorial and technical mistakes, if any are found.

Table of Contents

Abstract ii

Preface iii

Table of Contents iv

List of Tables vi

List of Figures vii

Acknowledgements viii

1 Introduction 1

2 The Impact of Supply Chains on Firm Level Productivity 4

2.1 Introduction 4
2.2 Research Background 8
2.3 Measuring Productivity 10
2.4 Productivity Spillovers 13
2.5 Extensions and Robustness Checks 28
2.6 Conclusions 37

3 The Strategic Role of Business Insurance in Managing Supply Chain Risk 39

3.1 Introduction 39
3.2 Literature Review 42
3.3 Model Preliminaries 43
3.4 Analysis 51
3.5 Extensions 63
3.6 Conclusion 72

4 Policy Incentives to Mitigate the Impact of Operational Tort Liability . . . 74
4.1 Introduction . . . 74
4.2 Literature Review . . . 78
4.3 Model . . . 79
4.4 Analysis . . . 84
4.5 Comparative Statics . . . 90
4.6 Extensions, Discussion and Conclusion . . . 93

5 Conclusions . . . 97

Bibliography . . . 99

Appendices

A Technical Results and Variable Definitions of Chapter 2 . . . 107
A.1 The Olley-Pakes Approach . . . 107
A.2 The Formation of Supply Chain Networks . . . 108
A.3 Variables Definition and Construction . . . 114

B Proofs for Chapter 3 . . . 116

C Proofs for Chapter 4 . . . 122

List of Tables

2.1 Estimates of the production function 12
2.2 Decile distribution of log-TFP 13
2.3 Summary statistics about the customer-base of each firm. . . 22
2.4 Summary statistics about the firm's networks. 22
2.5 Table of summary statistics 23
2.6 Model estimates for equation 2.3. 25
2.7 Summary statistics about the supplier-base of each firm. . . 30
2.8 Model estimates for equation (2.9) 31

3.1 Optimal tiered and simple contracts - Case $\mathcal{OC}$ 68
3.2 Optimal tiered and simple contracts - Case $\mathcal{SC}$ 68
3.3 Simulation of optimal tiered contracts for different damage variances (case $\mathcal{SC}$) 69
3.4 Simulation of optimal tiered contracts for different damage variances (case $\mathcal{OC}$) 69
3.5 Simulation of optimal insurance contracts with deductibles. . 70

A.1 Model estimates for the selection equation. 112
A.2 Model estimates for the linear-in-means equation. 113

List of Figures

2.1 Determinants of productivity. 5
2.2 Direct and indirect spillover effects. 6
2.3 Illustrative example of the model parameters 14
2.4 Estimation of indirect and direct spillover effects. 16
2.5 Networks and linear independence. 18
2.6 An illustration of Matrix **H** (from Figure 2.3). 21
2.7 Estimates of exogenous spillovers. 26

3.1 Model timeline 49
3.2 First best level of insurance. 52
3.3 Optimal parameters - case $\mathcal{SC}$, for $\beta = 0.4$, $x = 100$,$\pi = 45$. . 57
3.4 Welfare parameters - case $\mathcal{SC}$ 58
3.5 Optimal contract parameters - case $\mathcal{OC}$, for $\beta = 0.4$, $x = 100$,$\pi = 45$. 60
3.6 Welfare parameters - case $\mathcal{OC}$ 60
3.7 Contracting regions for the general case. 62
3.8 Optimal contract parameters for varying levels of α. 71

4.1 Model timeline 80
4.2 An illustrative example using an inverse demand function, $\Delta(Q) = 1 - Q$. 83
4.3 Illustration of the model regions. 85
4.4 A visualization of Lemma 12 88
4.5 Illustrative example with inverse demand function $P(Q) = 1 - Q$. 91
4.6 Optimal policy as a function of λ. 93

A.1 A sample network, where $c_{12} = c_{32} = 1$ and $c_{ij} = 0$ everywhere else. 110

Acknowledgements

My time at UBC has been rewarding and special. I am deeply grateful to Professor Harish Krishnan for his continual support. His tremendous dedication during my time in the Ph.D program was determinant. Harish introduced me to the field of supply chain management, and devoted a great deal of his time in helping me become a better researcher. Not only my thesis, but also my academic aptitude and writing style has improved substantially as a result of his effort.

I would also like to thank Professor Tim Huh, Professor Mahesh Nagarajan and Professor Ralph Winter, who served as committee members. They provided invaluable suggestions for the improvement of this thesis. Their guidance extends far beyond this document. Each of them helped me develop various academic skills. I am truly indebted to them.

I also wish to thank Professor Thomas Lemieux and Professor Robin Lindsey, who served as university examiners, and Professor Vishal Gaur, who served as the external examiner for this thesis. Their comments were very helpful during the final stages of this thesis. Their suggestions will help me improve these essays during the publication process. I am very appreciative for their kind help.

Many thanks to other faculty members at the Sauder School of Business for their support during various stages of this process. Finally, I would like to thank to those individuals who indirectly helped me pave the road that led towards the end of this process. This includes the Trent University International Program, for giving me the financial resources to come to Canada, and the University of British Columbia for its continual financial support during graduate school. Not to mention my friends, who kept nourishing my spirit.

My deepest thanks go to my family and to Hana. You were always there when I needed, and for this I couldn't be more thankful. I dedicate this thesis to you.

Chapter 1

Introduction

This thesis explores, through three independent essays, issues related to supply chain management. The first essay focuses on firm-level productivity (Chapter 2), while the remaining two essays focus on risk management issues (Chapters 3 and 4).

Firm-Level Productivity

The literature has shown large and persistent differences in firm-level productivity. Simply put, the economy is divided into two groups of firms: (i) those that have successfully managed to implement an efficient production process and; (ii) those that are lagging tremendously in the productivity distribution.

This gap has non-trivial implications. Firm-level productivity is one of the most reliable indicators, not only of firm success, but also of firm survival. As Krugman (1997) puts it, “productivity isn’t everything, but in the long-run it is almost everything.” This implies that productive firms are destined to be at the forefront of the economy. And those firms that are lagging are ultimately poised to go out of business.

What causes some firms to be so productive and, at the same time, other firms to be so unproductive? In other words, what drives firm-level productivity? Researchers have shown dozens of factors that can potentially affect this metric. For example, the level of market competition (Syverson, 2011), the level of IT (Brynjolfsson and Hitt, 2003) - even the weather conditions (Cachon et al., 2013)- can play a determinant role. The influence of supply chains partners, however, has not been carved out.

In Chapter 2 we explore the impact of vertical relationships on firm-level productivity. The main interest of this essay, however, is not to explore simply whether supply chain partners affect each other’s productivity. After all, when two firms interact in a supply chain, they engage in an intimate relationship characterized by collaboration, mutual dependency and mentoring. It is thus intuitive to expect that the productivity of a firm is affected by its choice of partners.

Our main interest in this essay is to explore the issue of *"how?"*. Specifically, what are the characteristics, actions or mechanisms through which partners in a supply chain end up affecting each other? Or more formally, through which channels does productivity spill over along the supply chain? For example, are firms affected via the operational, geographic or financial characteristics of the partners? Or, after holding these characteristics constant, are firms affected by virtue of interacting with more productive partners (e.g. via knowledge transfers)?

We explore this question through econometric methods. But we must first overcome several econometric challenges. For example, highly productive firms also tend to be operationally efficient, financially healthy, and be located in desirable geographies. How can we then disentangle the marginal impact of each characteristic from the others? Second, firms are likely to choose partners based on their strategic desire to become more productive. This will give rise to endogeneity in the formation of supply chain networks, and lead to biases in our results. Our essay develops a methodology to overcome the above econometric challenges and obtain consistent results.

Risk Management

The second part of this thesis focuses on supply chain risk management. We place special emphasis on high-impact/low-probability events (e.g. product recalls, oil spills, etc.). These types of events expose firms to financial distress, illiquidity or even bankruptcy. A firm, in the face of this possibility, must balance its exposure not only to operational risk, but also to financial risk.

To hedge financial risk, firms have access to financial instruments offered by third-party institutions. For example, depending on the situation firms can purchase business insurance or receive public subsidies. These financial instruments allow a firm to decrease its exposure to financial risk and, at the same time, to optimally manage its operational risk. Through two separate essays we study optimal risk management strategies in the presence of external financing instruments.

In Chapter 3 we explore risk management in the presence of business insurance. We show that insurance sometimes may serve a purely strategic role within the supply chain. Specifically, we show that when a large (and wealthy) firm collaborates with a small (and wealth constrained) partner, contractual tools alone leave the wealth-constrained firm with inefficiently low incentives to exert effort because it is unable to take on the appropriate level of liability. The opposite happens with the wealthy firm. Insurance can serve as a credible commitment mechanism that allows the wealthy firm to commit not to exert "too much" effort to prevent operational accidents. Interestingly, this mechanism improves the efficiency of the supply chain. By committing not to exert too much effort, the wealthy firm takes away the incentives of the wealth-constrained firm to free-ride on the "excessive" efforts of the wealthy firm. This balances the provision of effort in the supply chain and improves overall efficiency.

In Chapter 4 we focus on industries characterized by the potential for injury or harm to third parties due to operational accidents. In this type of setting the costs of liability can be significant. Not rarely, this condition leads to excessive firm exit in some industries, which leaves the markets monopolized (and decreases market inefficiency).

To prevent market failures governments often resort to offering incentives to the firms. There are two popular incentive schemes: ex ante and ex post subsidies. Ex ante subsidies offer incentives to the firms *prior* to commencing operations, to defray the costs of making operations more reliable. For example, governments can offer grants to purchase safety equipment or to hire skilled personnel.

Ex post subsidies are given *after* firms commence operations, and are conditional on the occurrence of an accident. These types of subsidies decrease the firm's exposure to the ex post costs of an accident (e.g. clean up costs, litigation costs). For example, governments sometimes offer liability caps to limit the cost of the accident for the firm.

In this essay we explore the conditions under which it is socially optimal for a social planner to offer ex ante or ex post subsidies. We find that when the planner values reliability over competition, the socially optimal policy is to offer ex ante subsidies alone. In the converse scenario, we show, it is optimal for the planner to offer a combination of ex ante and ex post incentives.

Chapter 2

The Impact of Supply Chains on Firm Level Productivity

2.1 Introduction

Syverson (2011) reviews an extensive literature that shows large and persistent differences in firm-level productivity, even "within narrowly defined industries." But what explains these differences? Syverson notes two broad categories of factors that drive firm-level productivity: (i) factors that are internal to the firm (such as research and development, managerial practices and talent) and; (ii) factors that are external to it (such as regulations and product market competition).

Though researchers have made great strides in determining the drivers of this productivity dispersion, the role of supply chains remains unclear. Even so, we can reasonably conjecture that supply chain relationships play a non-trivial role. After all, firms in a supply chain collaborate, communicate and influence each other's processes. Consider Wal-Mart that, by implementing a sophisticated distribution system, improved the flow of information and, consequently, the efficiency of its supply chain partners (Brynjolfsson, 2003). Similarly, Dell created an integrated network that improved the ability of its suppliers to better match supply and demand (Fillard et al., 2011).

In this essay we explore the role of supply chain linkages in influencing firm-level productivity, by considering two key *channels* through which productivity can spill over across firms.

First, a firm can benefit from interacting with partners that have "favorable" characteristics (independent of the partners' productivity). In the literature, these are known as *exogenous* channels. As a case in point, consider a firm's geographic location. When a supplier is located in a favorable region, it may be able to ship inputs more efficiently due to the existence of better transportation infrastructure, commercial regulations or climatic conditions. A supplier's location by itself, therefore, can affect the productivity of its customers.

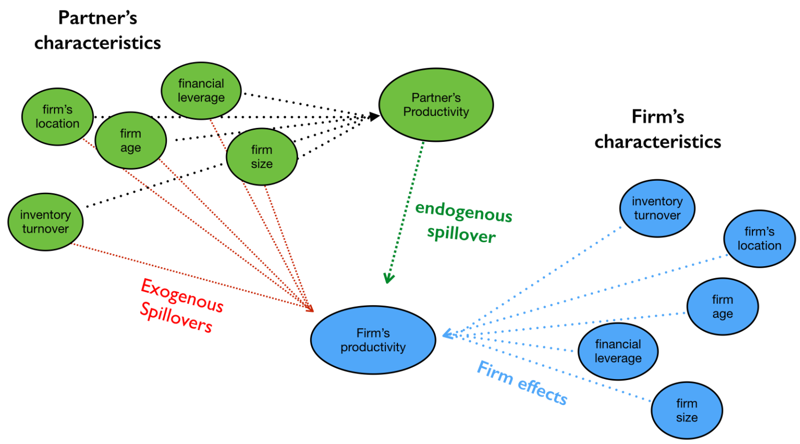


Figure 2.1: Determinants of productivity.

Second, firms can benefit from interacting with productive partners (independent of the partners' characteristics). This is known as the *endogenous* channel. As in the Wal-Mart and Dell anecdotes, productive firms can influence the operations of their peers through mentoring or collaboration. And even in the absence of mentoring or collaboration, firms may learn from, and adopt, the good practices of their partners.

In this context, a firm's productivity is thus affected by three types of effects: (i) by the firm's own characteristics; (ii) by the characteristics of its partners (through the exogenous channels) and; (iii) by the productivity of its partners (through the endogenous channel) - see Figure 2.1.

By estimating these three types of effects, we obtain a nuanced picture about how partners influence each other's productivity. Consider, for example, a given characteristic of a firm. We can determine the direct and the indirect impact of this characteristic on the productivity of the supply chain partners. To illustrate this point, suppose that a change in inventory turnover affects a firm's own productivity. A change in the inventory turnover can, in addition, affect the productivity of the partners in two ways: (i) *directly* by the change in inventory turnover (the exogenous channel) and; (ii) *indirectly,* because of the change in the productivity of the firm (the endogenous channel) - see Figure 2.2.

To estimate the impact of supply chains on firm-level productivity, we collect a sample of approximately 27,000 supply chain relationships. These data are from publicly-traded firms in the U.S. and are available due to a requirement that if a customer exceeds 10% of a firm's annual revenue, the

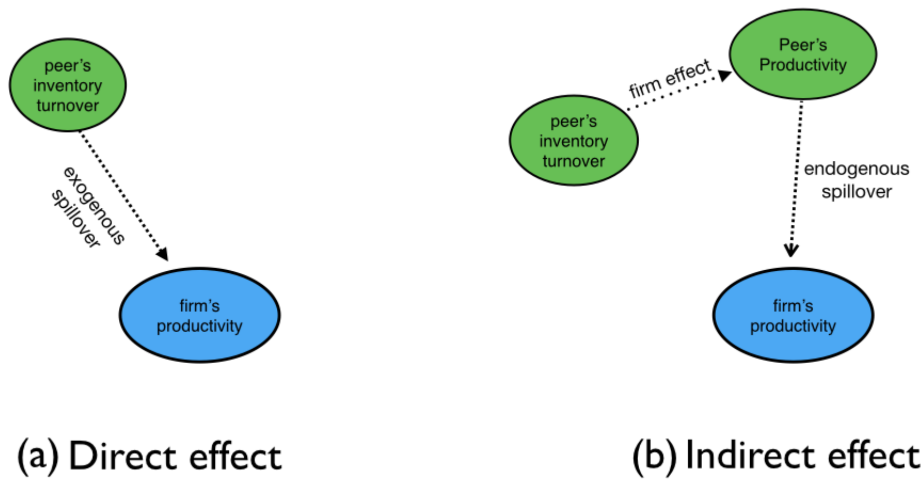


Figure 2.2: Direct and indirect spillover effects.

firm must disclose these sales and the name of the customer. We merge this dataset to key idiosyncratic information about each firm, from the Compustat database.

Our estimates find that the endogenous spillover effect is significantly larger than any of the exogenous spillover effects. This means that interacting with productive firms is relatively more important than interacting with firms with "favorable" characteristics. We find several other interesting results (some of which are highlighted below):

- The size of a firm has two counteracting effects. First, larger firms are more productive themselves (this can be explained by the presence of scale economies). Therefore, firms *indirectly* benefit from interacting with large partners. However, holding the partners' productivity fixed, firms *directly* benefit from having smaller partners. This can be explained by the idea that a firm has more influence over the managerial decisions of smaller customers (or suppliers), and this can benefit the firm.

- The effect of firm age on its partners is inversely U-shaped. Previous findings in the literature had already found an inverse U-shaped relationship between firm age and productivity (Van Biesebroeck, 2005; Fernandes, 2008). But we show that, in addition, the spillover effect follows this same pattern. This implies that it is more efficient to interact with partners that are neither too young nor too old. Specifically, firms benefit most from interacting with partners aged 17-20 years.

- Firms located in the U.S. are more productive than foreign firms, and this indirectly benefits the partners of U.S. firms. However, keeping the partners' productivity levels fixed, it is beneficial to interact with foreign firms. This last result is related to the trade literature, which suggests that U.S. firms benefit from technology spillovers via imports (Keller and Yeaple, 2009).

- While a firm's inventory turnover affects its partners directly *and* indirectly, financial leverage only affects them indirectly. This result hints that a firm's productivity is more susceptible to the operational, than to the financial characteristics of its partners.

Before arriving at the above results, we had to deal with several identification problems that afflict the "peer-effects" literature. The most prominent is the *reflection problem*, which arises due to two identification issues: (i) the "correlated environment problem" and; (ii) the "entanglement problem". The correlated environment problem arises because firms that form links often share common geographic, economic or technological environments. Correlated productivity levels might thus represent the impact of (unobserved) common shocks, and not the impact of spillover effects. But even when these shocks are absent, the reflection problem does not disappear. This is due to the entanglement problem. This problem arises because the endogenous and the exogenous channels are inherently entangled. In other words, whenever the characteristics of a firm change so does its productivity, i.e. these effects are perfectly collinear. This condition makes it extremely difficult to disentangle the true impact of a firm's exogenous characteristics from the impact of the firm's productivity level on the partners. Therefore, this type of identification is often considered one the most challenging econometric tasks in the literature (Angrist and Pischke, 2009; Jackson, 2010).

We use a novel identification strategy to overcome the reflection problem. This strategy is based on Bramoullé, Djebbari and Fortin (2009) , who show that we can overcome the reflection problem when the network of relationships satisfies a "partially-overlapping" structure. Using this result, we exploit the structure of the supply chain networks to identify our model. Our approach uses a series of generalized two-stage least squares estimators, which extract information from multiple echelons in the supply chain networks. Our identification approach is strong because we rely on the internal structure of our data (and not on external instruments) to derive our estimates.

To test the robustness of our main results, we re-estimate them by varying some model constructs. We were particularly concerned with the fact that a firm may be inclined to choose partners with specific characteristics, e.g. geographic location or firm size. Hence, we need to consider the possibility that our results are correlated with the choice of supply chain partners, and this would bias our estimates. To address this concern, we re-estimate our results by controlling for this factor. We also vary other constructs, such as the specification of the production function, the variable definitions, the industry classification, etc. Our conclusions are robust to these checks.

This manuscript has two main contributions. From a practical standpoint, our results help us understand how productivity is channeled across the supply chain. We provide evidence about the influence of various characteristics, and we also show that interacting with productive partners (the endogenous channel) is the main source of productivity spillovers - more so than any exogenous channel. Our results can help practitioners at the time of building supply chain relationships.

The second contribution is that we are (to our knowledge) the first to jointly identify endogenous and exogenous spillover effects using data on firm networks. To accomplish this, we overcome several identification issues by using recent techniques drawn from the peer effects literature. We provide a novel framework to estimate spillover effects across supply chains. This framework can guide future research on similar problems. As such, this manuscript bridges the literature on peer effects and supply chains.

The remainder of this article is organized as follows. In the next section we review the literature. We then estimate the production function in §2.3 and, using this function, the productivity of each firm. In §2.4 we estimate the influence of supply chain linkages on firm-level productivity. In §2.5 we study a few extensions and robustness checks. We conclude in §2.6.

2.2 Research Background

This essay is related to two streams of empirical research: (i) firm-level productivity and; (ii) operational performance and supply chain linkages.

A survey of the literature on firm-level productivity is provided by Syverson (2011). We can divide this literature into two groups. The first group uses data from a single industry and, in this case, productivity is often measured through inputs and outputs that are specific to the industry. For example, Kellogg (2011) collects data on the oil-drilling industry to estimate productivity gains among firms that collaborate through long-term contracts. Cachon et al. (2013) use data from automobile plants to estimate the impact of severe weather on productivity. Chandra et al. (2012) use Medicare data to explain productivity differences across U.S. hospitals.

The second group does not focus on studying the productivity of a particular industry but, rather, on making conclusions that apply across large sectors of the economy. The problem is that inputs and outputs vary tremendously across industries. To avoid inconsistencies, researchers often gauge productivity using the "value-added" approach. In this approach inputs consist of capital and employment, and output is measured as value-added. Our essay uses this approach. To construct some variables, we follow the methodology adopted by Brynjolfsson and Hitt (2003), who use the Compustat dataset to measure the relationship between firm-level productivity and IT.

When we measure productivity, we must worry about simultaneity and selection biases (these biases are explained in §2.3). We control for these issues by adopting a widely used identification strategy developed by Olley and Pakes (1996).

There is also a stream of empirical research dedicated to studying the impact of supply chain linkages on a firm's operational performance. For example, Kalwani and Narayandas (1995) report that suppliers observe higher returns on investment when they are engaged in long-term relationships with their customers. Hendricks and Singhal (2005) study the effect of supply chain disruptions on long-run stock price performance. In the global sourcing literature, Jain, Girotra, and Netessine (2013a) study the impact of sourcing relationships on inventory. The authors show that when firms source from foreign suppliers, their inventory investments increase. Lieberman and Demeester (1999) study productivity growth when suppliers and manufacturers collaborate through Just-In-Time delivery. The authors use data on the Japanese auto industry.

We contribute to this literature by studying the different channels through which productivity spills over across firms. To our knowledge, this type of analysis has not been done previously, perhaps because of the serious identification challenges encountered in this literature. The most prominent identification issue is the reflection problem, a problem first studied by Manski (1993). The reflection problem has become a central issue when estimating spillover effects. Fortunately, the literature has come a long way over the past years. A review of these developments can be found in Blume et al. (2010).

In this essay we exploit recent results found by Bramoullé, Djebbari, and Fortin (2009), who show that it is possible to overcome the reflection problem if the network of relationships satisfies a particular structure (which involves partially overlapping interactions). Our data satisfies the necessary structure and, therefore, we can control the reflection problem. We use the estimators proposed by Lee (2007) to obtain our estimates. These estimators are asymptotically optimal for estimating spillover effects in network interactions.

2.3 Measuring Productivity

2.3.1 Econometric Specification

We use Total Factor Productivity (TFP) as our measure of productivity. TFP is desirable because it is invariant to the intensity of use of observable input factors (Syverson, 2011). To obtain our estimates, we use an approach similar to Imrohoroglu and Tuzel (Imrohoroglu and Tuzel, 2014).

Consider a log-linear Cobb-Douglas production function[1]

$$y_{irt} = \alpha_{rt} + \beta_k k_{irt} + \beta_l l_{irt} + \rho_{irt} + \varepsilon_{irt} \tag{2.1}$$

where y_{irt} is the natural logarithm of value-added for firm i in industry r and year t, while k and l represent the log of capital and labour. Parameter α_{rt} measures the industry fixed-effect on output; we classify firms into an industry according to their 3-digit SIC code. The term ρ_{irt} represents the firm's TFP, which can be interpreted as the relative productivity rank of a firm within its industry. The term ε_{irt} is normally distributed random shock. Therefore, if we let $\hat{\alpha}_{rt}, \hat{\beta}_l, \hat{\beta}_k$ denote the estimates of the production function, we have that

$$\hat{\rho}_{irt} = y_{irt} - \hat{\alpha}_{rt} - \hat{\beta}_l l_{irt} - \hat{\beta}_k k_{irt}$$

is the estimated log TFP of firm i at time t.

2.3.2 Identification Strategy

To measure log-TFP, we use the value-added approach, where inputs consist of capital and labour and output is measured as value-added. Although this is one of the most popular approaches to measure productivity, two

[1] We obtain this function after taking logarithms of $Y = AK^{\beta_K} L^{\beta_L} e^{\varepsilon}$.

key problems arise when estimating TFP through OLS: simultaneity and selection biases.

A *simultaneity* bias arises because the input factors and the observed productivity level are simultaneously determined. In these cases, there is correlation between the inputs and the error term. For example, if a firm observes a favorable shock (that is unobserved by the econometrician), the firm knows that it needs less labour to produce a given level of output. As a result, the firm will decide to hire less, and OLS will overestimate β_l.

A *selection* bias arises because a firm's profitability is correlated with its level of capital stock, which is fixed in the short term. That is, a firm with larger capital stock is less likely to exit the market (despite low productivity draws). This is because the firm expects to earn greater profits in the future. Given the negative correlation between capital stock and the exit probability, productivity is also correlated to the capital stock, and OLS underestimates β_k.

To control for both of these biases, we use an estimation method proposed by Olley and Pakes (1996). We explain this methodology in Appendix A. Although this approach has become a standard one in the productivity literature, a drawback is that we must use *positive* capital investments as a proxy variable. In principle the OP approach can still be used in these situations, although it requires that we ignore all data points with non-positive capital investments. Fortunately, in our sample less than 2.6% of the firms report non-positive capital investments.[2]

There are other "garden variety" problems with the value-added approach (see Griliches and Mairesse 1997). For example, there is lack of information on the quality dimensions and on the utilization of these variables, and productivity is sensitive to output prices. Fortunately, productivity estimates are "likely quite robust to [these] measurement peculiarities. The inherent variation in establishment micro-data is so large as to swamp any small measurement-induced differences in productivity metrics (Syverson, 2011)." As such, we treat these problems as minor data limitations (like virtually every paper in the literature). But in this point we do more than assuming, and we test the robustness of our productivity estimates in §2.5.3.

[2] In some studies, the Olley-Pakes approach is undesirable because the proportion of observations with non-positive capital investments is large. To deal with this problem, Levinsohn and Petrin (2003) created a very similar methodology that uses intermediate materials (instead of capital investments) to solve the simultaneity bias. This methodology, however, does not control for the selection bias.

Elasticity (log)	Olley-Pakes	OLS
Capital (β_k)	0.4492	0.4138
	(0.0041)	(0.0019)
Labour (β_l)	0.5986	0.6305
	(0.0067)	(0.0016)
Scale elasticity	1.0191	1.0435

Table 2.1: Estimates of the production function

2.3.3 Data Description

To estimate TFP, we use data from the Compustat database, which contains financial information about publicly-traded firms. Each firm-year observation reports data on annual sales, levels of inventory, assets, liabilities, etc. We use datasets from the NBER-CES Manufacturing Industry Database to retrieve price deflators for capital stock, materials and output. We also retrieve data from the U.S. Bureau of Labor Statistics to obtain labour related data (i.e. average compensation costs and deflator prices). Our methodology to construct the variables can be found in the Appendix section.

Before building our dataset, we applied a filtering mechanism to discard useless observations. First, we dropped all observations reported before 1962 (due to well-known reporting biases). We also discarded firms with SIC code 99, as these are unclassified establishments. Because TFP is an industry-specific measure, these observations are useless. Finally, we deleted observations reporting non-positive value added, number of employees, capital investment gross property, sales or plant and equipment. Using the data above, we constructed an unbalanced panel containing 22,133 distinct firms and 137,864 firm-year observations. Note that, because we use data from Compustat, our analysis focuses on publicly-traded firms.

2.3.4 Estimates

We report the elasticity estimates in Table 2.1 (using the Olley-Pakes approach). For comparison purposes, we also report the OLS estimates of the production function. Though not reported, our regression includes dummies to control for industry fixed effects.

Our results show that OLS underestimates the capital stock elasticity and overestimates the labour elasticity. These results allow us to appreciate the impact of simultaneity and selection biases that we discussed earlier. We can also observe that firms face (mild) increasing returns to scale, given that

	Percentile					
	10	**20**	**40**	**60**	**80**	**90**
log(TFP)	0.281	0.436	0.579	0.69	0.82	0.944
$90^{th} - i^{th}$**percentile ratio**	1.941	1.662	1.44	1.289	1.132	

Table 2.2: Decile distribution of log-TFP

the scale elasticity is greater than one (i.e. $\beta_k + \beta_l > 1$).

As mentioned above, we test the robustness of our estimates in §2.5.3. In addition, we compared our results with those found in similar studies. To this end, we looked at the decile distribution of log(TFP) and the inter-decile productivity ratio (see Table 2.2). This ratio determines the degree to which a firm in the 90^{th} percentile of the productivity distribution is more efficient than an n^{th} percentile firm. For example, a ratio equal to two would imply that a 90^{th} percentile firm produces twice as much output (with the same degree of measured inputs) as an n^{th} percentile plant. Syverson (Syverson, 2004) reports that the $90^{th} - 10^{th}$ percentile ratio is equal to 1.92, a ratio that is almost identical to ours. In our case, this ratio is equal to 1.941. Imrohoroglu and Tuzel (Imrohoroglu and Tuzel, 2014) report a $90^{th} - 20^{th}$ dispersion of 1.8, while, in our case, the dispersion is equal to 1.662.

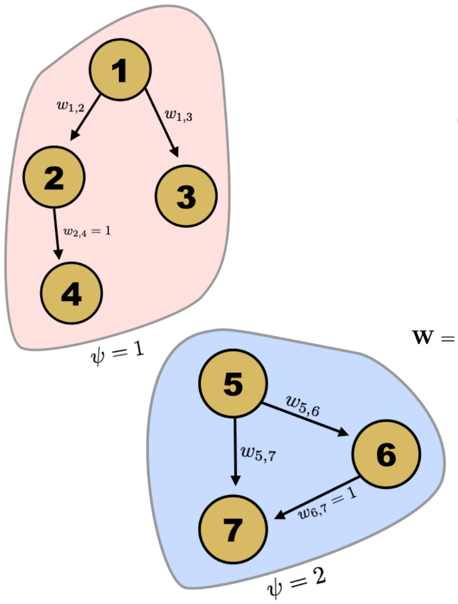
2.4 Productivity Spillovers

In this section, we construct an econometric model to jointly estimate the endogenous and the exogenous spillover effects on TFP. Due to data limitations (which are explained in §2.4.5), we restrict our main analysis to the firm's customer-base. In §2.5.1, however, we estimate spillover effects for a firm's supplier-base.

The modeling presented below borrows terminology from Sacerdote (2001) and from Bramoullé, Djebbari, and Fortin (2009).

2.4.1 Setup and Notation

Consider a sample of N *firm-year* observations, $\mathcal{N} = \{1, 2, \ldots, N\}$. Each $i \in \mathcal{N}$ is characterized by a value for TFP_i and a K-vector $\mathbf{X}_i = [x_{1i}, \ldots, x_{Ki}]$ of characteristics that could affect TFP_i. Each observation is also characterized by a set of customers. Let $\mathbf{C}$ be an $N \times N$ matrix with typical element c_{ij}, where c_{ij} equals one if j is a customer of i, and zero otherwise. We assume that $c_{ii} = 0$. Let $\mathbf{W}$ be an $N \times N$ matrix characterizing the strength of



$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 0 & w_{12} & w_{13} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & w_{24} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{56} & w_{57} \\ 0 & 0 & 0 & 0 & 0 & 0 & w_{67} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 2.3: Illustrative example of the model parameters

the relationships between a firm and each of its customers. This matrix has typical element

$$w_{ij} = \begin{cases} \frac{c_{ij} \cdot (sales_{ij})}{\sum_{j=1}^{N} c_{ij} \cdot (sales_{ij})} & \text{if } c_{ij} = 1 \\ 0 & \text{otherwise} \end{cases}$$

where $sales_{ij}$ is the revenue received by firm i from customer j. Using graph theory, we can represent firms as nodes, and their relationships through weighted edges. As such, $\mathbf{W}$ is the weighted adjacency matrix of the graph.

Two firms belong to the same network if they can be connected through an undirected path of relationships, that is, if they are (weakly) connected in the graph. We use ψ to index the networks, and assume that there are Ψ networks in total, each with size L_ψ.[3] See Figure 2.3 for an illustration for a network with $N = 7$ firms and $\Psi = 2$ networks (the networks have size $L_1 = 4$ and $L_2 = 3$ respectively).

[3]Because each $i \in \mathcal{N}$ represents a firm-year observation, we may be able to find two networks that are comprised by the same firms, but at different points in time.

2.4.2 Econometric Specification

We specify a linear-in-means model, where

$$\begin{aligned} TFP_i &= \beta + \underbrace{(\gamma_1 x_{1i} + ... + \gamma_K x_{Ki})}_{\text{firm effects}} + \underbrace{(\theta_1 x_{1i}^C + ... + \theta_K x_{Ki}^C)}_{\text{exogenous spillovers}} \\ &+ \underbrace{\omega TFP_i^C}_{\text{endogenous spillover}} + u_i \end{aligned} \tag{2.2}$$

Here, $x_{ki}^C \equiv \sum_j w_{ij} x_{kj}$ denotes the weighted average characteristics of i's customers and $TFP_i^C \equiv \sum_j w_{ij} TFP_j$ denotes their average productivity. The error term is represented by u_i, which we assume to be normally distributed with $E\left[u_i | \mathbf{X}_i, \mathbf{W}_i\right] = \delta \mathbf{X}_i$ for some K-vector δ.

The vector $\gamma = [\gamma_1, ..., \gamma_K]'$ represents the firm effects; $\theta \equiv [\theta_1, ..., \theta_K]'$ represents the exogenous spillover effects, i.e. the impact of a customer's characteristic on a firm's TFP and; ω represents the endogenous spillover effect, i.e. the impact of a customer's TFP on a firm's TFP. Finally, δ represents the correlated effects. If $\delta \neq \mathbf{0}$, this implies that the customers and the firm are affected by common (economic, geographic or technological) shocks.

If we let $\mathbf{TFP}$=$[TFP_1.....TFP_N]'$, $\mathbf{X} = [\mathbf{X_1 ... X_N}]'$ and $\mathbf{1}$ be an N×1 vector of ones, the matrix representation of our model is

$$\mathbf{TFP} = \mathbf{1}\beta + \mathbf{X}\gamma + \mathbf{WX}\theta + (\mathbf{W})\mathbf{TFP}\omega + \mathbf{u} \tag{2.3}$$

where $E[\mathbf{u}|\mathbf{X}, \mathbf{W}] = \mathbf{X}'\delta$.

These estimates give very detailed descriptions about the spillovers effects. We can estimate the *direct* impact of a customer's k^{th} characteristic on its supplier (through the exogenous spillover effect, θ_k). We also estimate how a given characteristic affects a firm's own TFP (through γ_k), and in turn, how this change *indirectly* affects the supply chain partners through the endogenous effect, ω. Therefore, the indirect spillover effect is $\gamma_k \omega$. For example, in figure 2.4 we look at a scenario where the firm has one customer. If firm j increases in size (by one unit), i's TFP will receive a *direct* spillover equal to θ_{size}. At the same time, a unit change in the size of the customer will affect the productivity of this customer by γ_{size}. This change, in turn, will affect i's TFP firm through the endogenous effect, ω. The size of the *indirect* spillover effect is $\gamma_{size}\omega$. Therefore, the total spillover effect is $\theta_{size} + \gamma_{size}\omega$.

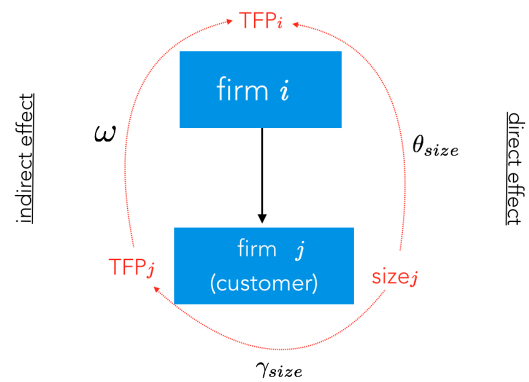


Figure 2.4: Estimation of indirect and direct spillover effects.

Unfortunately, the estimation of these effects is considered one of the most daunting econometric tasks, and in many cases an impossible one. One of the most challenging problems is the reflection problem.

2.4.3 The Reflection Problem

In his seminal paper, Manski (1993) shows that our linear-in-means model cannot be identified through OLS. This is because of two identification issues: (i) the "correlated environment" problem and; (ii) the "entanglement" problem. The first issue arises because, in a supply chain, firms often share common geographic, economic or technological environments. Correlated shocks are present when $\delta \neq \mathbf{0}$. If the presence of these effects is not controlled for, the econometrician may establish a spurious relationship between the productivity of a firm and the firm's customers.

But even when this problem is absent, i.e. $\delta = 0$, it is not possible to identify our model through OLS. This is because of the entanglement problem, which is the more complicated problem. This issue arises because the endogenous effect cannot be disentangled from the exogenous effects. To see this, observe that given equation (2.3), we can write

$$\mathbf{E}[\mathbf{TFP}|\mathbf{X},\mathbf{W}] = \mathbf{1}\beta + E[\mathbf{X}|\mathbf{W}]\gamma + \mathbf{WX}\theta + \mathbf{E}[(\mathbf{W})\mathbf{TFP}|\mathbf{X}]\omega \tag{2.4}$$

Therefore, we can eliminate $\mathbf{(W)TFP}$ by rewriting (2.3) as

$$\begin{aligned} \mathbf{TFP} &= \beta(\mathbf{I}-\omega\mathbf{W})^{-1}\mathbf{1} + (\mathbf{I}-\omega\mathbf{W})^{-1}(\mathbf{X}\gamma + \mathbf{WX}\theta) \qquad (2.5)\\ &\quad +(\mathbf{I}-\omega\mathbf{W})^{-1}\mathbf{u} \\ &= \mathbf{1}b_1 + \mathbf{Xb_2} + \mathbf{WXb_3} + \mathbf{u_o} \qquad (2.6) \end{aligned}$$

Equation (2.5) is known as the reduced form equation. The reduced form equation and equation (2.3) are informationally equivalent.

The reduced form estimates $(b_1, \mathbf{b_2}, \mathbf{b_3})$ cannot be mapped onto $(\beta, \gamma, \theta, \omega)$ because there is insufficient information. This is because the TFP and the exogenous regressors are linearly dependent, i.e. as the exogenous characteristics vary so does TFP. For this reason, the exogenous effects cannot be identified apart from the endogenous effects.

Fortunately, the reflection problem does not completely close the door for identification. There are some alternatives to dealing with this problem. In this essay we use a novel identification strategy that exploits partially overlapping network interactions. This approach uses the structural information of our model instead of relying on exogenous instruments. This feature makes our estimation approach particularly strong.

2.4.4 Identification Strategy

Bramoullé, Djebbari, and Fortin (2009) show that identification is possible if the network interactions satisfy a structure involving partially overlapping "peer" groups. This is because each agent has a peer group that is different from the peer group of every other agent. Under some conditions (that we explain below), we can exploit these differences by instrumenting the exogenous "peer of peer" effects on the peer effects, which allows us to disentangle the exogenous effects from the endogenous effects. This approach does not impose stringent assumptions on the structure of the data, which makes our identification technique robust.

Bramoullé, et al.'s approach can be applied under two scenarios: one where correlated effects are absent, and one where the correlated effects are fixed across the network. We derive our estimates under these two assumptions.

No Correlated Effects

In this section, we assume that $\delta = \mathbf{0}$ or, equivalently, that $E[\mathbf{u}|\mathbf{X}, \mathbf{W}] = 0$. In the absence of correlated effects, Proposition 1 in Bramoullé, Djebbari and Fortin (2009) show that if the identity matrix, **I,** and the matrices **W** and $\mathbf{W}^2$ are linearly independent, the reduced form estimates from equation (2.6) can be mapped onto $(\beta, \gamma, \theta, \omega)$. This means that our model can be identified. For example, in network (a) from Figure 2.5 identification is not possible because $\mathbf{I}$,$\mathbf{W}$ and $\mathbf{W^2}$ are linearly dependent. In network (b) identification is possible because the matrices exhibit linear independence

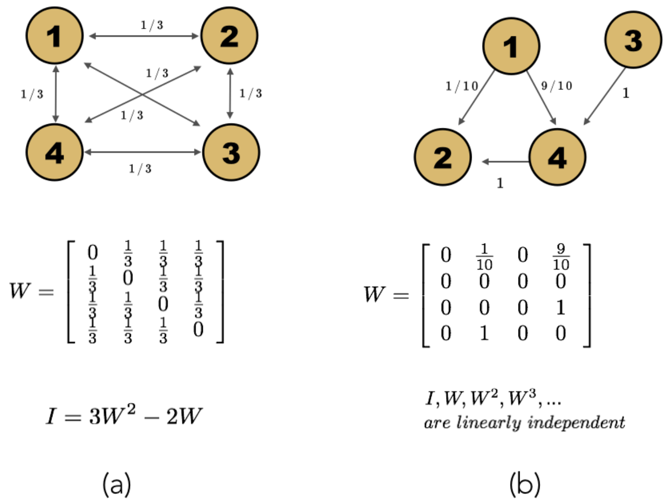


Figure 2.5: Networks and linear independence.

The intuition behind this result is the following: matrix $\mathbf{W^2}$ describes the weighted relationships between a firm and the "customers of its customers."[4] If $\mathbf{I}$, $\mathbf{W}$ and $\mathbf{W^2}$ are linearly independent, then any $\mathbf{W^nX}$ (for $n = 2, 3, \ldots$) can serve as an identifying instrument. That is, the third or fourth echelon of customers can also be used to instrument for the first echelon.

In Figure 2.5 we illustrate this insight. In the network from Figure 2.5(a) identification is not possible because there is a complete overlap in the network.[5] In other words, all firms derive one third of their revenue from each partner. Conversely, in the network from Figure 2.5(b) it is possible to identify the model. This is because the (partially overlapping) network structure of 2.5(b) allows us to extract the unique impact of a firm on its partners. For example, the "influence" (i.e. the weight) of Firm 2 is higher on Firm 4 than on Firm 1; the influence of Firm 4 is higher on Firm 3 than on Firm 1, etc. We exploit these differences to tease out the exact impact of each characteristic on the supply chain partners.

To check that $\mathbf{I}, \mathbf{W}$ and $\mathbf{W^2}$ are linearly independent, we use Corollary 1 in Bramoullé, Djebbari and Fortin (2009). According to this corollary, linear independence is guaranteed if the diameter of the network is greater than

[4]Similarly, $\mathbf{W^3}, \mathbf{W^4}, \ldots$ represent the relationship between the firms and lower customer echelons in the network.

[5]The "star" and "ring" networks, and bipartite networks are other types of structures characterized by this type of linear dependence.

three; from the data, we found that the diameter of our network is twelve.[6] Hence, we can identify our estimates.

After verifying linear independence, we proceed to identify equation (2.3). To this end we use a series of generalized Two-Stage Least Squares (2SLS), based on an approach proposed by Lee (Lee, 2007). This estimator is desirable because it is asymptotically optimal under i.i.d. errors.[7] To apply this estimator, we follow two steps:

Step 1 We begin by estimating an over-identified 2SLS model. We use $\mathbf{X}$, $\mathbf{WX}$, and $\mathbf{W^2X}$ as first-stage instruments (to instrument for $(\mathbf{W})\mathbf{TFP}$). Then, we use $(\mathbf{W})\mathbf{TFP}$, $\mathbf{X}$ and $\mathbf{WX}$ as regressors in the second-stage equation. This specification allows us to obtain the estimates $(\beta^{2SLS}, \gamma^{2SLS}, \theta^{2SLS}, \omega^{2SLS})$.

Step 2 We use the estimates obtained in step 1 to estimate the expectation of equation (2.5)- the reduced form equation:

$$\mathbf{E}[(\mathbf{W})\mathbf{TFP}^{\mathbf{2SLS}}|\mathbf{X},\mathbf{W}] \\ = \mathbf{W}(\mathbf{1}-\omega^{\mathbf{2SLS}}\mathbf{W})^{-1}[\beta^{2SLS}\mathbf{1}+\mathbf{X}\gamma^{\mathbf{2SLS}}+\mathbf{WX}\theta^{\mathbf{2SLS}}]$$

Next, we specify the second 2SLS model. This model uses $\mathbf{E}[(\mathbf{W})\mathbf{TFP}^{\mathbf{2SLS}}|\mathbf{X},\mathbf{W}]$, $\mathbf{X}$ and $\mathbf{WX}$ as instruments (once again, to instrument for $(\mathbf{W})\mathbf{TFP}$). We use $(\mathbf{W})\mathbf{TFP}$, $\mathbf{X}$ and $\mathbf{WX}$ as second-stage regressors. After regressing this model, we obtain the Lee estimates $(\beta^{Lee}, \gamma^{Lee}, \theta^{Lee}, \omega^{Lee})$.

Correlated Effects

In the presence of correlated effects (i.e. when $\delta \neq \mathbf{0}$), the estimators above yield biased estimates. To deal with this problem we consider, for a given network ψ, the following equation

$$\mathbf{TFP}_\psi = \beta_\psi\mathbf{1} + \mathbf{X}_\psi\gamma + \mathbf{WX}_\psi\theta + \omega(\mathbf{W})\mathbf{TFP}_\psi + \mathbf{u}_\psi \tag{2.7}$$

[6]Note that the diameter of a network is defined as the “longest shortest path” between any two vertices. Therefore, to satisfy Corollary 1 in Bramoullé, Djebbari and Fortin (2009), it suffices to find a “shortest path” that is larger than three.

[7]Our model controls for heteroskedasticity, i.e. we use the generalized version of Lee’s Estimators. As a result, our estimator loses the optimality property, but maintains its consistency property. The loss of optimality does not seem to make a significant difference, given that our model estimates are very robust.

Here, β_ψ represents unobserved shocks that are common to each member of the network. We thus replace the assumption that $\mathbf{E[u_\psi|X_\psi] = 0}$ by the weaker assumption that $\mathbf{E[u_\psi|X_\psi] = X'}\delta$ but $\mathbf{E[u_\psi|X_\psi,}\beta_\psi\mathbf{] = 0.}$ This identification strategy assumes that correlated effects are present within the members of network ψ. In other words, we allow for the possibility that members of a given network are affected by common shocks (e.g. climatic, technological, economic, etc.).[8]

To identify equation (2.7), we define the square matrix $\mathbf{H}$ of size N, where each entry $h_{ij} = \frac{1}{L_\psi}$ if $i, j \in \psi$ and $h_{ij} = 0$ otherwise. This matrix allows us to average out the fixed network effects (see figure 2.6 for an illustration). Using this matrix, we transform our model by letting

$$\begin{aligned}(\mathbf{I} - \mathbf{H})\mathbf{TFP} = (\mathbf{I} - \mathbf{H})\mathbf{X}\gamma + (\mathbf{I} - \mathbf{H})\mathbf{WX} \\ \theta + \omega(\mathbf{I} - \mathbf{H})(\mathbf{W})\mathbf{TFP} + (I - H)\mathbf{u}\end{aligned} \tag{2.8}$$

Proposition 5 in Bramoullé, Djebbari, and Fortin (2009) shows that whenever matrices $\mathbf{I}$, $\mathbf{W}$, $\mathbf{W^2}$and $\mathbf{W^3}$ are linearly independent, then equation (2.8) can be identified. This condition is more stringent than the case where correlated effects are absent. This is due to the fact that, to control for correlated effects, some information is lost. Because the diameter of our network is greater than 3, Corollary 1 in Bramoullé, Djebbari and Fortin also guarantees linear independence of these matrices.

To identify our model in the correlated effects case, we again use Lee's estimators. In this case, the first step of our estimation consists in specifying a 2SLS regression. We use $(\mathbf{I} - \mathbf{H})\mathbf{X}$, **(I-H)WX** and $(\mathbf{I} - \mathbf{H})\mathbf{W^2X}$ as first-stage instruments, to instrument for $(\mathbf{I} - \mathbf{H})(\mathbf{W})\mathbf{TFP}$. In the second-stage regression, we use $(\mathbf{I} - \mathbf{H})(\mathbf{W})\mathbf{TFP}$,$(\mathbf{I} - \mathbf{H})\mathbf{X}$ and $(\mathbf{I} - \mathbf{H})\mathbf{WX}$ as regressors. This model allows us to recover estimates $\gamma^{\mathbf{2SLS}}$, $\theta^{\mathbf{2SLS}}$ and ω^{2SLS}.

In the second step, we use these estimates to obtain the expectation of the reduced form equation, where

$$\begin{aligned}\mathbf{E}[(\mathbf{I} - \mathbf{H})(\mathbf{W})\mathbf{TFP^{2SLS}}|\mathbf{X}, \mathbf{W}] \\ = \mathbf{W}(\mathbf{1} - \omega^{\mathbf{2SLS}}\mathbf{W})^{-1}[(\mathbf{I} - \mathbf{H})\left(\mathbf{X}\gamma^{\mathbf{2SLS}} + \mathbf{WX}\theta^{\mathbf{2SLS}}\right)]\end{aligned}$$

We then perform a second 2SLS regression, but this time we use instruments $\mathbf{E}[(\mathbf{I} - \mathbf{H})(\mathbf{W})\mathbf{TFP^{2SLS}}|\mathbf{X}, \mathbf{W}]$, $(\mathbf{I} - \mathbf{H})\mathbf{X}$ and $(\mathbf{I} - \mathbf{H})\mathbf{WX}$ in the first-stage equation. In the second-stage, we use $(\mathbf{I} - \mathbf{H})(\mathbf{W})\mathbf{TFP}$, $(\mathbf{I} - \mathbf{H})\mathbf{X}$ and $(\mathbf{I} - \mathbf{H})\mathbf{WX}$ as regressors.

[8] Recall that networks are formed across time periods. For this reason, β_ψ also includes time fixed-effects.

$$\mathbf{H} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

Figure 2.6: An illustration of Matrix **H** (from Figure 2.3).

2.4.5 Data Description

The statement No. 14 of the Financial Accounting Standards Board (FASB) requires firms to disclose the revenue derived from sales to customers that exceed 10% of their annual revenue. The reports contain information about the principal firm (i.e. global identifier and company name), the year, the customer's name and the sales made to this customer. Compustat retrieves these relationships from annual 10-K filings, and stores this information in the business segments database. Note that this dataset only reports "major" customers, i.e. customers that exceed the 10% threshold. While it would have been desirable to have data on all customers, this is not a severe limitation. This is because very small customers are unlikely to play a significant influence on the firm. That is, supply chain relationships are important in so far as they represent a significant portion of a firm's annual revenue (e.g. more than 10%).

The source dataset is considerably messy and the reports do not contain global identifiers for the customer. To build this dataset, our first step was to perform a visual check. In this check, we noticed that some observations reported ambiguous statements about a customer, for example, by referring to it as "customer 1", or by reporting "2 customers" instead of detailing their names. In other cases, the firm's name was ambiguously abbreviated. If an observation fell into any of these cases, it was discarded. In the vast majority of the cases, however, the customer's name was fully spelled, or the abbreviation was clear enough. For example "Johs. & Johs." was clearly making reference to "Johnson & Johnson". In these cases, the observation was rewritten to match the original company name.

After performing our visual check, we matched the firms through a phonetic string algorithm. This algorithm allowed us to match the names reported by a firm to the Compustat database. If the word matching software was unable to properly match the company, or if the accuracy of the match

# of customers	observations (%)	Avg. revenue
0	8,013 (29.27%)	0
1	13,981 (51.07%)	19.89%
2	3,380 (14.17%)	30.71%
3	1,185 (4.33)%	43.06%
$\geq$4	318 (1.15%)	48.93%

Table 2.3: Summary statistics about the customer-base of each firm.

	Network Size (L_ψ)
Mean	286.56
Std. Deviation	281.84
Median	256
Minimum	2
Maximum	785
Number of Networks(Ψ)	4485

Table 2.4: Summary statistics about the firm's networks.

was below 90%, the firm was manually matched. Our manual match was successful in the vast majority of these cases.

After obtaining TFP estimates, we matched each firm to each of its customers. If there were data missing about the relationship, we discarded the customer. For example, in some cases the firm did not report the amount of sales made to one of its customers, and in other cases this amount was equal to zero or negative.

After cleaning the data, we kept all firm-year observations that either report a customer, or those that are reported by a customer. The resulting dataset contains 27,699 firm-customer relationships, $N = 26,336$ firm-year observations and $\Psi = 4,485$ networks ranging between 1976 and 2009. Our dataset covers 6,597 firms, which is approximately 28% of the Compustat universe. We use these firms to construct our interaction matrix $\mathbf{W}$.

Table 2.3 summarizes the customer-bases for all firms; Table 2.4 summarizes the network characteristics. From Table 2.3 we can observe that 29.27% of our sample consists of firms without “major” customers. Most of these firms are retailers, e.g. Wal-Mart and K-Mart or apparel stores, or consumer services firms. Approximately half of the firms report exactly one major customer. This customer represents, on average, 19.89% of the firm's annual revenue. About 14% of the firms report two major customers, and these customers represent 30.71% of the firm's annual revenue. On average,

VARIABLES	(1) N	(2) mean	(3) std.dev.	(4) p25	(5) p50	(6) p75	(7) p99
TFP	27,376	1.783	0.649	1.505	1.782	2.094	3.386
Total assets	27,376	6,854	33,901	41.69	249.9	2,136	116,672
Leverage	27,376	0.267	0.314	0.0763	0.227	0.373	1.159
Inv. Turnover	27,376	0.140	0.182	0.0543	0.120	0.190	0.518
Age	27,376	17.12	14.29	6	12	26	55
Region: West	27,376	0.223	0.416	0	0	0	1
Region: Midwest	27,376	0.213	0.409	0	0	0	1
Region: South	27,376	0.247	0.431	0	0	0	1
Region: Northeast	27,376	0.351	0.477	0	0	1	1

Table 2.5: Table of summary statistics

we capture 25.5% of the annual revenue from firms that report at least one customer.[9] From table 2.4 we can observe that the average customer belongs to a network of size 287, and that the largest network includes 785 firms.[10]

Vector of Characteristics To build the vector of firm characteristics ($\mathbf{X}_i$) we include firm age, size, financial leverage, inventory turnover and geographic region (at the beginning of the period). These characteristics are summarized in Table 2.5.

We proxy for firm age by observing the year in which the firm first appeared in Compustat; we also include the square of firm age to control for (observed) non-linearities. To control for the effect of firm size, we use the natural logarithm of total assets. We include financial leverage and inventory turnover to proxy for the financial and operational conditions of the firm. Financial leverage is defined as the ratio of total debt to the book value of total assets, and inventory turnover is defined as the ratio of annual net-sales to average inventory.[11] We also control for the square of inventory turnover

[9] Note that a significant fraction of the sales (which are not reported in these data) go to small consumers.

[10] To ensure that our estimates (particularly when correcting for network fixed effects) are not affected by the inclusion of small networks, we re-estimated them by excluding these networks. As we show in §2.5.3, our results are robust to the inclusion or exclusion of small networks.

[11] A more precise definition for inventory turnover uses the ratio of the costs of goods sold to inventory. However, some observations were missing information about the "cost of goods sold". But as we show in §2.5.3, our results are invariant from the choice of definition.

after observing a non-linear relationship between TFP and turnover. Finally, we control for geographic effects by dividing firms into five regions: Northeast, Midwest, South, West and Overseas. The first four categories represent the official Census-Bureau designated regions, and the (excluded) dummy Overseas includes international firms, or those firms located outside of mainland U.S. Approximately, 5% of the firms are located overseas.

The definitions above are based on standard accounting definitions and those used by similar studies (e.g. Patatoukas, 2011; Keller and Yeaple, 2009, Kalwani and Narayandas, 1995). However, some variables can be defined in alternative ways. To ensure that our results are not skewed by a particular definition, we re-estimated our model using alternative definitions. These robustness checks are explained in §2.5.3.

2.4.6 Estimates

We show our estimates in Table 2.6: in Column 1, we report the OLS estimates; in Column 2, we present the Lee estimates (without controls for correlated effects); in Column 3, we present the Lee estimates with controls for correlated effects.[12] A Hausman test reveals that the model in Column 3 yields the most robust estimates (the p-value for this test is smaller than 0.001). We summarize our main findings below.

Endogenous Effect (ω): The Lee estimators report the presence of large and positive endogenous effects. The estimate of the coefficient, ω, is equal to 0.5979 when we do not control for correlated network effects; the coefficient is equal to 0.6081 when we correct for this source of bias.[13] According to this result, if the average log-productivity of the customer-base increases by one standard deviation, the firm's log-TFP increases by about $3/5^{ths}$ of a standard deviation.

Note the OLS estimators show that the endogenous effects are negative. That is, when we fail to control for the reflection problem, our results show that the productivity of a firm decreases when it has a more productive customer-base. This comparison allows us to appreciate the severity of this identification issue.

Exogenous Effects: Below we examine the impact of the exogenous characteristics (we focus on interpreting the most robust estimates, i.e. the estimates from Column 3). We explain both the firm effects, γ, and the spillover effects, θ. Note that we illustrate the direct (θ) and indirect ($\omega\gamma$) spillover effects in Figure 2.7.

[12] In all three models, TFP is estimated through the Olley-Pakes approach.

[13] Both of these estimates are robust at the 1% level of confidence.

Effect	Variable	(1) OLS		(2) Lee		(3) Lee	
		Coeff.	***t-stat***	***Coeff.***	***t-stat***	***Coeff.***	***t-stat***
Endogenous effect (ω)	Cust. TFP	-0.0926**	-2.38	0.5979***	3.02	0.6081***	6.35
	Age	0.0079***	3.02	-0.0027	-0.72	0.0101***	6.19
	Age^2	-0.0002***	-5.04	0.0001*	1.96	-0.002***	-5.73
Exogenous	Size	0.0354***	4.55	-0.0660**	-2.46	-0.0567***	-4.42
spillover	Leverage	-0.0000	-0.41	0.0000	0.89	0.0001	0.72
effects (θ)	Inv. Turnover	0.9955***	10.00	-0.0313	-0.91	0.3551***	3.09
	Inv. Turnover^2	-0.2027***	-6.81	0.0169	0.61	-0.0085***	-3.00
	Region: West	0.3579***	8.88	-0.3982**	-2.20	-0.2621***	-3.01
	Region: Midwest	0.2638***	10.09	-0.1819**	-2.48	-0.0297	-0.66
	Region: South	0.2636***	9.06	-0.1048	-1.12	-0.0374	-0.84
	Region: Northeast	0.2719***	10.68	-0.1959**	-2.56	-0.0217	-0.50
	Age	0.0079***	4.08	-0.0167***	-11.27	0.0042***	3.63
	Age^2	-0.0003***	-9.20	0.0003***	9.89	-0.0002***	-7.12
	Size	0.1346***	58.67	0.0489***	9.06	0.1396***	76.84
	Leverage	-0.0559***	-2.99	-0.1619***	-7.26	-0.0592***	-4.16
Firm	Inv. Turnover	0.4619***	8.39	-0.1611***	-2.74	0.3989***	9.68
effects (γ)	Inv. Turnover^2	-0.0598***	-7.11	0.0336**	2.51	-0.0491***	-6.88
	Region: West	0.6554***	23.07	0.1000***	5.68	0.5838***	33.63
	Region: Midwest	0.5267***	21.60	0.0656**	2.44	0.5199***	33.36
	Region: South	0.3844***	23.09	0.0508***	3.87	0.3594***	28.16
	Region: Northeast	0.4890***	25.34	0.0835***	5.42	0.4523***	33.47
	Intercept			1.4819***	19.51		
Number of Observations		27,376		27,376		27,376	
R-Squared		0.8920		0.0003		0.8787	
Fixed Effects (Network level)		Yes		No		Yes	

*** p<0.01, **p < 0.05,*p <0.1

Table 2.6: Model estimates for equation 2.3.

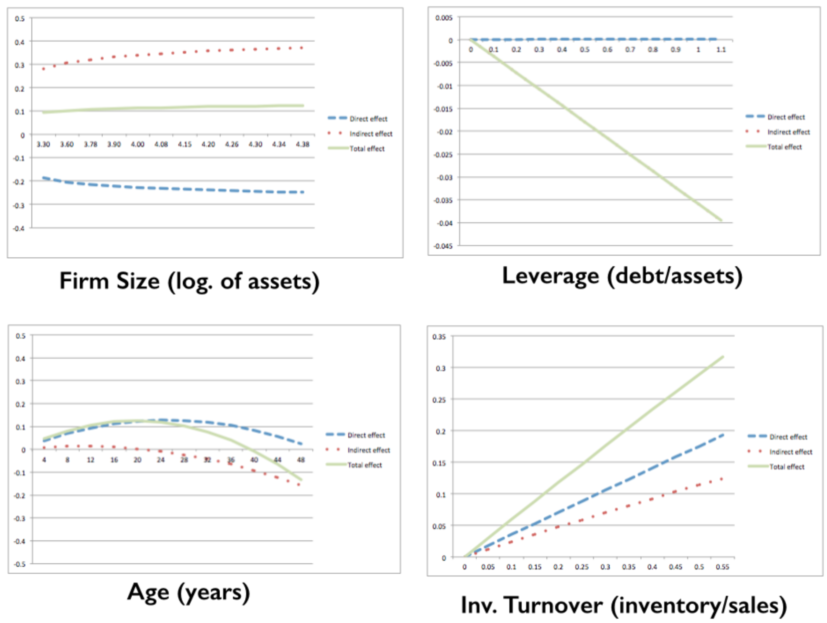


Figure 2.7: Estimates of exogenous spillovers.

Size The exogenous *firm* effect is positive ($\gamma_{size} = 0.1396$). This means that larger firms are more productive, which is likely due to scale economies in the production processes (see §2.3). As a result, firms *indirectly* benefit from interacting with larger customers. Note that the size of the indirect spillover effect is $\gamma_{size}\omega = (0.1396)\,(0.6081) = 0.085$. In contrast, the coefficient estimate for the *direct (exogenous) spillover* effect is negative ($\theta_{size} = -0.0567$). This means that, holding the partner's productivity fixed, firms benefit from having smaller customers. This result can be explained by the idea that when a firm contracts with smaller customers, it can exert more "influence" over them (and this can benefit the firm). The indirect effect dominates the direct spillover effect, implying that it is efficient to interact with larger customers. The total spillover effect is $\theta_{size}+\gamma_{size}\omega = 0.0283$.

Leverage The firm effect ($\gamma_{leverage}$) is negative, which can be explained by the idea that a leveraged firm has financial obligations, and these obligations constrain the firm's production possibility frontier. We find that the exogenous spillover effect, $\theta_{leverage}$, is approximately equal to zero. This means that, if we hold constant a customer's TFP, the customer's financial leverage causes no spillover effects. Therefore, a customer's financial leverage is only significant because it negatively impacts the customer's own TFP. The total spillover effect is equal to $\theta_{leverage} + \gamma_{leverage}\omega = 0 + (-0.0592)\,(0.6081) = -0.036$.

Age Our estimates from Column 3 show that the *firm* effect is inversely U-shaped, and peaks at age 14. Jensen et al. (2001), Van Biesebroeck (2005) and Fernandes (2008) find an inverse U-shaped relationship between firm age and productivity. Our results show that, in addition, the *spillover* effect is inverse U-shaped (this effect peaks at age 24), i.e. it is efficient to interact with customers that are neither too young nor too old. This result can be explained as follows: when a firm interacts with very young customers, these customers are inexperienced at handling their operations. For example, they may have unstable ordering cycles, which can trigger a large bullwhip effect across the supply chain. As the customers age, there is a learning-by-doing effect that mitigates these inefficiencies (Syverson 2011). But firms are less prone to deviate from their customary practices (to innovate) at an advanced age. Therefore, a firm that interacts with old customers may adopt, or be influenced by, the old-fashioned practices of their partners.

Inventory We show that both the firm effect and the spillover effect are positive. This means that firms with higher inventory turnover are more productive and, in addition, that firms benefit from having customers with high inventory turnover. A causal relationship between productivity growth and inventory reduction has been previously observed across automobile supply chains in Japan (Lieberman and Demeester, 1999).

Location Being located in the U.S. has a positive impact on TFP. But if we hold the partners' productivity fixed, a firm's productivity is positively affected when it has a larger proportion of foreign customers. Similar results are found by Keller and Yeaple (2009), who show that foreign enterprises provide positive productivity spillovers to U.S. firms. If we sum up the indirect and direct spillover effects, we find that (overall) it is beneficial to interact with customers located in the U.S. west coast. The spillover effect of geography, however, is not statistically significant.

Summary of Results

Our results show that a firm's productivity benefits when the firm interacts with customers that have: (i) high productivity levels; (ii) large size; (iii) age between seventeen and twenty; (iv) high inventory turnover and; (v) low financial leverage. The effect of geography is small, and not very robust.

We note three interesting results. First, firm age has two counteracting effects: while firms *indirectly* benefit from interacting with larger customers (because they are more productive), they *directly* benefit from interacting with smaller customers. We argue that this is because large firms exert more influence over smaller customers. Second, inventory turnover causes both direct and indirect spillover effects, but financial leverage only causes indirect spillovers. This hints at the fact that a firm's productivity is more susceptible to the operational conditions of its peers, than it is to their financial conditions. Third, after analyzing the magnitude of the effects, we find that the endogenous effect is the largest one. In other words, the endogenous channel is the primary channel through which productivity spills over - more so than any exogenous channel. This means that interacting with a productive firm is very beneficial, even if the firm has "counterproductive" traits (e.g. high financial leverage). It is less beneficial to interact with unproductive partners (even if they have favorable exogenous traits).[14]

2.5 Extensions and Robustness Checks

We now study some extensions to gauge the robustness of our model. In §2.5.1, we estimate productivity spillovers between a firm and its suppliers. In § 2.5.2, we re-estimate our results by controlling for selection biases in the formation of supply chains. We also perform various (minor) robustness checks, which are discussed in §2.5.3.

[14]This last statement must be interpreted with caution, because "beneficial" characteristics are also associated with high productivity. An alternative way of interpreting this result is the following: if a firm could change any aspect of its customers (holding everything else constant), increasing the customers' productivity would yield the highest benefits for the firm's productivity.

2.5.1 Extension 1: Supplier-Base Analysis

In the preceding sections, we looked at the impact of customers on their supplier's TFP. While it is true that we can invert these relationships to extract the firms' supplier bases, the resulting data overreport small firms as suppliers. This is because firms only report (in the 10-K filings) those customers that represent a large portion of their revenue; on average these customers are large. Large firms, on the other hand, are unlikely to report small firms.

To understand why this is problematic, consider the following stylized example. Suppose A is a small firm whose annual revenue is equal to \$100,000, and 20% of this revenue is obtained from firm B. Firm A will thus report firm B in the data. Now, suppose that firm X is a large firm whose annual revenue is equal to \$100 million, and also that firm X obtains 5% of its annual revenue from firm B. As a result, the data will not capture the relationship between firm B and firm X, even though the total value of trade between X and B is much larger than the value of trade between A and B.

For example, Walmart appears as a major customer for up to 106 firms in a single year, but each of these firms cover a tiny fraction of Walmart's supplier-base. This will likely give rise to a reporting bias when estimating the impact of suppliers on a firm's TFP.

Having acknowledged this limitation, we extend our study to consider the influence of suppliers on their customers' TFP. Although the results presented in this extension may yield interesting insights, they should be interpreted with caution (due to the limitations expressed above).

Econometric Model

We define $\mathbf{W_S}$ as the weighted interaction matrix of firm-supplier relationships. This matrix has typical element

$$w_{ij}^S = \begin{cases} \frac{c_{ji} \cdot (sales_{ji})}{\sum_{j=1}^{N} c_{ji} \cdot (sales_{ji})} & \text{if } c_{ji} = 1 \\ 0 & \text{otherwise} \end{cases}$$

where, recall, $sales_{ji}$ is the amount paid by firm i to supplier j. Based on this definition, we analyze the following structural equation:

$$\mathbf{TFP} = \beta \mathbf{1} + \mathbf{X}\gamma + \mathbf{W_S X}\theta + \omega(\mathbf{W_S})\mathbf{TFP} + \mathbf{u} \tag{2.9}$$

We identify equation (2.9) through the strategies described in §2.4.1.

	No. of suppliers (%)	Avg. cost
0	18,125 (66.21%)	0
1	5,920 (21.62%)	18.7%
2	1,382 (5.05%)	10.1%
3	612 (2.24%)	15.6%
≥4	1,337 (4.8%)	17.9%

Table 2.7: Summary statistics about the supplier-base of each firm.

In Table 2.7, we summarize the data on the (observed) supplier-base. This table also reports the cost of all purchases made by a firm to its supplier base (as a percentage of the firm's total purchases). To calculate this item we use the total cost of goods sold (COGS) from Compustat, and let $Cost_{S_i} = \frac{\sum_j sales_{ji}}{COGS_i}$. As we can observe in the table, the average cost of observed suppliers is relatively small, especially when compared with the customer-base. While major customers represent (on average) 25.5% of the firm's annual revenue, the observed suppliers represent 17% of the firm's annual purchases. We also note that the average supplier size is equal to 1.57, while the average size of a reported customer is equal to 6.63 (size is measured by the logarithm of total assets). This implies that, on average, a reported customer is four times larger than a reporting supplier. Also, while the customer bases include up to 9 "major" customers, the supplier bases sometimes comprise more than 100 suppliers.

Estimates

We present our model estimates in Table 2.8: we present OLS estimates in Column 1; in Column 2, we present the Lee estimates without controlling for correlated fixed-effects and; in Column 3, we present the Lee estimates with controls for correlated network fixed-effects. A Hausman test reveals that the model in Column 3 is the most robust one.

The Lee estimates show that the endogenous effect (ω) is positive. Even so, this estimate is small and statistically insignificant in the fixed-effects model. This could be due to the fact that the suppliers are small and, hence, do not represent a significant portion of the firm's purchases (recall the Walmart example above). The reported suppliers are thus unlikely to represent a significant influence on the firm. The qualitative properties of the exogenous effects are very similar to the estimates derived through the customer-base analysis.

Effect	Variable	(1) OLS Coeff.	(1) OLS t-stat	(2) Lee Coeff.	(2) Lee t-stat	(3) Lee Coeff.	(3) Lee t-stat
Endogenous effect	Sup. TFP	-0.3453***	(-8.03)	0.2930***	(2.83)	0.0027	(0.03)
Exogenous spillover effects	Age	0.0149***	(5.32)	-0.0004	(-0.20)	0.0100***	(3.72)
	Age^2	-0.0004***	(-7.33)	0.0001	(1.23)	-0.0002***	(-2.96)
	Size	0.0284***	(3.59)	-0.0492***	(-2.90)	-0.0279	(-1.61)
	Leverage	-0.0001	(-0.78)	-0.0001	(-1.26)	-0.0001	(-0.78)
	Inv. Turnover	-0.0483	(-0.42)	-0.6361***	(-5.91)	-0.3201***	(-2.60)
	Inv. $Turnover^2$	0.1656*	(1.72)	0.3200***	(5.51)	0.3013***	(3.69)
	Region: West	0.2689***	(4.48)	-0.3353***	(-3.37)	-0.0680	(-0.65)
	Region: Midwest	0.0902*	(1.82)	-0.3481***	(-4.17)	-0.1824**	(-2.09)
	Region: South	0.0959**	(2.09)	-0.2521***	(-4.17)	-0.1047	(-1.61)
	Region Northeast	0.0720*	(1.68)	-0.2967***	(-3.93)	-0.1773**	(-2.25)
Firm effects	Age	0.0155***	(7.60)	-0.0172***	(-10.64)	0.0157***	(14.91)
	Age^2	-0.0005***	(-15.75)	0.0003***	(8.97)	-0.0005***	(-23.65)
	Size	0.1445***	(41.87)	0.0385***	(8.58)	0.1412***	(61.63)
	Leverage	0.0332	(1.52)	-0.1773***	(-7.44)	0.0376***	(2.63)
	Inv. Turnover	0.8590***	(15.12)	-0.2764***	(-5.50)	0.8317***	(21.00)
	Inv. $Turnover^2$	-0.1159***	(-7.09)	0.0465***	(3.35)	-0.1135***	(-16.11)
	Region: West	1.0255***	(31.14)	0.1105***	(6.93)	1.0239***	(79.31)
	Region: Midwest	0.8812***	(29.38)	0.0249	(1.60)	0.8903***	(61.50)
	Region: South	0.5997***	(30.56)	0.0599***	(5.29)	0.5966***	(50.33)
	Region Northeast	0.7789***	(31.75)	0.0633***	(4.85)	0.7794***	(67.22)
	Intercept			1.7683***	(61.53)		
Number of Observations		27,376		27,376		27,376	
R-Squared		0.8809		0.0205		0.8763	
Fixed Effects (Network level)		Yes		No		Yes	

*** p<0.01, **$p < 0.05$,*p <0.1

Table 2.8: Model estimates for equation (2.9)

2.5.2 Extension 2: The Formation of Supply Chains

In the preceding sections we assumed that the formation of supply chains is an exogenous process. But in reality firms make *strategic* decisions about whether to select suppliers that are large or small, proximate or distant, etc. The underlying selection process will give rise to a non-randomly selected sample. We thus need to consider the possibility that spillover effects are correlated with the selection of supply chain partners. We refer to this problem as *partner selection bias*.[15]

To explain why this is a problematic issue, consider the following example: assume firm j is a customer of firm i and suppose that when a firm is located overseas (as opposed to the U.S.), this induces a positive shock on the partner's TFP. So if j relocates its production facilities from the U.S. to overseas, we expect the productivity of supplier i to receive a positive shock. But we can also expect that firm j's desire to purchase from firm i will change as a result of $j's$ relocation. For example, j will drop i as a supplier if the shipping costs increase significantly.

Fortunately, most of the supply chain relationships in our data linger through several years. Note that 95% of the firm-customer relationships that are found in period $t-1$ are also found during period t and, on average, a reported relationship lasts for over five years. This is likely due to the presence of long-term contracts, and implies that most of the observed relationships are the result of past decisions.

While the above condition drastically mitigates the impact of this problem, it does not completely eliminate it. To address this issue we build a *strategic network formation model* that draws on Goldsmith-Pinkham and Imbens (2013). This model allows us to derive consistent estimates in the presence of partner selection biases.[16]

This approach is analytically complex, and we explain it in the Appendix. Roughly speaking, we build a dynamic approach that considers two sequential processes. In the first process, firms select their supply chain partners based on a number of factors. We estimate these factors. And then, *conditional* on the selection process, we estimate the determinants of productivity

[15] A similar problem arises in sociological networks, where individuals have a tendency to befriend people with similar attributes. This is known as *homophily*.

[16] Note that this type of bias plagues (virtually) every study that investigates spillover effects across supply chain networks (including the ones cited in this manuscript). This problem has not been controlled in the past, mainly because there were no theoretical models to address it. As Bramoullé (2013) notes, "Goldsmith-Pinkham and Imbens propose one of the first convincing applications" to deal with this problem. For this reason, our attempt to control for partner selection bias has some limitations.

(i.e. the firm effects and the spillover effects).

To jointly estimate the factors that influence both processes, we use Monte-Carlo-Markov-Chain methods coupled with Bayesian estimation. We report details about our model, and the results in the Appendix.

Our main take-aways from this extension are the following. First, we find (from Table A.1 in the Appendix), that two firms are more likely to engage in a relationship if: (i) their industries are compatible;[17] (ii) the firms were already engaged in a relationship; (iii) the firms are in close geographic proximity and; (iv) the supplier has large size and age. Other factors, such as inventory turnover and financial leverage, did not play a statistically significant influence on the selection of supply chain partners.

Second, we find that there are mild partner selection biases that cause the endogenous effect to be slightly upward biased (see Table A.2 in the appendix). However, the magnitude of these biases is very small, and not nearly large enough to overturn our main conclusions. We also find that the exogenous effects are practically unaltered by the presence of these biases. This is likely due to the fact that supply chain relationships last for several years implying that, to a large extent, the selection of supply chain partners is exogenous to our model. As a result of this test, we are confident that our main results are robust to the presence of partner selection biases.

2.5.3 Other Robustness Checks

In this section we test the robustness of our main findings by varying some constructs from our main model: in §2.5.3, we vary model constructs of the production function; in §2.5.3, we vary the constructs of the model used to estimate spillover effects.

Due to space constraints, we do not present the tables for the results that are discussed below, but all tables are available from the authors. Also, note that the robustness checks presented below do not comprise an exhaustive list. We performed other (minor) robustness checks, which are not discussed in this section.

[17] We create a proxy to measure the compatibility between two industries by looking at the historical connectivity between every pair of industries. We find, for example, that the leather and the apparel industries are very compatible.

Productivity Estimation

Industry Classification

Productivity is the relative efficiency ranking of a firm within the industry it operates. Any productivity estimates are thus unavoidably sensitive to the classification scheme used to estimate them. To avoid inconsistencies across studies, the literature often uses the Standard Industry Classification (SIC) code. In this essay we follow this standardized approach by classifying firms into an industry according to their 3-digit SIC code.

We decided against using a 2-digit classification scheme, because this classification is too coarse. For example, under a 2-digit SIC code, we would have had to group together firms producing pickup trucks, and firms producing bicycles, into the same industry. A 4-digit SIC code, on the other hand, would have forced us to granulate firms into very specific industries. For example, we would have had to generate a separate industry for firms that manufacture leather luggage (SIC code 3161), and one for firms that manufacture leather purses (SIC code 3171).

To ensure that our main conclusions are not sensitive to the industry classification, we re-estimated our results by using a 2-digit and a 4-digit SIC code. We also used a 3-digit North American Industry Classification System (NAICS) code. We find that our estimates are very robust to industry classification. This model construct does not influence our main results.

Intermediate Materials

The classic Cobb-Douglas production function, $Y = AK^{\beta_K} L^{\beta_L}$, has two input choices: capital (K) and labour (L). But it is not uncommon to find different specifications for this function. A very popular alternative includes, in addition to labour and capital, a third input that measures *intermediate materials* (M). Under this specification the Cobb-Douglas function becomes $Y = AK^{\beta_K} L^{\beta_L} M^{\beta_M}$.

There is no inherent advantage with either functional form, and our choice was motivated by the fact that the related literature uses a two-factor functional form (e.g. Imrohoroglu and Tuzel, 2014; Kellogg, 2011; Van Biesebroeck, 2005, etc.).

We re-estimated our results by using a three-input production function, and found that the estimated input elasticities are $\hat{\beta}_K = 0.3681$, $\hat{\beta}_L = 0.4697$ and $\hat{\beta}_M = 0.2258$. We used the resulting estimates to perform a visual comparison of the firm's productivity under both types of functions. The relative ranking of firms does not seem to be affected by the choice of a

2-input or a 3-input production function. In other words, productive firms look productive (and unproductive firms look unproductive) regardless of the specification used. In the actual estimation of spillover effects, none of the results change their qualitative properties.

Estimation of Spillover Effects

Dynamic Spillover Effects

The endogenous spillover effect measures how a firm's productivity is affected by the productivity of its partners. As such, ω captures the effect of communication and collaboration with other firms, but also the result of learning and mentoring.

An important issue here is that learning effects are not always internalized by the firm immediately but, rather, they diffuse through time. In other words, it may take two to three years for a firm to learn and adopt the productive practices of their partners. But our model does not capture dynamic effects.

We tested the robustness of our model by studying dynamic spillovers. In this extension, we assume that a firm's productivity affects its partners' productivity with a lag. We study the following model

$$\mathbf{TFP_t} = \beta\mathbf{1} + \mathbf{X_t}\gamma + \mathbf{W_t}\mathbf{X_t}\theta + \sum_{\mathbf{n=0}}^{\mathbf{N}} \omega_{\mathbf{t-n}}(\mathbf{W_{t-n}})\mathbf{TFP_{t-n}} + \mathbf{u_t}$$

We use the above model to study the lagged effect of the endogenous channel. Note that the reflection problem does not affect the lagged variables.[18] Therefore, these lagged effects can be treated as exogenous regressors.

We regressed various specifications of the above model by setting N equal to 1, 2 and 3, and also by excluding the spillover effect at time t. We did find the presence of lagged effects. For example, by running a model with two lagged variables (i.e. by setting $N = 2$), we found that the endogenous spillovers were equal to $\omega_t = 0.2619$, $\omega_{t-1} = 0.07527$ and $\omega_{t-2} = 0.118218$.[19] All of these coefficients are statistically significant at the 1% level. If we drop the lagged effects, our regression (for this same sample) shows that $\omega_t = 0.3026$. All of our results are qualitative identical under both regressions.

Our main conclusion is that our qualitative results are robust to the introduction of lagged endogenous effects. However, as we see above, the

[18] For a discussion about this issue, see Aakvik et al. (2013).

[19] Note that this sample includes only those firms that appear for three consecutive years. The sample size is equal to 11,645.

learning effects that diffuse through time are non-trivial. These effects should be analyzed in a more careful manner, in a separate study.

Small Networks

A significant number of networks contain less than five firms. To make sure that our estimates are not affected by the inclusion of small networks (particularly when correcting for network fixed effects), we re-estimated them by excluding small networks. The results are not altered by the exclusion or inclusion of these networks.

Variable Definitions

We relied on standard accounting definitions to define the vector of exogenous characteristics, $\mathbf{X}_i$. Unfortunately there are multiple (valid) definitions for some characteristics. In these cases we adopted the most commonly used definition (in related literature), or the choice was data-driven.

To ensure that our main results are robust to the choice of definition, we re-estimate these results by using alternative definitions:

Inventory Turnover we defined inventory turnover as the ratio of net sales to inventory. However, it is more precise to define this variable as the ratio of the *Cost of Goods Sold* to the average of inventory.

Our choice was motivated by the fact that more observations report data on annual sales than they do on the cost of goods sold. This allowed us to drop fewer variables. Note that this is not a big issue, given that there is a 97.4% correlation between both definitions. In the actual estimation, our results were unaltered by the choice of definition.

Financial Leverage To measure financial leverage, we used the ratio of total debt (short plus long-term debt) to total assets. But in some studies leverage only includes long-term debt in the numerator (e.g. Cardella, 2013; Imrohoroglu and Tuzel, 2014). In other studies, leverage is measured by using ratio of book value of total assets to the book value of equity (e.g. Patatoukas, 2011). We re-estimated our results using these alternative definitions. The quantitative results are not altered by either definition.

Geographic Location We could control for geographic effects by dividing firms into states, or into one of the nine Census-Bureau designated divisions. We found that using the official census regions (West, Mid-West, South, North East) facilitates the exposition of our analysis, as opposed to dividing firms into states, and analyzing the separate impact of each dummy variable.

If we re-estimate our results by classifying firms into their home states, our results are qualitatively identical. As an aside note, it worth noting that states like California, New York and New Jersey have the most productive firms (on average).

2.6 Conclusions

We provide new evidence about the link between Total Factor Productivity and supply chains. Our main contribution is to identify the various channels through which productivity can spill over across firms. We identify the impact of a firm's productivity on the productivity of its supply chain partners (the *endogenous* channel), and also the impact of the firm's characteristics on these partners (the *exogenous* channels).

Our estimates yield several interesting results. Among other things, we find that interacting with productive firms (i.e. the endogenous channel) is the largest source of productivity spillovers along the supply chain, more so than any exogenous channel. We also find that a firm's productivity is more susceptible to the operational, than to the financial characteristics of the partners.

To arrive at the above results, we first had to deal with the reflection problem. This is often considered one of the most challenging econometric tasks. To this end, we used a novel econometric approach that exploits partially-overlapping network interactions. Note that without controlling for this problem, we would have found the opposite result, i.e. that interacting with productive partners hinders a firm's productivity (compare the estimates of the endogenous effect, ω, in Column 1 with ω in Column 3 - in Table 2.6).

Identifying the joint impact of these two types of channels is more than an interesting econometric exercise. Our estimates provide a precise description about how firms affect the productivity of their partners. These details can be useful for practitioners at the time of building (and managing) supply chain relationships. To our knowledge, such estimates do not exist in the literature, perhaps because of the identification challenges involved.

Thanks to the recent infusion of data about inter-firm relationships, there are several avenues for further research. For example, our model did not tackle the issue of spillover effects across competitors. The influence of competitors can be modeled by using the approach adopted by Bloom et al.

(2013). Another topic of research is to explore the impact of second- and third-tier suppliers (and customers) on the productivity of the firm. It would also be interesting to study the dynamic nature of the spillover effects, i.e. how the influence of partners persists over time. This essay contributes by providing a robust framework to estimate spillover effects across supply chains.

Chapter 3

The Strategic Role of Business Insurance in Managing Supply Chain Risk

3.1 Introduction

Firms routinely face the possibility of operational failures. For example, a contaminated input may cause a product recall, an industrial accident may shut down production or a toxic spill may cause an environmental hazard. These events often lead to serious financial consequences that can threaten the survival of the firm. The likelihood of a failure, however, can be mitigated through costly effort such as operational maintenance or quality control.

In a supply chain, preventing a failure involves coordinating the efforts of multiple firms. As a case in point, consider oil drilling operations. The likelihood of an oil spill depends both on the care taken by the driller and the oil well cementer. The efforts of the two firms are partially substitutable because, for example, the oil driller can increase its drilling care to compensate for a poorly cemented oil well. However, the safety actions taken by the driller are more effective when the well is appropriately built and maintained by the cementing provider (and vice versa).

Coordinating reliability efforts is complicated by two factors. First, the efforts of the firms may not be observable (e.g. the level of care taken by an operator in maintaining its equipment). This will give rise to moral hazard problems. And second, it may be impossible to identify the root-cause of the operational failure, which will lead to ambiguity about the degree of responsibility of each party. This is a frequent problem in operations involving "interdependent" systems (Kim and Tomlin, 2013). For example, an executive from EcoMotors argues that in the event of a product defect leading to a recall, "assigning responsibility for warranties gets messy. Was it just a part that was not designed properly? Was it the environment that the part was in, which typically is controlled by the automaker?" (Armstrong, 2003). In other cases the failure automatically destroys all evidence about the root-cause, e.g. an explosion leading to a major fire (Okes, 2009).

Firms can deal with the above two problems through contractual clauses by allocating the financial burden of a failure, i.e. the *ex post liability*, to those parties best positioned to prevent it. These clauses take the form of performance penalties, liability-sharing agreements or quality warranties, and can achieve the right incentives for effort provision. Contractual tools are often used to apportion the potential costs of various types of failures, including product recalls arising from quality defects (Chao et al., 2009), oil spills (Hewitt, 2008) and the accidental releases of other toxic materials (Gallagher, 2012).

But contractual agreements are not the only tool used by firms to allocate the financial liability of an operational failure. Firms also have the ability to transfer their liability away from the supply chain to third parties. For example, a firm can purchase business insurance to provide coverage for losses arising from an operational failure. As a result, firms have two mechanisms to deal with failures: (i) contractual incentives, which allocate financial liability <u>within</u> the supply chain and; (ii) insurance coverage, which transfers the liability <u>away</u> from the supply chain.

Most firms rely on some form of insurance coverage as part of their overall operational strategy. In some cases, insurance offers coverage for losses arising from uncontrollable factors, e.g. a natural disaster or a terrorist attack. But in many other cases insurance offers coverage for events where either the operator or the supplier, or both, can potentially affect the likelihood of the outcome. For example, in the U.S. most insurance companies offer services like equipment breakdown insurance or boiler and machinery insurance to cover small losses arising from equipment failures. They also offer services like product recall insurance, nuclear liability insurance or oil-spill insurance to cover losses arising from more serious events. These services cover a variety of industries including light & heavy manufacturing, utilities, steel machinery, mining & minerals, chemical products, hi-tech companies, etc. (RSA group, 2011).

When the firms can affect the likelihood of a failure, the use of insurance may lead to inefficiencies due to moral hazard. This is because insurance decreases overall incentives to ensure operational reliability, by leading to the "de-responsabilization of parties or agents in the supply chain" (Kogan and Tapiero, 2007). Why then do firms buy insurance coverage for these types of events? Would it not be more efficient to allocate all liability within the supply chain, to those firms who are in the best position to prevent an operational failure?

One answer to the above question is that insurance coverage serves a liquidity-enabling role, by allowing wealth-constrained firms to avoid the possibility of bankruptcy or illiquidity due to an event causing large losses. For example, producers of nuclear energy can exchange the prohibitively large costs associated with a nuclear meltdown for a manageable insurance premium. Therefore, if all firms in a supply chain are wealth-constrained and the overall wealth of all firms cannot cover the potential impact of a failure, they may need to seek third-party insurance. Insurance can play a second role in the presence of risk aversion, which is to allow risk-averse firms to transfer risk to a (risk-neutral) insurer. In other words, insurance improves the allocative efficiency of risk within the economy.

But could the above two roles entirely explain the use of business insurance in a supply chain? In this essay we show that, even if the supply chain has adequate wealth and firms are risk-neutral (i.e. in the absence of the above two roles), insurance may serve a purely strategic role within the supply chain. Specifically, we show that without insurance some firms in the supply chain may excessively free-ride on the efforts of other firms. The purchase of insurance can serve as a mechanism that allows a firm to credibly commit not to increase its effort and, thereby, to mitigate the free-riding problem. In other words, some firms in the supply chain can strategically use insurance as a commitment mechanism.

To see the strategic role played by insurance, consider a supply chain relationship where one of the firms has sufficiently large wealth to cover any potential losses arising from an operational failure, but the other firm has severe wealth constraints. In this case contractual agreements alone misallocate incentives along the supply chain. Specifically, contractual tools alone leave the wealth-constrained firm with inefficiently low incentives to exert effort because its is unable to take on the appropriate level of liability. Conversely, the "wealthy" firm ends up with excessively high incentives. Because effort is substitutable, any increase in the effort of the wealthy firm will further undercut the incentives of the wealth-constrained firm to exert effort. Insurance can play a strategic role by allowing the wealthy firm to credibly commit not to exert effort beyond some level and, in turn, the wealth-constrained firm has less incentive to free-ride. Insurance can therefore mitigate the distortion in effort provision and improve total welfare.

The above results are particularly relevant to supply chains characterized by an uneven wealth distribution. These types of relationships are becoming increasingly common as "larger corporations are looking to partner with small, specialized companies" (Business Development Bank of Canada, 2013). Also, in emerging economies it is not rare to see large, wealthy multinationals partnering with small and medium enterprises (Etemad et al., 2001). In these cases, our results imply that the uneven distribution of wealth would lead to firms being more likely to buy insurance as a strategic tool. In general, our results imply that insurance coverage and contractual incentives are not necessarily substitutes, but may complement each other.

3.2 Literature Review

This essay bridges two research streams: (i) supply chain contracts and; (ii) risk management and insurance.

Supply Chain Contracts: We are interested in the sub-stream that focuses on contracts coordinating reliability investments. Kim et al. (2007) and Chu and Sappington (2010) are prominent examples of this literature. Specifically, we consider a context where the operational outcome can be influenced by the efforts of both the operator and the supplier, and also that these efforts are unobservable. This is a setting characterized by double-sided moral hazard, which is also studied by Roels et al. (2010) and Jain et al. (2013b). Second, our model assumes that the root-cause of an operational failure cannot always be attributed to either party; similar assumptions are made by Saouma (2008) and Kim and Tomlin (2013).

Our model considers the case where firms have limited wealth in a principal-agent setting. This literature stems from Sappington (1983), who argues that "contracts in which the liability of one or more parties is explicitly limited are very common in practice." In the economics literature, this topic has been extensively explored; some of the most notable examples are Innes (1990), Holmstrom and Tirole (1997), Oyer (2000), Gromb and Martimort (2007) and Poblete and Spulber (2012). Some examples from the management literature include: Saouma (2008), who studies outsourcing relationships and warranties in a setting where the suppliers have limited wealth; DeVéricourt and Gromb (2014), who study capacity investments under limited liability and; Desiraju (2004), who studies intrabrand competition under the same assumption.

Risk Management and Insurance: The main goal of this essay is to contribute to a better understanding of the interaction between business insurance and supply chains contracting. This literature is limited. In the operations management literature, Dong and Tomlin (2012) study the interplay between business insurance and inventory management, and find that insurance can increase the marginal value of inventory and the overall value of emergency sourcing. Unlike this essay, Dong and Tomlin consider a model where firms obtain coverage for uncontrollable events, e.g. natural disasters or terrorist attacks. Therefore, the authors do not need to consider issues related to moral hazard, which are at the core of our model.

Unlike the operations management literature, there is an extensive literature in economics that explores the optimality of insurance in the presence of moral hazard. Winter (2000) summarizes this literature. Within this large literature, our model is most closely related to Tommasi and Weinschelbaum (2007) who study the robustness of principal-agent contracts to the introduction of third-party insurance. Our results differ from this paper in two key respects. First, while their analysis is driven by the assumption of risk aversion and single-sided moral hazard, we assume that the parties are risk-neutral and subject to double-sided moral hazard. Second, in their paper insurance opportunities are available for the agent, not for the principal; we assume that the principal has the ability to purchase insurance coverage. Due to this reason, they find that insurance decreases welfare while we show that insurance can improve total welfare in contractual relationships.

3.3 Model Preliminaries

3.3.1 Operational Features

A risk-neutral operator (O) receives revenue π from operating a system that requires the technical expertise of a risk-neutral supplier (S). Consider the following examples: (i) an oil driller that delegates all cementing operations to an oil well cementer or; (ii) an equipment operator that outsources all supervision and maintenance tasks to a service supplier. The system is subject to unexpected operational failures, e.g. a biohazard spill leading to an environmental accident, or a defective product that must be recalled. These failures lead to financial losses for the operator from property damages, clean-up costs, production interruption, etc. Let $X \in \{0, x\}$ be a random variable representing the "failure costs" borne by the operator. If $X = 0$, operations have performed as planned, and the operator incurs no losses. If $X = x > 0$, an operational failure has occurred, and the operator incurs losses equalling this amount.[20]

Failure Probability and Reliability Efforts:

To diminish the likelihood of a failure, the supplier and the operator can exert costly but unobservable effort. For example, the supplier may use

[20]In §3.5 we consider an extension where the failure costs have a continuous support.

better quality inputs or improve the design of its product or service. The operator can increase the level of care when performing operations, minimize the systems' exposure to strenuous conditions, hire skilled personnel, etc.

Let $e_S \geq 0$ and $e_O \geq 0$ denote the efforts of the supplier and the operator, respectively, where e_S and e_O are the dollar investments to improve operational reliability; e_S^* and e_O^* denote the optimal effort levels.

The probability of an operational failure is $F(e_S, e_O) = \frac{1}{(1+e_O)^{\beta}(1+e_S)^{1-\beta}}$, where parameter $\beta \in (0,1)$ represents the sensitivity of the failure probability to the operator's effort (relative to the supplier's). When β is low reliability is more sensitive to the supplier's effort and less sensitive to the operator's effort, and vice versa. Note that $F(e_S, e_O)$ is an inverse Cobb-Douglas function, which reflects the idea that the efforts of the operator and the supplier are substitutable and also collaborative in nature. Effort is substitutable because the operator can exert higher effort to compensate for situations where the supplier exerts lower effort, and vice versa. However, it is collaborative because the effort of one firm is enhanced when the other firm exerts high effort, i.e. $\frac{\partial^2 F(e_S, e_O)}{\partial e_S \partial e_O} > 0$. Cobb-Douglas functions are commonly assumed in papers involving collaborative effort in supply chains, including Roels et al. (2010) and Kim and Netessine (2013).

Failure Attribution:

We focus on settings where the root-cause of an operational failure cannot be identified. This assumption is often the norm in practice and is studied in a subfield of reliability analysis called *dependent failure analysis*. From this literature we have identified three settings where this assumption holds and, hence, where our model is applicable:

1. Cascading Failures: Large-scale operations often require the input of interdependent subsystems (e.g. power generation plants, oil and gas pumping, etc.), some of which need the managerial expertise of specialized suppliers. When there is a high level of interdependency, the subsystems are highly susceptible to experiencing a *cascading failure*, where the failure of one subsystem triggers the subsequent failure of other subsystems (Ericson, 2005). In these cases it is either impossible, or prohibitively costly to, to identify the subsystem responsible for the failure. Kim and Tomlin (2013) study liability allocation rules across systems that are susceptible to experiencing cascading failures.

2. Destructive Failures: In certain operational failures, evidence is damaged or destroyed following the event and, as Okes (2009) explains, this often makes it impossible to determine the root-cause of the problem. For example, following the explosion of a large wind turbine in Androssan, UK, "much of the evidence was burned, and Infinis [the wind farm operator] and Vestas [the supplier of wind turbines] disagree on which was the key initial cause of the destructive fire" (New Scientist, 2013).

3. Commingled and Homogeneous Goods: Whenever the efforts of the multiple firms contribute to the production of a good that is commingled, non-modular or homogeneous (e.g. chemical substances, food products), it is a challenge to determine the degree of responsibility of each party for the operational failure. For example, in 2007 a tainted food incident from ConAgra caused approximately 15,000 people in the U.S. to fall sick, but "ConAgra could not pinpoint which of the more than 25 ingredients in its pies was carrying salmonella" (New York Times 2009).[21] Also, in July 2013 an oil-cargo train derailment caused a massive explosion in Lac-Megantic, Quebec. To date, the authorities have not been able to determine if the explosion was caused by chemical contaminants in the oil (from a previous shipment), or because the oil itself contained high levels of flammable hydrogen sulphide gas (The Globe and Mail 2013).

3.3.2 Contracts

In the face of a potential failure, the operator needs to optimally apportion the liability for the accident. The operator can apportion some of the potential losses upstream to the supplier, through a contractual agreement. In addition, the operator can purchase insurance coverage to allocate some of the burden away from the supply chain to a third party insurer. We explain below the specifics of both agreements, (i) the procurement contract and (ii) the insurance policy.

Procurement Contract:

The operator procures the services of the supplier by offering a 'take-it-or-leave-it' contract, $T(w, y, X)$. This contract consists of a fixed-fee payment, $w \geq 0$, and a liability-sharing (or penalty) rate, $y \in [0, 1]$. The penalty rate splits the cost of the failure between the contracting parties: the supplier pays yX, and the operator pays the remainder. The cash flows of the contract (from the operator to the supplier) are $T(w, y, X) = w - yX$, and the

[21] More generally, the FDA (2010) states that whenever there is a contaminated food product "traceability has proved particularly difficult because of the complexity of the distribution system and the practice within pack houses supplying the US market of co-mingling produce."

expected transfer payment is $E_X\left[T\left(w, y, X\right)|e_S, e_O\right] = w - yE\left[X|e_S, e_O\right]$, where $E\left[X|e_S, e_O\right] = xF\left(e_S, e_O\right)$.

This contractual form is widely observed in practice, where the supplier gets a fixed payment but a proportion of any liability (contingent on the realization of performance) is subtracted (Kim et al., 2007). By reviewing different procurement agreements, we found numerous examples of contracts that adopt this apportioning mechanism. For example, in a distribution agreement signed in 2001 between the technology companies Lucent Technologies and Agere Systems, the parties agreed to split any costs arising from product defects and accidental releases of contaminants; Lucent agreed to be liable for 86% of the costs, and Agere for the remaining share (article VI, clause i).

Note that sometimes the parties will first agree to do a preliminary investigation to determine the root-cause of the failure, and the liability-sharing clause will only apply if the investigation is inconclusive.[22] While this behavior is certainly observed, we assume that the root-cause of a failure cannot be identified. But this assumption is not too restrictive because, as Moslelh et al. (1998) put it, "all too often investigations of failure occurrences... do not determine the root causes of failures."

Insurance Contract:

The operator has access to an insurance market to cover any losses arising from the operational failure.[23] The operator will thus negotiate an insurance policy with a third-party insurer. To model this relationship we reviewed numerous insurance policies and spoke with practitioners. We identified three key components (schedules) that characterize these policies: (i) a schedule of insured events and losses; (ii) a payments schedule and; (iii) a schedule of excepted causes.

1. Schedule of Insured Events and Losses: This schedule defines those events for which the operator can obtain coverage (e.g. an oil spill or a product recall), and the types of losses covered by the policy. In our model

[22] For example, in a chemical manufacturing contract between APP Pharmaceuticals and New Abraxis (2007, clause 12.9), the parties agreed to share the costs of a product recall only if the fault of the recall cannot be determined through preliminary tests performed by an independent FDA testing agency.

[23] Throughout the base model we assume that the operator has full control over the levels of insurance that are purchased in the supply chain. In reality, however, the supplier may have the option to purchase insurance to cover some of its losses. We analyze this case in Section 3.5.2.

we assume that both the operational failure and the failure costs are defined within this schedule.[24]

2. Payment Schedule: This schedule includes the cash flows of the policy. We consider an arrangement where the operator pays an *insurance premium* and, in exchange, obtains a *coverage* reimbursement. Let $v \in [0,1]$ denote the *coverage* level, which is equal to the proportion of the failure costs reimbursed by the insurer in the event of an operational failure. This reimbursement amounts to vx.

The *insurance premium*, P, is an upfront payment made by the operator to the insurer. In practice, the premium is quoted as the sum of three factors: (i) the actuarially fair premium, (ii) an additive load, and; (iii) a proportional load. The actuarially fair premium represents the expected reimbursement to the operator, the additive load accounts for the fixed transaction costs of providing insurance and the proportional load accounts for costs that are proportional to the amount of coverage. Therefore, the premium is equal to

$$\begin{aligned} P\left(v|e_S^*, e_O^*\right) \quad = \quad & \underbrace{vE\left[X|e_S^*, e_O^*\right]}_{\text{actuarially fair premium}} \quad + \quad \underbrace{l_{add}}_{\text{additive load}} \\ & + \underbrace{l_{prop}vE\left[X|e_S^*, e_O^*\right]}_{\text{proportional load}} \end{aligned} \tag{3.1}$$

This type of pricing scheme has also been modeled in the literature, for example, by Patel et al. (2005) and Dong and Tomlin (2012). The cash flows of the insurance contract (from the operator to the insurance provider) are equal to $I\left(v,P,X\right) \equiv P - vX$, and the expected cash flows are $E_X\left[I\left(v,P,X\right)|e_S,e_O\right] = P - vE\left[X|e_S,e_O\right].$

We make the following three assumptions. First, we assume that the insurer is knowledgeable about the product technology, i.e. the insurer is informed about the function $E\left[X|e_S,e_O\right]$, but cannot observe the effort provision of each firm. Second, we assume that the insurer can observe the procurement contract, (w,y), before setting P. Therefore, the insurance provider can infer the incentive compatible levels of effort, e_S^* and e_O^*. Both assumptions are standard in the insurance literature (see Winter 2000). Third, to focus on the intuition behind our results we let $l_{add} = l_{prop} = 0$, i.e. the insurance premium is equal to the actuarially fair level of coverage.

[24]For example, in a product recall insurance policy from AIG Insurance Ltd. (2010), the policy defines an insured event as "any *Product Recall* or *Government Recall* resulting from any: (a) Defect; (b) Malicious Product Tampering; or (c) Product Extortion." The same policy defines the insured losses as "all reasonable and necessary (a) Insured's Product Recall Costs; (b) Business Interruption Costs; (c) Replacement Costs. . . "

Our results do not depend on this assumption because both loads reflect transaction costs associated with purchasing insurance and, therefore, the only influence that these factors play in our model is to (uniformly) shift the results without adding any intuition.

Finally, note that the payment schedule of a policy generally includes more complicated payment arrangements, including deductible levels. In §3.5.3 we show that our results are robust to these complications.

3. Schedule of Excepted Causes: An insurance policy also defines a set of excepted causes for which the insurer is exempted from covering the insured. These exceptions generally include unacceptable actions from the insured party, or matters uninsured under the law.[25] We assume that the parties will not engage in any behavior that might lead to an excepted cause of a failure, e.g. unreasonable negligence, violations of the law, etc.

A possible misconception is that the insured party will not be able to obtain coverage unless the root-cause of the failure can be clearly identified. Note that under the law

> *The burden is on an insured to establish that the occurrence forming the basis of its claim is within the basic scope of insurance coverage. And, once an insured has made this showing, the burden is on the insurer to prove the claim is specifically excluded* (Supreme Court of California, 1989).

This means that to obtain coverage the policy holder only needs to show the existence of insured losses arising from an insured event. The insurer can *only* be exempted from making a reimbursement if it can convincingly show that the cause is excepted in the contract (e.g. if the insurer can show that the operator was unreasonably negligent, intentionally tampered with the product, etc.). And, as mentioned earlier, we abstract away from these issues.

3.3.3 Model Dynamics

In Figure 3.1 we illustrate the timeline of this model. In stage one the operator designs the procurement contract, $T(w, y, X)$, and chooses the level of insurance coverage, v. Next, the supplier accepts or rejects the deal offered by the operator. If the supplier accepts the arrangement, the operator purchases the insurance policy and makes the ex ante payments (i.e. the insurance premium and the fixed fee, P and w).

[25] In the *Product Recall Insurance* policy mentioned in Footnote 24, AIG defines a number of excepted causes, including "Uninsurable matters under the law", "Asbestos", "Intentional violation by the Insured of any governmental or regulatory requirements", etc.

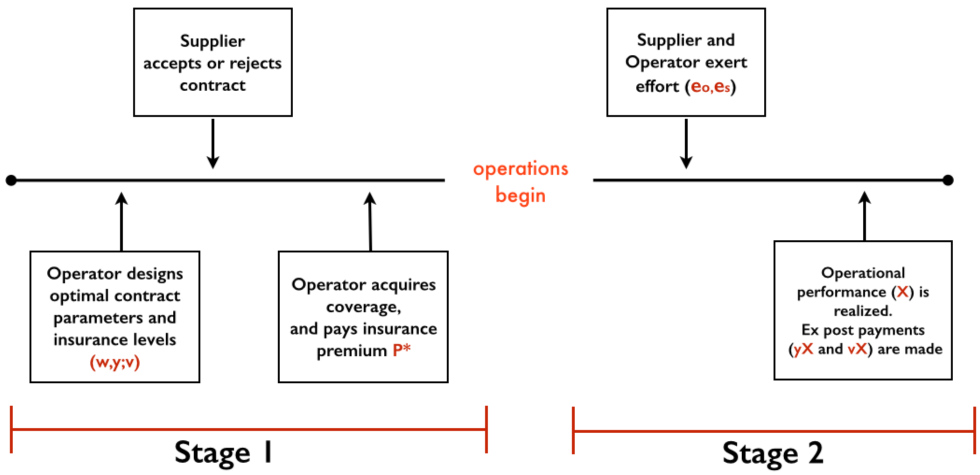


Figure 3.1: Model timeline

Operations begin in stage two. In this stage, the operator and the supplier exert effort levels e_O and e_S. Note that because we assume that the efforts of both parties are unobservable, the sequence in which effort is exerted will not affect the results analyzed below.[26]

After both parties have exerted effort, the operational performance, $X \in \{0, x\}$, is realized. The cash flows, yX and vX, are made at the end of the operational period.

3.3.4 Payoffs and Wealth Constraints

Payoff Functions

The supplier's and the operator's ex post profits are

$$
\begin{aligned}
\Pi_S\left(e_S, w, y, X\right) &= T\left(w, y, X\right) - e_S \\
\Pi_O\left(e_O, w, y, v, P, X\right) &= \pi - X - T\left(w, y, X\right) - I\left(v, P, X\right) - e_O
\end{aligned}
$$

[26]This is because, according to the *Principle of Interchange of Moves,* the order of play is "immaterial if one player does not have any information about the other player's action when making his choice." (Osborne and Rubinstein, 1994). In Section 3.5.1, however, we study a sequential game where the supplier is the first party to exert effort, and this effort is observable. Our results are robust to this setting.

Therefore, the expected profits for both firms are

$$\begin{aligned} E_X\left[\Pi_S\left(e_S, w, y, X\right)|e_O\right] &= E_X\left[T\left(w, y, X\right)|e_S, e_O\right] - e_S \\ E_X\left[\Pi_O\left(e_O, w, y, v, P, X\right)|e_S\right] &= \pi - E_X\left[X|e_S, e_O\right] - E_X\left[T\left(w, y, X\right)|e_S, e_O\right] \\ &\quad - E_X\left[I\left(v, P, X\right)|e_S, e_O\right] - e_O. \end{aligned}$$

Wealth Constraints:

We focus on scenarios characterized by low-probability but high-impact failures (e.g. a product recall, a biohazard accident, etc.). In this context firms are often unable to sustain large financial losses. We assume that at the outset (i.e. prior to any contractual agreement) the operator has wealth $W_O \geq 0$ and the supplier has wealth $W_S \geq 0$. Moreover, we assume that both parties will only consider contracts where they are guaranteed to end with non-negative wealth at the end of stage two. Therefore, the following conditions must be satisfied

$$W_O + \Pi_O\left(e_O, w, y, v, P, x\right) \geq 0 \qquad (\mathcal{WC_O})$$

and

$$W_S + \Pi_S\left(e_S, w, y, x\right) \geq 0 \qquad (\mathcal{WC_S})$$

where, $\Pi_O\left(e_O, w, y, v, P, x\right)$ and $\Pi_S\left(e_S, w, y, x\right)$ refer to the ex-post profits of the firms in the event of an operational failure (i.e. when $X = x$).

Firms generally write liability limits as a condition to enter into any contractual agreement; these constraints have been thoroughly studied in the literature (see §2). As Sappington (1983) explains, these clauses can be established as bankruptcy or insolvency provisions, where no further liability can be imposed on the firm if it reaches a state of insolvency. The liability limits can also be explicitly stated as fixed dollar amounts.[27] Saouma (2008) also uses these constraints to study outsourcing relationships, by arguing that "excessive liability resulting from a single faulty product can drive even large suppliers into bankruptcy." Similarly, Desiraju (2004) argues that a principal reason these clauses arise is due to "equity considerations that mandate the guarantee of an appropriate level of well-being" for the various parties involved in a contractual relationship.

[27] The following is a sample limited liability clause drawn from a contract for the Manufacturing of pharmaceutical inputs: *"Notwithstanding anything herein to the contrary, in no event will the GENERICO Indemnified Parties have any liability to NEW ALPHA or any of its Affiliates, or to any third party in connection with this Agreement, for monetary Damages in excess of \$100 million in the aggregate"* (Manufacturing Agreement between New Abraxis Inc. (New Alpha) and APP Pharmaceuticals, LLC. (Generico), November, 2007; Clause 12.9).

3.4 Analysis

In this section we study the interaction between liability-sharing through contracts and liability-sharing through insurance. We present a benchmark scenario in §3.4.1, by assuming that the operator and supplier are organized as a centralized entity. In §3.4.2 we consider the case where the firms operate as decentralized entities.

3.4.1 Benchmark Scenario: Centralized Supply Chain

When the operator and the supplier are organized as a centralized firm, a contract between these parties is unnecessary. The centralized supply chain thus finds the optimal level of effort and insurance coverage to maximize the total expected profits, $E[\Pi]$ (where $\Pi = \Pi_O + \Pi_S$). This supply chain has wealth $W = W_O + W_S$ and is unable to bear losses beyond this threshold. To find the optimal strategy for the centralized chain, we solve our model through backward induction.

Stage 2: Optimal Effort Levels:

In stage 2, the centralized chain jointly chooses effort levels to maximize

$$E_X\left[\Pi\left(e_S, e_O, v, P, X\right)\right] = \pi - \left(\left(1-v\right)E_X\left[X|e_S, e_O\right] + P + e_S + e_O\right)$$

The level of insurance coverage, v, and the insurance premium, P, have already been determined in stage 1. Note that since the premium is a fixed transfer to the insurer made in stage 1, the optimal levels of e_S and e_O do not depend on the premium P.[28]

Lemma 1. *The centralized supply chain sets effort levels $e_O^* = e_O^{\mathcal{FB}}$ and $e_S^* = e_S^{\mathcal{FB}}$ satisfying*

$$e_O^{\mathcal{FB}}\left(v\right) = \sqrt{\frac{x\left(1-v\right)\beta^{2-\beta}}{\left(1-\beta\right)^{1-\beta}}} - 1 \text{ and } e_S^{\mathcal{FB}}\left(v\right) = \sqrt{\frac{x(1-v)(1-\beta)^{1+\beta}}{\beta^{\beta}}} - 1 \quad (3.2)$$

Proof: All proofs are found in the appendix. ■

According to Lemma 1, the centralized supply chain sets effort levels such that these levels are proportional to the influence that each firm has in

[28] When the failure costs, x, are small enough, then e_O^* and e_S^* are non-positive. In these situations, none of the parties exert effort and the model becomes trivial. For this reason throughout this essay we focus on the case where x is large enough so that $e_S^* > 0$ and $e_O^* > 0$.

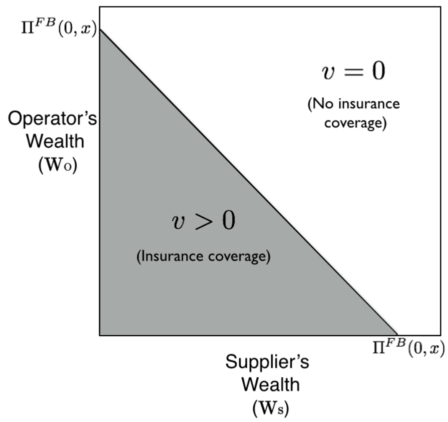


Figure 3.2: First best level of insurance.

mitigating the failure probability. To see this note that $\frac{e_S^{\mathcal{FB}}+1}{e_O^{\mathcal{FB}}+1} = \frac{1-\beta}{\beta}$, where β and $1-\beta$ represent the relative sensitivities of the failure probability to the operator's and supplier's efforts respectively. This ratio represents the first-best allocation of effort.

Stage 1: Optimal Insurance Coverage:

The centralized supply chain chooses the profit maximizing level of insurance coverage, subject to the constraint that its ex post wealth must be non-negative. Recall that the insurance premium is priced in accordance with the scheme specified in equation (3.1).

Let $\Pi^{\mathcal{FB}}(v,X) \equiv \Pi\left(e_S^{\mathcal{FB}}(v), e_O^{\mathcal{FB}}(v), v, P\left(v|e_S^{\mathcal{FB}}(v), e_O^{\mathcal{FB}}(v)\right), X\right)$ represent the profit function evaluated at the optimal effort levels. Hence, in stage 1 the centralized supply chain solves

$$\max_{v\in[0,1]} E_X\left[\Pi^{\mathcal{FB}}(v,X)\right]$$

subject to $\Pi^{\mathcal{FB}}(v,x) + W \geq 0$. The following proposition characterizes the first best level of insurance coverage.

Proposition 2. *If* $W+\pi \geq x+\sqrt{\frac{x}{\beta^\beta(1-\beta)^{1-\beta}}}-2$*, the centralized supply chain purchases no insurance coverage, i.e.* $v^*=0$*. If* $W+\pi < x+\sqrt{\frac{x}{\beta^\beta(1-\beta)^{1-\beta}}}-2$*, the centralized supply chain purchases insurance coverage* $v^* = v^{\mathcal{FB}} > 0$*, where* $v^{\mathcal{FB}}$ *solves* $\sqrt{\frac{x}{(1-v^{\mathcal{FB}})\beta^\beta(1-\beta)^{1-\beta}}} = W+\pi+2-x\left(1-v^{\mathcal{FB}}\right)$.

We illustrate this proposition in Figure 3.2. The centralized supply chain only purchases insurance if, given the realization of an operational accident, the potential losses are large enough to exceed the entire wealth of the chain, i.e. if $W + \Pi^{\mathcal{FB}}(0, x) = W + \pi - x - \sqrt{\frac{x}{\beta^{\beta}(1-\beta)^{1-\beta}}} + 2 < 0$. In other words, in a centralized setting insurance only plays the liquidity-enabling role, that is, to ensure the financial viability of the supply chain.

3.4.2 Decentralized Supply Chains

We now consider the case where the operator and supplier operate as decentralized entities. As in §3.4.1, we solve this model through backward induction.

Stage 2: Incentive Compatible Effort Levels:

In stage 2 the operator and supplier independently set the profit-maximizing levels of effort. The contract between the operator and supplier, (w, y), and the insurance contract, (v,P), have been determined in stage 1.

The supplier and the operator have best response functions $\tilde{e_S}|_{e_O} \equiv \arg\max_{e_S \geq 0} E_X \left[\Pi_S\left(e_S, w, y, X|e_O\right)\right]$ and $\tilde{e_O}|_{e_S} \equiv \arg\max_{e_O \geq 0} E_X \left[\Pi_B\left(e_O, w, y, v, P, X|e_S\right)\right]$. We derive the Nash equilibrium effort levels in Lemma 3.

Lemma 3. *The operator and the supplier exert effort levels e^*_O and e^*_S, where*

$$e^*_O(y, v) = \sqrt{\frac{x\left(\beta\left(1-y-v\right)\right)^{2-\beta}}{\left(\left(1-\beta\right)y\right)^{1-\beta}}} - 1 \quad \textit{and} \quad e^*_S(y, v) = \sqrt{\frac{x((1-\beta)y)^{1+\beta}}{(\beta(1-y-v))^{\beta}}} - 1$$

The effort levels depend on the penalty rate and the level of insurance coverage, y and v. This leads us to our next result, Lemma 4, which describes the failure probability as a function of these parameters. This lemma helps us simplify our analysis in subsequent sections.

Lemma 4. *Let $\Phi(y, v) \equiv F\left(e^*_S(y, v), e^*_O(y, v)\right)$ denote the failure probability as a function of the penalty, y, and the level of insurance coverage, v. We have that*

$$\Phi(y, v) \quad = \quad \left[x\left(\beta\left(1-y-v\right)\right)^{\beta}\left(\left(1-\beta\right)y\right)^{1-\beta}\right]^{-1}$$

Stage 1: Optimal Contracts

In this stage we derive the optimal operator-supplier contract, w and y, and the optimal level of insurance, v. The operator's problem is to solve

$$\max_{w\geq 0, y\in[0,1], v\in[0,1-y]} E_X\left[\Pi_O\left(e_O^*, w, y, v, P\left(v|e_S^*, e_O^*\right), X\right)|e_S^*\right]$$

subject to

$$E_X\left[\Pi_S\left(e_S^*, w, y, X\right)|e_O^*\right] \geq 0 \tag{$\mathcal{IR}$}$$

$$\Pi_S\left(e_S^*, w, y, x\right) + W_S \geq 0 \tag{$\mathcal{WC_S}$}$$

$$\Pi_O\left(e_O^*, w, y, v, P\left(v|e_S^*, e_O^*\right), x\right) + W_O \geq 0 \tag{$\mathcal{WC_O}$}$$

$\mathcal{IR}$ represents the individual rationality constraint for the supplier; $\mathcal{WC_S}$ and $\mathcal{WC_O}$ represent the wealth constraints for the supplier and for the operator. Note that the incentive compatibility constraints have already been embedded in the problem, through $e_S^* = \arg\max_{e_S} E_X\left[\Pi_S\left(e_S, w, y, X|e_O^*\right)\right]$ and $e_O^* = \arg\max_{e_O} E_X\left[\Pi_O\left(e_O, w, y, v, P, X|e_S^*\right)\right]$. To simplify our analysis we first present the results for three special cases:

- **Special case $\mathcal{UC}$ - unconstrained wealth:** we assume that $W_O = \infty$ and $W_S = \infty$.
- **Special case $\mathcal{SC}$ - Supplier with wealth constraints:** we assume that $W_O = \infty$ but $W_S < \infty$.
- **Special case $\mathcal{OC}$ - Operator with wealth constraints:** we assume that $W_O < \infty$ but $W_S = \infty$.

After analyzing these three special cases we present the general results, where $W_S \leq \infty$ and $W_O \leq \infty$.

Special Case $\mathcal{UC}$: Unconstrained Wealth

Assume that both parties are financially unconstrained, i.e. $W_S = W_O = \infty$. In this case, the wealth constraints $\mathcal{WC_O}$ and $\mathcal{WC_S}$ are never binding. We characterize the optimal contracts in Proposition 5.

Proposition 5. *When $W_O = \infty$ and $W_S = \infty$, the optimal contracting parameters are given by $v^* = v^{\mathcal{UC}}$, $y^* = y^{\mathcal{UC}}$ and $w^* = w^{\mathcal{UC}}$, where*

- $v^{\mathcal{UC}} = 0$

- $y^{\mathcal{UC}} = \begin{cases} 0.5 & \text{if } \beta = 0.5 \\ \frac{1-\beta^2-\sqrt{\beta(1-\beta)(1+\beta)(2-\beta)}}{1-2\beta} & \text{if } \beta \neq 0.5 \end{cases}$

- $w^{\mathcal{UC}} = xy^{\mathcal{UC}}\Phi\left(y^{\mathcal{UC}}, 0\right) + e_S^*\left(y^{\mathcal{UC}}, 0\right)$

In the absence of wealth constraints the decentralized supply chain does not purchase insurance coverage, i.e. $v^{\mathcal{UC}} = 0$. This is because insurance externalizes liability away from the supply chain and, therefore, reduces the incentives of each firm to invest in reliability. As a result, the operator finds it more profitable for the supply chain to internalize all the financial liability, especially given that the supply chain has adequate wealth to self-insure.

Since the failure probability depends both on the efforts of the operator and supplier, the operator transfers some of the liability to the supplier, by optimally choosing a penalty rate equal to $y^{\mathcal{UC}}$. As a result, the operator internalizes a proportion of the liability equal to $1-y^{\mathcal{UC}}$, and the supplier internalizes a proportion equal to $y^{\mathcal{UC}}$.

Under this penalty the ratio of efforts is equal $\frac{e_S^{\mathcal{UC}}+1}{e_O^{\mathcal{UC}}+1} = \sqrt{\left(\frac{1-\beta}{\beta}\right)^3\left(\frac{1+\beta}{2-\beta}\right)}$, where $e_O^{\mathcal{UC}} \equiv e_O^*\left(y^{\mathcal{UC}}, v^{\mathcal{UC}}\right)$ and $e_S^{\mathcal{UC}} \equiv e_S^*\left(y^{\mathcal{UC}}, v^{\mathcal{UC}}\right)$. The ratio is different from the first best ratio (recall that $\frac{e_S^{\mathcal{FB}}+1}{e_O^{\mathcal{FB}}+1} = \frac{1-\beta}{\beta}$), because the operator must coordinate the supply chain efforts by accounting for the double-sided moral hazard problem. For this reason, the penalty adjusts the effort levels so that the *marginal* effort of each firm is proportional to its influence in mitigating the failure probability. Observe that $\frac{\partial F\left(e_S^{\mathcal{UC}}, e_O^{\mathcal{UC}}\right)/\partial e_S}{\partial F\left(e_S^{\mathcal{UC}}, e_O^{\mathcal{UC}}\right)/\partial e_O} = \frac{1-\beta}{\beta}\left(\frac{e_O^{\mathcal{UC}}+1}{e_S^{\mathcal{UC}}+1}\right)$.[29]

Special Case $\mathcal{SC}$: Supplier with Wealth Constraints

We next look at the case where the supplier has finite wealth, but the operator has no wealth constraints, i.e. $W_S < \infty$ and $W_O = \infty$. This implies that constraint $\mathcal{WC}_{\mathcal{O}}$ is never binding.

In Proposition 6 we present our main results for this case. This proposition shows the existence of two key wealth thresholds for the supplier, denoted by W_S^I and W_S^{II} (where $W_S^I > W_S^{II}$). The threshold W_S^I is such that when the supplier's wealth W_S is greater than W_S^I, constraint $\mathcal{WC}_{\mathcal{S}}$

[29] This ratio is analogous to the one shown by Bhattacharyya and Lafontaine (1995) in their seminal paper on the nature of share contracts under double-sided moral hazard.

does not bind. When this happens, the operator's problem reduces to the unconstrained case ($\mathcal{UC}$). If on the other hand $W_S < W_S^I$, the supplier's wealth constraint binds at optimum. This means that the supplier is unable to sustain large losses. Given this inability, the operator optimally reduces the penalty rate that the supplier bears. The operator, however, *only* finds it optimal to purchase insurance coverage when W_S is less than the threshold W_S^{II}.

Proposition 6. *Assume that* $W_S < \infty$ *and* $W_O = \infty$. *Define* $W_S^I \equiv y^{\mathcal{UC}} x \left(1 - \Phi^{\mathcal{UC}}\right)$ and $W_S^{II} \equiv x\left(1-\beta\right)\left(1-2m\right)$, where m *satisfies* $\frac{m}{1-2m} = \frac{1-\beta}{1-\Phi(m,0)}$. *The optimal contracting parameters are given by* $v^* = v^{\mathcal{SC}}$, $y^* = y^{\mathcal{SC}}$ *and* $w^* = w^{\mathcal{SC}}$*, where*

- $$v^{\mathcal{SC}} = \begin{cases} 0 & \text{if } W_S \geq W_S^I \\ 0 & \text{if } W_S \in [W_S^{II}, W_S^I) \\ \max\left\{ \frac{(1-\beta)\left(1-2y^{\mathcal{SC}}\right)}{2-\beta}, \frac{x(1-\beta)\left(1-2y^{\mathcal{SC}}\right)-W_S}{x(1-\beta)+W_S/y^{\mathcal{SC}}} \right\} & \text{if } W_S < W_S^{II} \end{cases}$$

- $$y^{\mathcal{SC}} = \begin{cases} y^{\mathcal{UC}} & \text{if } W_S \geq W_S^I \\ \frac{W_S}{x(1-\Phi^{\mathcal{SC}})} & \text{if } W_S \in [W_S^{II}, W_S^I) \\ \max\left\{ \frac{W_S}{x(1-\Phi^{\mathcal{SC}})}, \frac{\Phi^{\mathcal{SC}}(1-\beta)}{2-\beta(1+\Phi^{\mathcal{SC}})} \right\} & \text{if } W_S < W_S^{II} \end{cases}$$

- $$w^{\mathcal{SC}} = \begin{cases} w^{\mathcal{UC}} & \text{if } W_S \geq W_S^I \\ xy^{\mathcal{SC}} + e_S^*\left(y^{\mathcal{SC}}, 0\right) - W_S & \text{if } W_S \in [W_S^{II}, W_S^I) \\ xy^{\mathcal{SC}} + e_S^*\left(y^{\mathcal{SC}}, v^{\mathcal{SC}}\right) - W_S & \text{if } W_S < W_S^{II} \end{cases}$$

where $\Phi^{\mathcal{SC}} \equiv \Phi\left(y^{\mathcal{SC}}, v^{\mathcal{SC}}\right)$.

We illustrate the above results in Figure 3.3. When $W_S \geq W_S^I$, the operator sets a penalty rate equal to $y^{\mathcal{UC}}$ and purchases no insurance coverage (as in the unconstrained case). If $W_S < W_S^I$ the supplier is unable to sustain large financial losses and, therefore, the operator has to decrease the penalty from $y^{\mathcal{UC}}$ to $y^{\mathcal{SC}}$. When the operator decreases the penalty, it takes away from the supplier a share of the failure costs. Note that this share is equal to $\left(y^{\mathcal{UC}} - y^{\mathcal{SC}}\right) X$.

The operator now needs to decide whether to: (i) absorb this share, by retaining it or; (ii) transfer some of it away from the supply chain, by purchasing insurance coverage. If $W_S \in [W_S^{II}, W_S^I)$, the operator chooses to retain this share, i.e. to internalize the liability. However, if the wealth

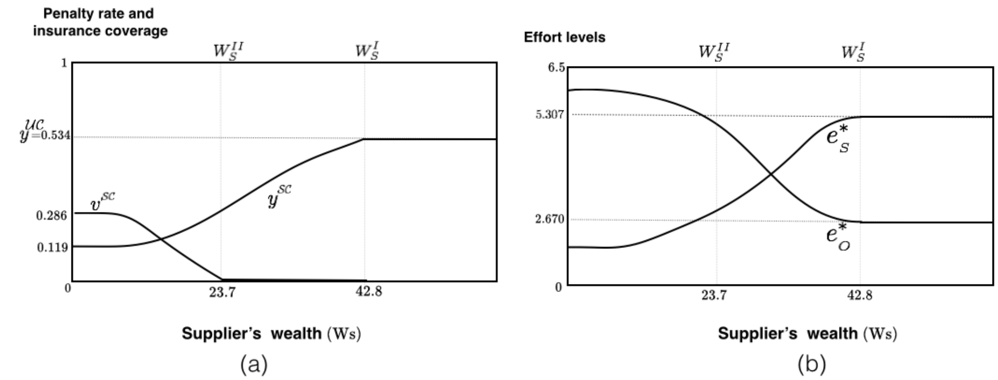


Figure 3.3: Optimal parameters - case $\mathcal{SC}$, for $\beta = 0.4$, $x = 100$,$\pi = 45$.

constraint of the supplier is very stringent, i.e. when $W_S < W_S^{II}$, the operator optimally transfers some of this share to the insurer by purchasing coverage.

It may seem optimal for a risk-neutral and wealth-unconstrained operator to always internalize the liability (instead of purchasing insurance). After all, when the operator internalizes the liability it has higher incentives to exert effort. On the other hand, when the operator purchases insurance, these incentives are reduced and the likelihood of an operational failure increases.

So why does the operator purchase insurance when the supplier's wealth constraint is very severe? The intuition is as follows. Because the efforts of the operator and supplier are partially substitutable, any increase in the effort of the operator has a negative externality on the incentives of the supplier. In other words, when the operator chooses to keep the extra share of the liability, the supplier knows that the operator has higher incentives to increase effort (in stage 2). Therefore, the supplier has an incentive to free-ride on the efforts of its counterpart and, as a result, it further reduces its own effort levels. This implies that the increase in the operator's effort is partially offset by a reduction in the effort of the supplier. In other words, when the operator internalizes the extra share of the liability, the supply chain is subject to an effort distortion.

Consider instead what happens when the operator chooses to externalize the failure costs through insurance. In this case the overall incentives to exert effort are reduced. Hence, the operator needs to trade-off the effort distortion caused by the free-riding problem with the reduction in the overall incentives to exert effort (caused by insurance). Our results show that when the supplier's wealth constraint is not stringent, i.e. when $W_S \in [W_S^{II}, W_S^{I})$,

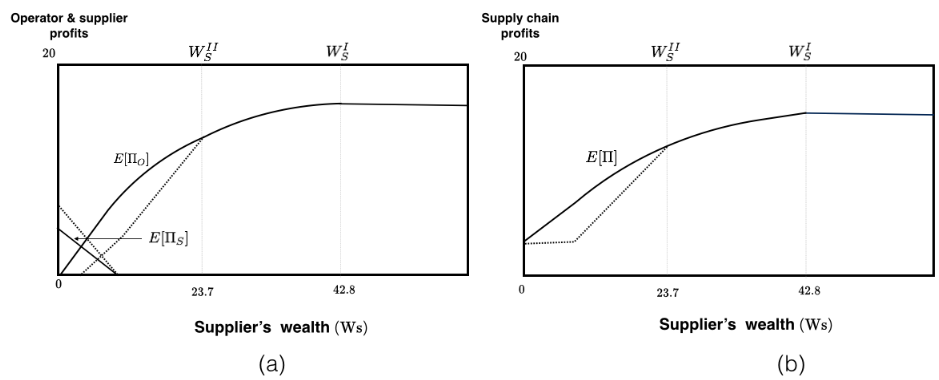


Figure 3.4: Welfare parameters - case $\mathcal{SC}$

the effort distortion in the supply chain is preferred to the dampening in efforts. In other words, it is better to internalize all liability. When the supplier's wealth constraint is very stringent however, i.e. when W_S is less than the threshold W_S^{II}, it is better to dampen the effort incentives by purchasing insurance than to further distort the supply chain efforts (see Figure 3.3). The operator can achieve this by seeking insurance as a commitment not to increase effort. This effectively reduces the supplier's incentives to free-ride and, therefore, mitigates the effort distortion in the supply chain.

Note that the operator may choose to buy insurance, not as a mechanism to ensure the financial viability of the supply chain, but rather as a credible commitment mechanism not to increase effort. Insurance allows the operator to better coordinate effort along the supply chain. This is what we refer to as the strategic role of business insurance.

As we can see from Figure 3.4, the introduction of insurance opportunities increases the welfare for the operator (relative to a scenario where insurance is not available). Conversely, the supplier's profits decrease with insurance. This is because insurance decreases the free-riding opportunities for the supplier. We can also observe that the overall supply chain profits increase.

Special Case $\mathcal{OC}$: Operator with Wealth Constraints

We next move to the case where the operator has finite wealth, but the supplier has no wealth constraints, i.e. $W_O < \infty$ and $W_S = \infty$. Therefore, constraint $\mathcal{WC_S}$ never binds. In Proposition 7 we present our main results for case $\mathcal{OC}$, where we show that (as in case $\mathcal{SC}$) insurance may be purchased for strategic reasons. The intuition behind this result, however, is slightly different from the one presented in the previous case.

In the proposition below, we again show the existence of two wealth thresholds. This time the thresholds, W_O^I and W_O^{II}, are for the operator. The operator's wealth constraint binds if W_O is less than W_O^I and, in this situation, the operator optimally increases the penalty rate that the supplier bears. The operator purchases insurance only when $W_O < W_O^{II}$.

Proposition 7. *Assume that* $W_O < \infty$ *and* $W_S = \infty$. *Define* $W_O^I \equiv x\Phi^{\mathcal{UC}}\left(2y^{\mathcal{UC}}(1-\beta)+\beta\right) + x\left(1-y^{\mathcal{UC}}\right) - 2 - \pi$ *and* $W_O^{II} \equiv x\Phi(n,0)(2n(1-\beta)+\beta)+x(1-n)-2-\pi$, *where* n *satisfies* $\frac{(n(1-\beta))^{\beta-1}}{x(\beta(1-n))^{\beta}} = \left(1+\beta\left(\frac{1-2n}{1-n}\right)\right)^2$. *The optimal contracting parameters are given by* $v^* = v^{\mathcal{OC}}$, $y^* = y^{\mathcal{OC}}$ *and* $w^* = w^{\mathcal{OC}}$, *where*

- $$v^{\mathcal{OC}} = \begin{cases} 0 & \text{if } W_O \geq W_O^I, \\ 0 & \text{if } W_O \in [W_O^{II}, W_O^I) \\ \left(\frac{1}{1-\beta}\right)\left(\frac{y^2}{2} - \frac{(1-y)(W_O+2-xy)(1-\beta)-x}{x\beta}\right)^{\frac{1}{2}} & \text{if } W_O < W_O^{II} \\ -\beta + 2y\left(\frac{3}{4}+\beta\right) & \end{cases}$$

- $$y^{\mathcal{OC}} = \begin{cases} y^{\mathcal{UC}} & \text{if } W_O \geq W_O^I \\ \Phi^{\mathcal{OC}}(2(1-\beta)+\beta)+1-\frac{2+W_O}{x} & \text{if } W_O \in [W_O^{II}, W_O^I) \\ \frac{\Phi^{\mathcal{OC}}-1-\beta\left(1-v^{\mathcal{OC}}\right)}{\Phi^{\mathcal{OC}}-2} & \text{if } W_O < W_O^{II} \end{cases}$$

- $$w^{\mathcal{OC}} = \begin{cases} w^{\mathcal{UC}} & \text{if } W_O \geq W_O^I \\ x\Phi^{\mathcal{OC}}y^{\mathcal{OC}} + e_S^*\left(y^{\mathcal{OC}}, 0\right) & \text{if } W_O \in [W_O^{II}, W_O^I) \\ x\Phi^{\mathcal{OC}}y^{\mathcal{OC}} + e_S^*\left(y^{\mathcal{OC}}, v^{\mathcal{OC}}\right) & \text{if } W_O < W_O^{II} \end{cases}$$

where $\Phi^{\mathcal{OC}} \equiv \Phi\left(y^{\mathcal{OC}}, v^{\mathcal{OC}}\right)$.

We illustrate Proposition 7 in Figure 3.5. When the operator's wealth constraint binds, the operator is unable to bear large financial losses. To satisfy this constraint, the operator needs to transfer away a higher share of the failure costs (relative to the unconstrained case $\mathcal{UC}$). The operator can transfer this share to the supplier, by increasing the supplier's penalty rate above $y^{\mathcal{UC}}$ or, alternatively, transfer this share away from the supply chain through insurance.

Similar to case $\mathcal{SC}$, when the supplier internalizes the extra share of the liability (due to a higher penalty rate) not only does the supplier have larger

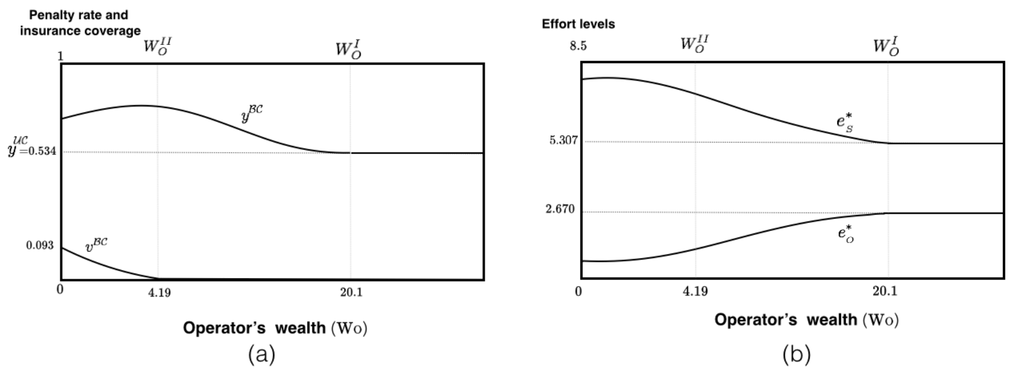


Figure 3.5: Optimal contract parameters - case $\mathcal{OC}$, for $\beta = 0.4$, $x = 100$,$\pi = 45$.

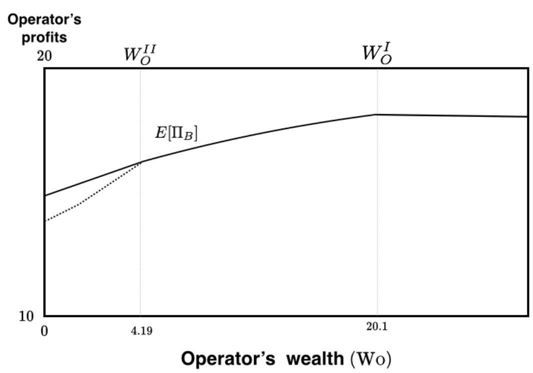


Figure 3.6: Welfare parameters - case $\mathcal{OC}$

incentives to increase effort, but the operator also has a larger incentive to free-ride on the supplier's efforts. This implies that any increase in the efforts of the supplier is partially offset by a further decrease in the efforts of the operator. Therefore, the operator faces a trade-off between distorting the efforts in the supply chain (by increasing the supplier's penalty rate, and exacerbating the free-riding problem) and dampening the supplier's incentives (by purchasing insurance). However, in case $\mathcal{OC}$ the operator is the party that free-rides on the effort of the supplier. This is unlike case $\mathcal{SC}$, where the supplier is the free-rider.

But if the operator is the one free-riding, why would it want to mitigate the free-riding problem? The intuition is as follows. Note that the operator coordinates the supplier's efforts through the penalty rate, y, and compensates these efforts through a fixed-fee transfer w. Therefore, by optimally choosing y and w, the operator extracts all rents in the supply chain. When the free-riding problem becomes severe enough, the effort distortion becomes very large. Because of this inefficiency the operator is able to extract fewer rents from the supplier.

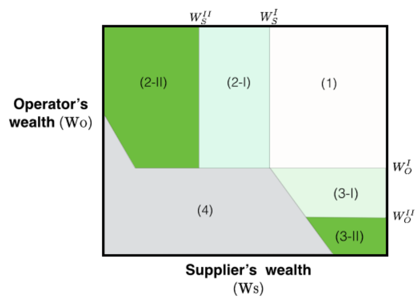
Due to the argument above, the operator optimally trades off the distortion in the supply chain efforts (by transferring the liability to the supplier) or the dampening of the supplier's incentives (by purchasing insurance). When the operator's wealth constraint is not stringent, a distortion of the supply chain efforts is preferred to a dampening of the supplier's incentives. For this reason, the operator chooses to increase the penalty rate, instead of purchasing insurance. However, when $W_O < W_O^{II}$, a dampening in the supplier's incentives is preferred to an effort distortion (and the operator purchases insurance coverage). In Figure 3.6 we can observe that when $W_O < W_O^{II}$, the operator's (and, hence, the supply chain's) welfare is improved through insurance.

The General Case: Supplier and Operator with Wealth Constraints

We now assume that both the supplier and the operator are subject to wealth constraints, i.e. that $W_S \leq \infty$ and $W_O \leq \infty$. In Proposition 8 we show the existence of four regions. In the first region, none of the wealth constraints bind. This leads to the unconstrained case $\mathcal{UC}$. In the second and third regions, one of the wealth constraints is binding but the other is not. These lead to the cases $\mathcal{SC}$ and $\mathcal{OC}$.

By looking at the general model, however, we must consider a new region. This is the region where both constraints bind at optimum. In this region the operator purchases insurance, not for strategic reasons, but rather to ensure the financial viability of the supply chain. In other words the operator's problem is not feasible without insurance. This is unlike the other regions, where the supply chain can run operations without insurance coverage.

Proposition 8. *Assume that $W_S \leq \infty$ and $W_O \leq \infty$ and define $\tilde{W}(y,v) \equiv x(1 + \Phi(y,v)(\beta(1-y-v) + y(1-\beta))) - \pi$. The optimal contracting parameters are*



	No insurance	Insurance purchased (strategic role)	Insurance purchased (liquidity-enabling role)
$\mathcal{WC}_O$ and $\mathcal{WC}_S$ redundant	(1)		
$\mathcal{WC}_S$ binding	(2-I)	(2-II)	
$\mathcal{WC}_O$ binding	(3-I)	(3-II)	
$\mathcal{WC}_S$ and $\mathcal{WC}_O$ binding			(4)

Figure 3.7: Contracting regions for the general case.

- $$y^* = \begin{cases} y^{\mathcal{UC}} & \text{if } W_S > W_S^I \text{ and } W_O > W_O^I \\ y^{\mathcal{SC}} & \text{if } W_S \leq W_S^I \text{and } W_O > \tilde{W}\left(y^{\mathcal{SC}}, v^{\mathcal{SC}}\right) - W_S \\ y^{\mathcal{OC}} & \text{if } W_O \leq W_O^I \text{and } W_S > \tilde{W}\left(y^{\mathcal{OC}}, v^{\mathcal{OC}}\right) - W_O \\ \frac{W_S}{x(1-\Phi^*)} & \text{otherwise} \end{cases}$$

- $$v^* = \begin{cases} v^{\mathcal{UC}} & \text{if } W_S > W_S^I \text{ and } W_O > W_O^I \\ v^{\mathcal{SC}} & \text{if } W_S \leq W_S^I \text{and } W_O > \tilde{W}\left(y^{\mathcal{SC}}, v^{\mathcal{SC}}\right) - W_S \\ v^{\mathcal{OC}} & \text{if } W_O \leq W_O^I \text{and } W_S > \tilde{W}\left(y^{\mathcal{OC}}, v^{\mathcal{OC}}\right) - W_O \\ 1 - y\left(1+\beta\right) & \text{otherwise} \\ -\frac{W_S + W_O + \pi - x}{x\beta\Phi^*} & \end{cases}$$

- $$w^* = \begin{cases} w^{\mathcal{UC}} & \text{if } W_S > W_S^I \text{ and } W_O > W_O^I \\ w^{\mathcal{SC}} & \text{if } W_S \leq W_S^I \text{and } W_O > \tilde{W}\left(y^{\mathcal{SC}}, v^{\mathcal{SC}}\right) - W_S \\ w^{\mathcal{OC}} & \text{if } W_O \leq W_O^I \text{and } W_S > \tilde{W}\left(y^{\mathcal{OC}}, v^{\mathcal{OC}}\right) - W_O \\ x\Phi^* y^* + e_S^*\left(y^*, v^*\right) & \text{otherwise} \end{cases}$$

We illustrate these results in Figure 3.7. In region 1 the wealth constraints are non-binding. In region 2, $\mathcal{WC}_O$ does not bind, but $\mathcal{WC}_S$ binds at optimum. Note that in sub-region 2-I, we have that $W_S \in [W_S^{II}, W_S^I)$. Therefore, by Proposition 6, the operator does not purchase insurance. Conversely, in sub-region 2-II we have that $W_S < W_S^{II}$ and the operator purchases insurance. In region 3, the supplier's wealth constraint is non-binding, but the operator's constraint is binding. In sub-region 3-I we have that $W_O \in [W_O^{II}, W_O^{I})$ and, by Proposition 7, the operator does not purchase insurance. Conversely, in sub-region 3-II we have that $W_O < W_O^{II}$ and the operator purchases insurance. In region 4, insurance serves a liquidity-enabling role.

3.4.3 Summary of Results

In §3.4.1 we show that a centralized supply chain only purchases insurance as a way to ensure its financial viability (i.e. the liquidity-enabling role). In other words, the centralized chain purchases insurance coverage if and only if the potential costs of an operational failure exceed the wealth of the chain, $W = W_S + W_O$.

We show that the liquidity-enabling role also arises in a decentralized supply chain (see region 4 in Figure 3.7). However, in a decentralized setting insurance is purchased even in situations where the supply chain has enough wealth to cover any ex post losses associated with an operational failure (see regions 2-II and 3-II). In these cases, the traditional explanations for why firms purchase insurance are absent, but nonetheless insurance is optimally purchased. Our essay shows that in these cases insurance serves as a commitment mechanism to mitigate a free-riding problem. The free-riding problem arises when one of the firms has wealth constraints, but the other firm has sufficient wealth to sustain large losses.

3.5 Extensions

3.5.1 Sequential Effort

In the base model we assume that the efforts of the supplier and operator are unobservable and, as we explain in Section 3.3.3, we can treat the efforts as simultaneous - even if these efforts are sequentially exerted. This is by the *Principle of Interchange of Moves* (Osborne and Rubinstein, 1994).

But the assumption of unobservability does not always hold. As a case in point, consider a supplier that is in charge of designing an equipment that is subsequently managed by the operator: the supplier will be the first party to exert effort, i.e. in the design of the equipment, and the operator will follow, i.e. by exerting effort when operating the equipment. The equipment operator may be able to infer the effort of the supplier by inspecting the quality of the equipment.[30] In these situations it is reasonable to assume a Stackelberg relationship in the exertion of the efforts.

We now study an extension where the supplier leads in the exertion of effort, and the operator follows after observing the effort of the supplier (i.e. a Stackelberg dynamic). We begin by presenting the incentive compatible levels of effort in Lemma 9.

Lemma 9. *If the efforts of the parties are Stackelberg (and the supplier leads), the operator and the supplier exert effort levels* $e^*_{0,seq}$ *and* $e^*_{S,seq}$*, where*

$$e^*_{O,seq}(y,v) = \sqrt{\frac{x\left(\beta\left(1-y-v\right)\right)^{2-\beta}}{\left(\left(\frac{1-\beta}{1+\beta}\right)y\right)^{1-\beta}}}-1$$

$$e^*_{S,seq}(y,v) = \sqrt{\frac{x\left(\frac{(1-\beta)}{1+\beta}y\right)^{1+\beta}}{\left(\beta\left(1-y-v\right)\right)^{\beta}}}-1$$

From the lemma above, we can verify that the ratio of efforts is equal to

$$\frac{e^*_{O,seq}+1}{e^*_{S,seq}+1} = \frac{(1+\beta)\,\beta}{(1-\beta)}\left(\frac{(1-y-v)}{y}\right) \tag{3.3}$$

Note that the ratio of efforts obtained in the base model (from Lemma 3) is equal to $\frac{e^*_O+1}{e^*_S+1} = \frac{\beta}{(1-\beta)}\frac{(1-y-v)}{y}$. By comparing these two ratios, we can notice that when the supplier leads in the exertion of effort, there is an additional distortion in ratio of these efforts (i.e. $\frac{e^*_{O,seq}+1}{e^*_{S,seq}+1} = (1+\beta)\left(\frac{e^*_O+1}{e^*_S+1}\right)$). This means that for given levels of y and v, the Stackelberg advantage allows the supplier to free-ride on the operator to a larger degree. We refer to this as a *Stackelberg distortion*.

So how does this assumption affect the results of base model? We find numerically that if the supplier is the wealth-constrained party (i.e. case $\mathcal{SC}$), the levels of insurance are greater in a Stackelberg relationship. Conversely, if the operator is the wealth constrained party (i.e. case $\mathcal{OC}$), the levels of insurance are higher when the efforts are simultaneously exerted.

This result is intuitive. To understand why, recall from the base model that in case $\mathcal{SC}$, the wealth constraints of the supplier cause a distortion

[30] Note that we are assuming that effort is observable *but* unverifiable in a court of law. This means that effort is not contractible. This is a commonly observed assumption (see page 38 in Tirole, 1988). If the efforts of the firms are verifiable, these efforts can be contracted and the model becomes trivial.

that *increase*s the ratio of efforts between the operator and the supplier. So when the relationship is Stackelberg, this distortion is multiplied by the Stackelberg distortion. Hence, the overall distortion is larger, and more insurance is needed to correct this problem.

Conversely, recall that in case $\mathcal{OC}$ the wealth constraints of the operator cause a *decrease* in the ratio of effort between the supplier and the operator. But when effort is sequential, the Stackelberg distortion moves in the opposite direction, and counteracts the distortion generated by the wealth constraints. As such, less insurance is needed to correct this problem.

In all scenarios the supply chain profits decrease when the effort is sequentially exerted, given that the supplier (who is the agent) gains an advantage on the operator.

3.5.2 Insurance Opportunities for the Supplier

In the base model we do not consider a scenario where the supplier (who is the agent) is able to purchase insurance. This assumption allows us to focus on the coordinating role that insurance plays for business operators. But this assumption may not often hold in practice, as buying insurance is an option that is also available to other parties in the supply chain.

To address this issue, we study a setting where both the supplier and the operator have the option of (independently) obtaining insurance coverage from third-parties.[31] In this extension we seek to address two questions: (i) when, and for what reasons, would the supplier buy insurance? and; (ii) how does this possibility affect the strategic role of insurance for the operator? The following proposition allows us to answer both questions.

Proposition 10. *If the supplier has the option of purchasing insurance coverage:*

1. The supplier will choose coverage level $v_S^* = \frac{\beta y}{1+\beta}$.
2. Under coverage level v_S^*, the ratio of efforts between the operator and supplier will be identical to the ratio given by equation (3.3) from section 3.5.1.

[31] Note that in cases where (i) multiple parties are involved in an operational failure, and (ii) the degree of fault is ambiguous, insurance companies have 'knock-for-knock' agreements, where the insurers agree to reimburse the losses for which their respective policy holders are responsible in their procurement agreements, regardless of fault (Lilleholt et al., 2012).

The above proposition tells us that the supplier will purchase insurance regardless of whether the parties are wealth-constrained or not. Specifically, the supplier will use insurance as a commitment not to exert effort, which will drive the operator to increase its own effort. However, the use of insurance by the supplier will not be used to mitigate distortions in the efforts of the supply chain but, rather, to further distort these efforts in its favour. This is because the supplier is only concerned about its own profits, not about the efficiency of the supply chain. The supplier thus uses insurance as a device to increase its capability to free-ride at the expense of the operator.

Part 2 of the above proposition tells us that when the supplier buys insurance, the ratio of efforts is identical to the case where the efforts are sequential. In other words, the supplier gains a Stackelberg advantage on the operator by purchasing insurance. This is because insurance allows the supplier to commit to a lower level of effort. This means that the analysis of this extension is qualitatively similar to the one from §3.5.1. Specifically, we find that the distortion generated by the supplier's insurance coverage increases the need for insurance for the operator in case $\mathcal{SC}$, but decreases it in case $\mathcal{OC}$. The reasoning is similar to the one presented in §3.5.1.

The supply chain profits decrease in all cases when the supplier has access to insurance coverage. This is because the supplier (i.e. the agent) gains a strategic advantage on the operator and worsens the free-riding problem in the supply chain. For this reason, it may be reasonable for an operator to try to impose a condition that supplier not purchase insurance, which is consistent with the assumption made in the base model.

3.5.3 Stochastic Failure Costs

In this section we relax the assumption that the failure costs are ex ante known. To this end, let $X \in \{0, [x_l, x_h]\}$ represent the failure costs. If $X = 0$, operations have run as planned and the operator does not incur any costs. If $X > 0$ a failure has occurred, and the costs are equal to $x \in [x_l, x_h]$, where $0 < x_l < x_h$. Assume that X has a probability atom at 0. Specifically, $Pr\left(X = 0\right) = 1 - F\left(e_S, e_O\right)$ where $F\left(e_S, e_O\right)$ is defined as in the base model.

By looking at the case where the failure costs are stochastic, we can explore other settings. For example, tiered contracts (i.e. contracts where the penalty levels depend on the realized costs) are common in practice but not considered in our main model. Second, under this assumption we can study the role of insurance deductibles in the insurance contract.

Tiered Contracts:

Jain et al. (2013b) show that in the presence of financial constraints, contracts with tiered penalties are significantly more powerful in mitigating double-moral hazard. This is because the design of tiered penalties allows the operator to coordinate efforts without causing an excessive financial burden on the wealth-constrained firm. For example, if the supplier has wealth constraints, the operator can optimally increase the supplier's penalty for small failures, and decrease the penalty for large failures. The *expected* penalty for the supplier thus remains unchanged and, at the same time, the financial distress of the supplier is mitigated. For this reason tiered contracts are frequently used in practice.[32]

We assume that the operator designs two tiers, $y_1 \in [0,1]$ and $y_2 \in [0,1]$, and a tier threshold $x_T \in [x_l, x_h]$. When $X \leq x_T$, the operator penalizes the supplier with a penalty rate equal to $y_1 X$. When $X > x_T$, the size of the penalty is $y_2 X$. The cash flows of the procurement contract (from the operator to the supplier) are

$$T\left(w, y_1, y_2, x_T, X\right) = \begin{cases} w - Xy_1 & \text{if } X \leq x_T \\ w - Xy_2 & \text{if } X > x_T \end{cases}$$

We study this model by deriving some analytical results and by running several numerical simulations. We first do comparative statics by varying the wealth of the parties, under the assumption that the failure costs are uniformly distributed. A set of results is illustrated in Tables 3.1 and 3.2. In these tables we present the optimal parameters under a contractual structure involving tiered contracts and insurance $(x_T, y_1, y_2; v)$. For comparison purposes, we also present the optimal parameters under a structure involving a simple penalty rate and insurance $(y; v)$. In the right-most column of these tables we present $\Delta\Pi$, which shows by how much does the welfare of the supply chain increase when the operator uses tiered contracts. We obtain the following results.

First, we find the operator can efficiently coordinate efforts through tiered penalties and, at the same time, mitigate the impact of the wealth constraints

[32] For example, in a joint operations agreement signed in 2001 between The Union Oil of California, the operator, and Ivanhoe Energy, the service supplier, the contractor uses tiered penalties for any costs incurred in the event of oil spills, blowouts, fires, etc. The operator will be compensated for:

(A) 5 % of total costs through $100,000; plus

(B) 3 % of total costs in excess of $100,000 but less than $1,000,000; plus

(C) 2 % of total costs in excess of $1,000,000. (Section III, clause 2)

W_O	x_T	y_1	y_2	v	y	v	$\Delta\Pi$
110	0	0.5	0.5	0	0.51	0	0
90	82.34	0.05	0.59	0	0.60	0	0.32
70	100	0.41	0.70	0	0.71	0	1.09
60	100	0.53	0.76	0	0.69	0.06	1.49
50	100	0.60	0.75	0.08	0.67	0.16	2.06
40	100	0.61	0.68	0.23	0.63	0.29	2.57

Table 3.1: Optimal tiered and simple contracts - Case $\mathcal{OC}$

W_S	x_T	y_1	y_2	v	y	v	$\Delta\Pi$
80	0	0.5	0.5	0	0.5	0	0
70	99.71	0.80	0.39	0	0.38	0	0.38
60	100	0.57	0.28	0	0.27	0	1.04
50	100	0.46	0.23	0	0.23	0.05	1.39
40	100	0.17	0.17	0.06	0.17	0.13	1.53
30	100	0.23	0.11	0.17	0.11	0.21	1.85

Table 3.2: Optimal tiered and simple contracts - Case $\mathcal{SC}$

(i.e. the free-riding problem). However, when the wealth of the operator, or the supplier, is small enough, a tiered contract is unable to eliminate excessive free-riding among the contractual parties. In these cases the operator still purchases insurance coverage, but the amount of coverage is smaller.

Second, we study the sensitivity of our results to the variance of the failure costs.[33] We find that the role of insurance decreases when the variance of the failure costs increases (see Tables 3.3 and 3.4). This is because when the failure costs have a large variance, the operator has more flexibility to distribute the failure costs through tiers. As such, tiered penalties can be used more effectively when the failures costs have a large variance.

Insurance Deductibles:

In this section we study the robustness of our model to more complex insurance contracts, by allowing the operator to not only choose the level of insurance coverage, v, but also a deductible level, $d \geq 0$. The insurance cash

[33]To this end, we symmetrically shift x_l and x_h in opposite directions, so that $x_l + x_h$ remains unchanged. For example, we look at a scenario where the expected failure costs are equal to 100, but we change the variance of the failure costs. To do this, we look at various cases: $x_l = x_h = 100$; $x_l = 75$ and $x_h = 125$; $x_l = 50$ and $x_h = 150$ and; $x_l = 25$ and $x_h = 175$.

x_l	x_h	x_T	y_1	y_2	v
25	175	93.73	0.59	0.73	0
50	150	95.48	0.63	0.75	0.05
75	125	98.63	0.68	0.75	0.10
100	100	100	0.76	0.76	0.14

Table 3.3: Simulation of optimal tiered contracts for different damage variances (case $\mathcal{SC}$)

x_l	x_h	x_T	y_1	y_2	v
25	175	93.72	0.25	0.13	0.07
50	150	95.54	0.25	0.16	0.11
75	125	98.13	0.25	0.20	0.14
100	100	100	0.25	0.25	0.17

Table 3.4: Simulation of optimal tiered contracts for different damage variances (case $\mathcal{OC}$)

flows are thus equal to $I(v,d,P,X) = P - \max\{vX - d, 0\}$. If we assume that the conditional probability of X, given $X > 0$, is uniform, then the cash flows of the insurance contract are given by $E_X[I(v,d,P,X)|e_S,e_O] = P - F(e_S,e_O)\int_{x_l}^{x_h} \frac{\max\{vx-d,0\}}{x_h-x_l}dx$

We analyzed this model both analytically and numerically. We found that insurance deductibles are useless when the supplier has binding wealth constraints, but the operator's wealth constraint is non-binding (i.e. case $\mathcal{SC}$). In this case, there always exists an optimal insurance contract where the deductible is equal to 0. When the wealth constraint of the operator is binding, and the wealth constraint of the supplier is non-binding (i.e. special case $\mathcal{OC}$), positive deductibles allow the operator to mitigate free-riding problem more efficiently. Specifically, we find that insurance deductibles increase the range where insurance is strategically purchased. This is because a positive deductible allows the operator to seek high levels of insurance when the losses exceed its financial wealth. However, when these losses are small enough, the (risk-neutral) operator does not benefit from insurance. Therefore, the operator uses a deductible to mitigate the impact of moral hazard, and seek high levels of insurance when needed (see Table 3.5). When the operator is wealth-unconstrained, and the supplier is wealth-constrained, this role is absent. The operator only benefits from deductibles when its wealth constraints are binding.

3.5.4 Alternatives to the Wealth Constraints

The wealth constraints used in the base model assume that the parties will not enter the contractual relationship unless they are guaranteed to end up with non-negative wealth under all contingencies. These constraints are common in practice and have been extensively analyzed in the literature. In many cases, however, firms cannot entirely avoid the possibility of

W_O	y	v	d	y	v	$\Delta\Pi$
120	0.51	0	0	0.51	0	0
100	0.60	0	0	0.60	0	0.32
80	0.53	0.46	59.54	0.71	0	0.70
70	0.54	0.46	48.71	0.69	0.06	1.49
60	0.55	0.45	36.80	0.67	0.16	2.06
50	0.55	0.44	23.24	0.63	0.29	2.57

Table 3.5: Simulation of optimal insurance contracts with deductibles.

bankruptcy but this does not prevent them from engaging in operational activities.

Rather than avoiding the possibility of bankruptcy entirely, firms sometimes adopt other criteria to maximize their profits and, at the same time, minimize their exposure to insolvency risks. To model this behavior, the literature has considered a number of alternative approaches. In this subsection we consider two popular approaches: the financial distress approach and the cost of bankruptcy approach.

Financial Distress Approach: According to the financial distress approach, firms may be averse to a state of insolvency, but are willing to tolerate this possibility if the likelihood is sufficiently small. In other words, the firm will engage in a contractual relationship only if the probability of insolvency is below a *tolerance threshold* $\alpha \in [0, 1]$. We write these constraints as follows:

$$\Pr\left[\Pi_O\left(e_O, w, y, v, P, X\right) + W_O \leq 0\right] \leq \alpha_O$$
$$\Pr\left[\Pi_S\left(e_S, w, y, X\right) + W_S \leq 0\right] \leq \alpha_S$$

We can easily verify that when the levels of tolerance (α_O and α_S) are equal to zero, the financial distress constraints are equivalent to the wealth constraints. Through numerical results we show that when the tolerance to distress increases (i.e. when α_S and α_O increase), the levels of insurance are smaller (see, Figure 3.8 for an illustration). In other words, the wealth constraints can be seen as an extreme version of the financial distress approach. Overall, there are no qualitative changes in the results.

Cost of Bankruptcy: Both the wealth constraints and the financial distress constraints reflect the idea that firms only consider decisions where the risk of insolvency is either small or absent. Other papers consider this possibility by incorporating the costs of negative wealth into the objective

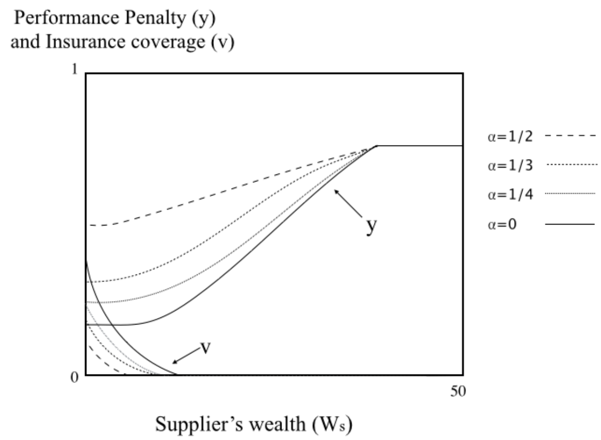


Figure 3.8: Optimal contract parameters for varying levels of α.

function, rather than expressing them as constraints (See Greenwald and Stiglitz (1990), Swinney et al. (2011) and references therein). Under this approach, we assume that the supplier and the operator maximize the utility functions,

$$\begin{aligned} U_S\left(e_S, y, w, D_S, X\right) &\equiv \Pi_S\left(e_S, w, y, X\right) \\ &\quad -D_S\psi_S\left(e_S, y, w, W_S, X\right) \\ U_O\left(e_O, w, y, v, P, D_O, X\right) &\equiv \Pi_O\left(e_O, w, y, v, P, X\right) \\ &\quad -D_O\psi_O\left(e_O, w, y, v, P, W_O, X\right) \end{aligned}$$

where D_S and D_O are the (exogenously given) bankruptcy costs for each firm, and $\psi_S\left(e_O, w, y, W_S, X\right) \equiv \Pr\left\{\Pi_S\left(e_S, w, y, X\right) + W_S \leq 0\right\}$ and $\psi_O\left(e_O, w, y, v, P, W_O, X\right) \equiv \Pr\left\{\Pi_O\left(e_O, w, y, v, P, X\right) + W_O \leq 0\right\}$ are the bankruptcy probabilities. U_O and U_S are known as *integrated objective functions.*[34] To analyze this setting, we consider a model where the operator seeks to find

$$\max_{w\geq 0, y\in[0,1], v\in[0,1]} E_X\left[U_O\left(e_O^*, w, y, v, P\left(v|e_S^*, e_O^*\right), D_O, X\right)|e_S^*\right]$$

[34] This approach captures the idea that reaching a state of insolvency brings out non-trivial costs. For example, the firms may be forced to sell their illiquid assets at low prices to repay their debts. An insolvent firm will also have to pay for auditor and litigation fees in the event of filing bankruptcy.

subject to

$$\begin{aligned} & E_X\left[U_S\left(e_S^*, y, w, D_S, X\right) | e_O^*\right] \geq 0 \\ e_S^* &= \arg\max_{e_S \geq 0} E_X\left[U_S\left(e_S, y, w, D_S, X\right) | e_O^*\right] \\ e_O^* &= \arg\max_{e_O \geq 0} E_X\left[U_O\left(e_O, w, y, v, P, D_O, X\right) | e_S^*\right] \end{aligned}$$

We performed numerical simulations using this model and found that when the cost of bankruptcy is large for one party, but small for the other, the operator optimally buys insurance for its strategic value. Consider the case where $D_O = 0$. If D_S is large, the supplier will weigh in the costs of bankruptcy at the time of exerting effort. To mitigate this inefficiency, the operator optimally decreases the penalty rate, which decreases the probability of bankruptcy for the supplier. In turn, this will generate a distortion in the effort of both parties. When D_S is too large, the operator is forced to purchase insurance as a way to commit not to increase effort and, thereby, to decrease the distortion in the effort among the parties. The intuition is similar to the base model.

Insurance will be purchased if the cost of bankruptcy is high for the operator and low for the supplier, and vice versa. However, if the costs of bankruptcy are high for both parties, insurance will not be purchased. This is because the distortion in the efforts of the operator is counteracted by the distortion in the efforts of the supplier. In this case, the supply chain efficiency decreases, but the efforts are not distorted (i.e. there is no excessive free-riding problem).

3.6 Conclusion

In this essay we study a context where the interdependent (and unobservable) efforts of firms in a supply chain can mitigate the likelihood of an operational failure. We show that firms may purchase insurance for strategic reasons. This happens when one of the firms has severe wealth constraints, but the other firm has sufficiently large wealth to cover potential losses. In this situation, contractual incentives alone leave the wealth-constrained firms with inefficiently low incentives to exert effort, and the "wealthy" firms with excessively high incentives. Because effort is substitutable, the wealth-constrained firm, which is aware of this incentive distortion, excessively free-rides on the efforts of the wealthy firm. Insurance coverage can mitigate this problem by transferring the failure costs (i.e. the financial liability) away from the supply chain. Specifically, insurance allows the "wealthy" firm to credibly commit not to increase effort and this, in turn, decreases the incentives of the wealth-constrained firm to free-ride.

The stylized model presented in this essay ignores some important operational features. For example, in our model the firms are able to decrease the likelihood of a failure, but not the magnitude of the failure costs (in the event of a failure). In the literature, the first type of effort is known as preventive effort, while the latter is known as contingency effort. If we relax this assumption, our main results would be affected if the contingency efforts are effective enough to decrease the failure costs to a point where the wealth constraints become non-binding. However, this is highly unlikely in many contexts, e.g. a nuclear meltdown, an oil spill or a product recall. In these contexts, an operational failure often causes financial distress or even bankruptcy. The effectiveness of contingency efforts may, therefore, be very limited.

In this essay we demonstrate that business insurance may allow the supply chain to operate more efficiently. Our results are particularly relevant to situations where large firms contract with considerably small suppliers. In these scenarios, a contractor would purchase business insurance, even if the contractor is wealth-unconstrained and risk-neutral. Through insurance, the contractors can prevent the wealth-constrained suppliers from excessively free-riding on its reliability efforts. A similar situation arises when small contractors hire large suppliers. This implies that the availability of insurance has non-trivial implications for supply chain contracting. To our knowledge, this role of insurance has not been previously highlighted in the literature. These results contribute to bridging the supply chain contracting and risk management literatures.

Chapter 4

Policy Incentives to Mitigate the Impact of Operational Tort Liability

4.1 Introduction

In industries characterized by the potential for injury or harm to third parties due to operational accidents, the costs of tort liability can be significant.[35] As a result of these costs, some firms may leave the market, and others may be discouraged from entering. Consider the following examples. In the 1980's, the Diphtheria, Pertussis and Tetanus (DPT) vaccine, given to children, allegedly caused some cases of severe neurological damage. After a series of multi-million dollar lawsuits against the DTP vaccine producers, the insurance premium for vaccine liability rose dramatically. As a result, all but one of the vaccine manufacturers exited the market, and the vaccine price went up by 6000 percent (Manning, 1994; Danzon and Sousa Pereira, 2011). Similarly, over the last decade, major nuclear energy suppliers, including Westinghouse and General Electric, were reluctant to enter the Indian energy market due to the costs associated with a potential nuclear accident (Bloomberg 2011). More recently, in 2013, a freight train from the Montreal, Maine & Atlantic (MM&A) Railway suffered a derailment in Lac-Mégantic, Quebec, resulting in the death of 47 people due to the explosion of 74-freight cars containing crude oil. After the crash, the railway company was unable to bear the liability and clean-up costs, and filed for bankruptcy. This led to concerns in several cities in Maine, where the MM&A was the sole railway company (Portland Press Herald, 2013).

In situations like these, government intervention has often been necessary

[35] In this essay, we use Shavell's (2009) definition of an accident as a "harmful outcome that neither injurers nor victims wished to occur, although the injurer... might have affected the likelihood of the outcome." We also use Shavell's definition of tort liability as a "legal obligation of a party who causes harm to make a payment to the victim of the harm."

to encourage market entry or to deter exit of firms. But there is widespread disagreement as to how firms should be incentivized. First, governments can provide incentives through *ex ante* subsidies. These subsidies influence decisions that firms take prior to commencing dangerous operations. For example, the U.S. Nuclear Decommissioning Trust provides funds to support the safe decommissioning of reactors, and the Nuclear Waste Program Act of 1982 subsidizes the disposal of radioactive waste. *Ex post* subsidies help mitigate the financial damages caused by the accident, by offering funds to share the costs of a clean-up, or by limiting the firm's exposure to liability. For example, both the U.S. Price-Anderson Indemnity Act of 1957 and the India Nuclear Liability Act of 2010 were created to protect suppliers against nuclear accident liability. These acts impose a maximum cap on any accident. Similarly, following the vaccination crisis of the 1980's, the Vaccine Injury Compensation Program created a no-fault liability system, which protects producers of pediatric vaccines from liability costs. In the oil and gas industry, the Oil Spill Liability Trust manages a $1.6 billion fund to help companies pay for the clean-up costs of oil spills.

In much of the law and economics literature, ex ante incentives are considered superior to ex post incentives because of the moral hazard concerns associated with the latter. This logic follows the idea that the firm must internalize the full impact of any potential accidents to take the efficient level of care to prevent accidents. By this view, offering ex post incentives will effectively subsidize dangerous operations leading to an increased probability of an accident. Because ex ante subsidies are an up-front payment to the firm, these do not diminish the financial impact of an accident for the firm and therefore do not induce moral hazard. Hence, ex ante subsidies are seen as the best way to induce sufficient market entry.

However, as noted in the examples given earlier, both ex ante and ex post subsidies are observed in practice. In this essay we demonstrate that when there is information asymmetry about the firms' ability to prevent accidents, the provision of ex post subsidies may be efficient. Some firms may have an inherent advantage in improving the reliability of their operations, and this capability is either unobservable to a social planner (or prohibitively costly to observe). This issue is central in the safety and regulation literature. For example, Antle (1996) has an extensive discussion on how uncertainty about the ability of food manufacturers (to control their safety of their products) has led to market failure and inefficient regulation. This is because, when a social planner does not have perfect information about the firms' ability to reduce the likelihood of an accident, ex ante subsidies to invest in reliability will go to all firms regardless of their ability to curtail operational accidents.

This is socially inefficient because the entry decision of high-ability firms, who receive these subsidies, may be unaffected by these incentives.

This essay seeks to characterize the conditions under which it is socially optimal to use ex ante subsidies or ex post liability protection (or both). To answer this question, we model a market for a homogeneous good. This good can be supplied by two risk-neutral firms. The market is characterized by asymmetric information: both hidden information and hidden action. A firm can either be a *high*-ability or *low*-ability type and this information is private. A high-ability firm can exert reliability improvements more efficiently (or, alternatively, at a lower cost) than a low-ability firm. In addition, these investments are unobservable to the social planner and, as a result, they are subject to moral hazard.

The social planner has to determine a policy to maximize social welfare. He does so by choosing the level of ex ante subsidies and ex post liability protection. The firms, depending on their type, choose whether to operate in (or exit) the market. The firms that stay in the market receive ex ante subsidies and invest in reliability. After operations begin, the firm(s) earn profits and consumers gain utility from the good. At this stage, an accident may occur, and if it does, the firm will be liable for its share of the damage; this share is determined by the terms of the ex post incentives that had been specified at the outset.

We demonstrate that in some cases it will be optimal to induce market entry only by high-ability firms, and in other cases it is efficient to induce entry by all firms, independent of their ability to prevent operational accidents. This depends on three factors: (a) the level of market competition; (b) the potential accident costs and; (c) the opportunity cost of public funds. Specifically, if the welfare gains of increased competition are low, the accident damages are high, or the opportunity cost of public funds is high, then it is socially efficient to induce *only* high-ability firms to stay in the industry. In the converse scenarios it is optimal to induce both high- and low-ability firms to stay in the industry.

Our results show that when it is socially optimal to induce only high-ability firms to stay in the market, the optimal policy will offer ex ante subsidies but no ex post liability protection. This is because, by offering no ex post liability protection and optimally choosing ex ante subsidies, the social planner can ensure that low-ability firms will not find it optimal to enter the market. If, on the other hand, it is socially optimal to induce both high- and low-ability firms to stay in the market, the optimal policy will offer a combination of ex ante and ex post incentives. Some level of ex post liability protection is efficient to provide incentives for low-ability firms to enter the market. This is because relying on ex ante subsidies alone would mean that higher ex ante subsidies would have to be offered to induce entry by low-ability firms. However, this provides excessive subsidies to high-ability firms. The social planner may therefore find it optimal to offer some level of ex post subsidies instead of higher ex ante subsidies, as a way to provide an incentive for lower ability firms to enter the market.

One key conclusion of the above analysis is the following. In some cases, the value of having multiple firms in the market is relatively large, or the costs of an operational accident are relatively low. In these scenarios, the social planner is willing to trade reliability for the benefits of increased market competition. Here, the social planner will induce low-ability firms to stay in the market and, as a result, he will offer both ex ante and ex post subsidies. In the converse scenario (i.e., if the costs of an accident are relatively high, or the benefits of market competition are low), then the social planner will value reliability more than market entry. In this case, the social planner will only be willing to induce market entry by firms with high-ability and, as a result, he will only use ex ante subsidies.

This essay contributes to a long-standing debate about government intervention in industries that engage in "dangerous" operations. Coglianese (2010) argues that "ex post liability, while useful, does not always by itself provide socially optimal level of risk control. As such, preventative risk regulation will be needed, and the core questions remain about how stringent should such regulation be, and what form such regulation take. Risk regulation research will continue to be needed to provide conceptual clarity to the normative basis for risk standards". In this essay, we explore how government regulation, social welfare and operational safety are intertwined. This essay is of interest for policy makers, as it demonstrates the optimal combination of ex ante and ex post mechanisms when entry and exit decisions are relevant. We also show that the optimal intervention depends on the level of industry competition, the extent of the accident damages, or the opportunity cost of public funds. For the operations management community, this essay is a first step to understanding how policy tools affect, not only the incentives to exert operational safety but also the structure of the industry, i.e. the number and type of competitors that firms will face. More broadly, this essay bridges the literatures on operational risk management and tort law.

4.2 Literature Review

Calabresi (1970) is considered a seminal contribution to the economics and tort law literature. He describes three types of accident costs: primary costs (direct accident losses), secondary costs (social costs of spreading - or concentrating - the accident liability)[36] and tertiary costs (judicial costs). In our essay, we analyze optimal policies in the presence of both primary and secondary costs. Specifically, we study whether it is optimal to spread accident losses, through ex post liability protection, as a way to reduce the social costs of industry exit (even if this increases the expected primary costs).

Calabresi's work is the preamble to a large debate on liability regulation. For example, Viscusi and Moore (1993) argue that liability protection encourages beneficial innovation in R&D intensive industries and, at the same time, Burk and Boczar (1993) show that liability protection is beneficial in the biotechnology industry, considering that manufacturers are exposed to other types of financial risks. However, Danzon and Sousa Pereira (2011) and Manning (1994) argue that liability protection has been either ineffective, or even harmful, in the vaccine industry. Similarly, Trebilcock and Winter (1997) argue that liability protection has detrimental effects on safety incentives and, unless regulation is a perfect substitute, it should not be considered in the nuclear industry.

In this essay, we focus on a context that combines asymmetric information and market exit. We show that the joint use of ex post liability protection and ex ante subsidies may be optimal, even in the presence of moral hazard. In this sense, our paper is related to Kolstad et al. (1990) and Schmitz (2000), who look at the interaction between ex post liability and ex ante regulation. Like these papers, we show that ex ante and ex post instruments are not always perfect substitutes. However, unlike these papers, we develop our model in a multiple-firm setting and, as a result, we focus on industry wide equilibria and social welfare. Moreover, the papers above focus on primary and tertiary costs, while our paper focuses on the interaction between primary and secondary accident costs.

[36] Secondary costs refer to the externalities that are generated when the accident liability is assigned to a specific party (or distributed across a group of parties). For example, suppose that courts rule in favour of making firms liable for all workplace accidents. As a result of this ruling, firms may take two types of actions: (1) they may take precautionary measures (e.g. buy safety equipment their employees), or (2) they may decrease hiring rates for blue-collar jobs. Both actions will decrease primary accident costs. However, the second type of action may also increase unemployment in the economy, and should be accounted for as a secondary cost.

Our paper also contributes to a small sub-field in the operations management (OM) literature, one that looks at the interaction between policy incentives and operations. For example, Bakshi and Gans (2010) model the provision of government incentives in homeland security, by studying the Customs-Trade Partnership Against Terrorism (C-TPAT). Arora et al. (2008) analyze public policy in e-operations, in the face of potential vulnerabilities to cyber security. Plambeck and Wang (2009) investigate the impact of e-waste regulation on new product introduction, and analyze the optimality "fee-upon-sale" and "fee-upon-disposal" policies. Kraft et al. (2013) look at both NGOs and government intervention, in industries where firms manage hazardous materials. Anand and Giraud-Carrier (2013) analyze and compare three popular mechanisms to regulate pollution, by looking at their impact on social welfare. Cheung and Zhuang (2012) look at a similar context to the one analyzed by our model. In their paper, the authors look at the impact of competition on optimal regulatory policies, in industries managing dangerous operations. The authors find that the optimal level of regulation is stricter under high degrees of competition. Our focus, in this essay, is different in two ways. First, we focus on the interaction between ex ante and ex post incentives, while Cheung and Zhang focus on government enforcement (i.e governments ensuring that firms comply with regulatory standards). Second, unlike this essay, the authors analyze optimal policies under *given* (i.e. exogenous) market structures. We focus on the way policy incentives affect not only operational decisions, i.e. investments in reliability, but also the structure of the market.

4.3 Model

There is a market for a homogeneous good, which can be supplied by two risk-neutral firms ($i = 1, 2$). The firms face a decision on whether to operate in the market or to stay out. If a firm decides to operate, it faces the possibility that it may cause an operational accident (e.g. a nuclear disaster). If an accident occurs, the firm is liable for the damages caused. To prevent this occurrence, the firm can invest in reliability. For example, it can conduct preventive maintenance, purchase safety equipment or sample production batches for quality control. In reality, some firms are "better" at exerting reliability improvements. For example, established firms may have a learning-by-doing advantage over newer entrants. We assume that the firms fall into one of two categories: those with *high-ability* and those with *low-ability* at making reliability improvements.

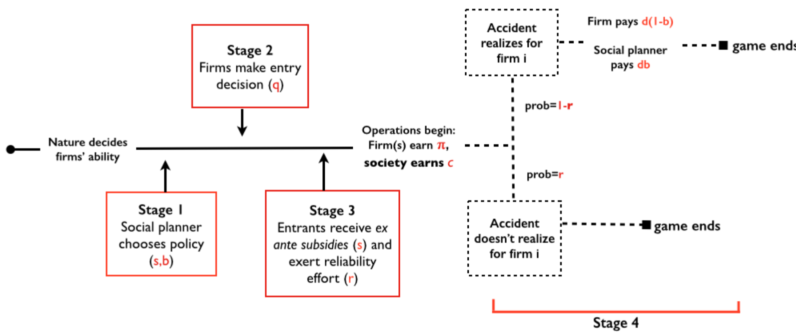


Figure 4.1: Model timeline

Firms make reliability investments up to a point where the marginal expected costs of an accident and the marginal investment costs are equal. In some cases, however, the ex-post costs of an accident are high which would, in turn, require high reliability investments. These costs may discourage some firms (especially those with low-ability) from operating in the market.

If a firm chooses not to operate, social welfare decreases because there is less competition in the market. The social planner can correct this market failure by offering incentives, which encourage firms to stay in the industry. We consider two types of incentives: *ex ante subsidies* (to defray the costs of investing in reliability) and *ex post liability protection* (to defray accident costs). The social planner's problem is to determine the policy that maximizes social welfare, where social welfare is equal to producer surplus plus consumer surplus minus the cost of public funds. Note that the government doesn't know the exact ability type of the firms, but knows that the proportion of high-ability firms is p.

Figure 4.1 summarizes the timeline of the model. This timeline can be divided into four stages, which we explain in detail below.

Stage 1: The Social Planner's Policy

First, the social planner chooses the optimal policy, by offering either ex ante subsidies or ex post liability protection (or both). This policy is denoted by the vector (s, b), where $s \geq 0$ represents the level of ex ante subsidies and $b \in [0, 1]$ represents the level of ex post liability protection, i.e. the proportion of ex post costs paid by the social planner.[37]

We let $\lambda \in (0, 1)$ denote the marginal cost of public funds. In the public policy literature, this term is also known as the shadow price of social funds (e.g., Laffont and Tirole, 1996; Dahlby, 2008). This term captures the fact that public funds allocated to this industry have alternative uses, i.e. there is an opportunity cost associated with providing incentives using public funds (Jones, 2005).

Stage 2: Entry Decision and Market Welfare

After observing the social planner's policy, the firms determine their optimal entry strategy. The firms can operate, stay out or play a mixed strategy (i.e. operate with some probability). We let $q_i \in [0, 1]$ denote the entry probability for firm i. We assume that there are no fixed costs for taking either decision. For this reason, our results are invariant to whether the firms are potential entrants or incumbents.

The firms' decisions will map into a market structure: duopoly (if both firms operate), monopoly (if only one firm operates) and complete market failure (if both firms stay out). Let $X \in \{D, M, 0\}$ denote the market structure, where D represents a duopoly, M represents a monopoly and 0 represents complete market failure. In other words, we have a mapping from the firms' entry decisions to the market structure, $(q_1, q_2) \rightarrow X$, such that

$$X(q_1, q_2) = \begin{cases} D & \text{with probability } q_1 q_2 \\ M & \text{with probability } q_1(1 - q_2) + q_2(1 - q_1) \\ 0 & \text{with probability } (1 - q_1)(1 - q_2) \end{cases} \tag{4.1}$$

If only one of the firms operates in the market, it will receive monopoly revenue $\rho_M > 0$ and if both operate, each will receive duopoly revenue $\rho_D \in \left[0, \frac{1}{2}\rho_M\right]$. In the monopoly case, consumer surplus is denoted by $c_M > 0$ and, in the duopoly case, consumer surplus is $c_D > c_M$. If both firms exit the market, then the firms and the consumers will receive 0 profits and surplus, i.e. $\rho_0 = 0$ and $c_0 = 0$. As a result, market surplus will equal $\rho_M + c_M$ in a monopoly, $2\rho_D + c_D$ in a duopoly and 0 if the market fails completely.

The values for the above parameters are determined by a downward sloping demand function $\Delta(P)$. For example, if we consider the demand function $\Delta(P) = 1 - P$, we can verify that the monopoly outcome yields $\rho_M = \frac{1}{4}$ and $c_M = \frac{1}{8}$. Our model does not explicitly assume that (in a duopoly) the firms

[37] We could, alternatively, model *ex post liability protection* by considering a liability cap. Both specifications are mathematically equivalent in our model.

compete in a specific fashion, e.g. a Cournot or Bertrand competition. However, to understand the implications of market competition on the optimal policy, we introduce a parameter $\kappa \in [0,1]$. Here κ describes the degree of competition among the duopolists. It ranges from perfect collusion ($\kappa = 0$) to Bertrand (perfect) competition ($\kappa = 1$). If $\kappa = 0$, then $\rho_D = \frac{1}{2}\rho_M$ and $c_D = c_M$ and, if $\kappa = 1$, then $\rho_D = 0$ and $c_D > c_M$. Moreover, we assume that for any $\kappa \in [0,1]$, $c'_D(\kappa) > 0$, $\rho'_D(\kappa) < 0$. See figure 4.2 for an illustration.

We assume any increase in competition increases market surplus, i.e. decreases deadweight loss. Observe that $2\rho'_D(\kappa) + c'_D(\kappa)$ measures the increase in surplus (or the decrease in deadweight loss) as κ increases. Therefore, we assume that $2\rho'_D(\kappa) + c'_D(\kappa) \geq 0$ or equivalently, $\frac{c'_D(\kappa)}{2|\rho'_D(\kappa)|} > 1$. Moreover, $c_D(1)$ captures the surplus of the good when there is no deadweight loss.

To guarantee that our model is well-behaved, we assume that the monopoly surplus $(c_M + \rho_M)$ yields at least half of the maximum market surplus which, in turn, guarantees us that $2c_M + 2\rho_M \geq c_D + 2\rho_D$, or equivalently, that $2c_M - c_D + 2(\rho_M - \rho_D) \geq 0$. Note that this assumption is not very restrictive, since it is satisfied by "well-behaved" demand functions (e.g. functions that are continuous, differentiable and downward sloping).[38]

Stage 3: Reliability Investments

The firms that enter the market will receive ex ante subsidies s, and make reliability investments. For example, the operator of a nuclear energy facility can improve the material quality of its containment areas, or increase the levels of monitoring, as a way to ensure that there are no radioactive leakages. These efforts are only observed by the firm and denoted by $r \in [0,1]$.

In addition to hidden action, we study a setting characterized by hidden information. That is, we consider the case where some firms have greater *ability* to improve the reliability of their operations. For simplicity, we only consider two types of firms, those with high-ability (H) and those with low-ability (L). When a firm has low-ability, the cost of effort is given by the function $\frac{r^2}{2\alpha_H}$ and if the firm has high-ability, the effort function is $\frac{r^2}{2\alpha_L}$, for some $\alpha_H \geq \alpha_L > 0$. The parameters α_H and α_L measure the degree of asymmetry between a high- and a low-ability firm.

[38]For example, note that for a given demand function of the form $D(P) = (A - BP)^m$, given $A, B > 0$ and $m \geq 0$, the profit maximizing price for a monopolist will be equal to $P = \frac{A}{B(m+1)}$. Therefore, the ratio of monopoly surplus to total surplus will be equal to $\frac{m - m^{m+1}}{(m+1)^{m+1}} + 1 > \frac{1}{2}$.

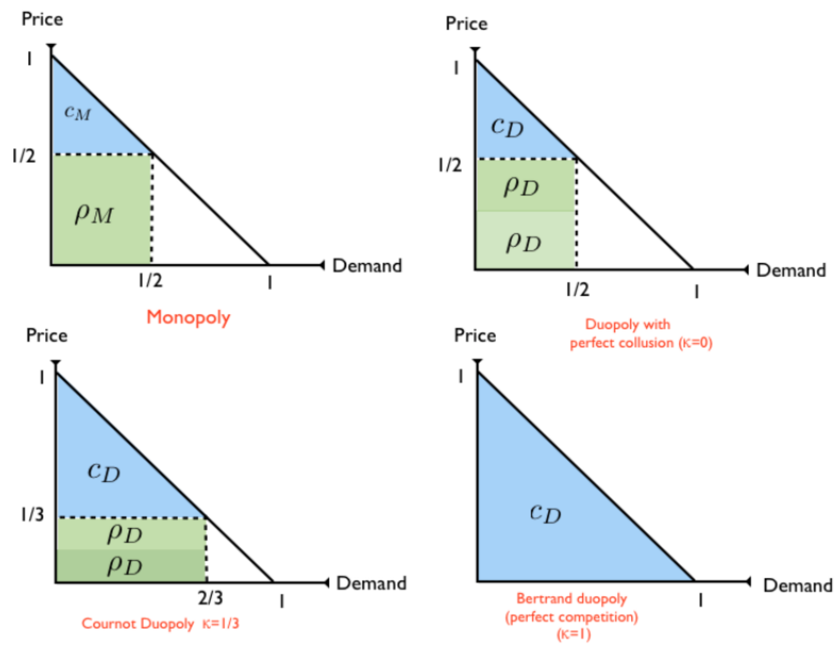


Figure 4.2: An illustrative example using an inverse demand function, $\Delta(Q) = 1 - Q$.

Let τ_i denote the ability type of firm i. Each firm's type is private information, but the social planner and firm i have a prior about firm j's type. Specifically, their prior that the probability that firm j has high ability is equal to $p \in [0, 1]$, and this prior is accurate.

Stage 4: Operations and Potential Accidents

After reliability investments are made, operations will begin and accidents may occur. The probability that an accident occurs is equal to $1 - r$. The damages caused are assumed to be equal to $d \in (0, \infty)$, where d encompasses a variety of costs that range from clean-up to litigation expenses. We assume that d is deterministic. In section 4.6, we discuss the case where d is a random variable. In accordance with the policy offered by the planner, the firms will pay $d(1 - b)$ and the social planner will pay db.

To avoid analyzing trivial scenarios, we assume that the accident damages (d) are high enough so that $2\rho_D + c_D < 2d$ and $\rho_M + c_M < d$. This allows us to avoid cases where it is optimal to have firms in the market, regardless of their reliability investments. In other words, we assume that it is optimal not to have a firm in the market, as opposed to entering the market and causing an operational accident with certainty.

4.4 Analysis

Analyzing our model through backward induction, we first describe the optimal level of reliability effort (in §4.4.1). Next, we determine the optimal entry decision (§4.4.2) and, finally, we characterize the socially optimal policy (in §4.4.3).

4.4.1 Stage 3: Reliability Investment Decision

In this stage, each firm will be either running operations, or be out of the market. Let $\Pi_i^{\tau_i}(r_i|X,s,b)$ denote the stage 3 profits for firm i (whose ability type is equal to τ_i). We have that

$$\begin{aligned}\Pi_i^{\tau_i}(r_i|X,s,b) \quad &= \quad \underbrace{\rho_X}_{\text{Revenue}} \quad - \quad \underbrace{(1-b)\,d\,(1-r_i)}_{\substack{\text{Expected}\\ \text{liability}}} \quad - \quad \underbrace{\frac{r_i^2}{2\alpha_{\tau_i}}}_{\substack{\text{Costs of reliability}\\ \text{investments}}} \\ &\quad + \quad \underbrace{s}_{\substack{\text{Ex ante}\\ \text{subsidies}}}\end{aligned} \tag{4.2}$$

If the firm is out of the market, then $\Pi_i = 0$. In stage 3, the firm's problem is to find

$$\max_{r_i \in [0,1]} \Pi_i^{\tau_i}(r_i|X,s,b)$$

Note that if firm i has ability τ_i, it will exert reliability investments

$$\hat{r}_i^{\tau_i}(b) = \begin{cases}(1-b)\,d\alpha_{\tau_i} & \text{if } \alpha_{\tau_i} \leq \frac{1}{d(1-b)}\\ 1 & \text{if } \alpha_{\tau_i} \geq \frac{1}{d(1-b)}\end{cases}$$

Moreover, when the firms are not offered liability protection, i.e. when $b = 0$, note that the firms fully internalize the accident costs and they will exert the level of reliability effort that minimizes primary accident costs. However, when ex post liability protection is offered, i.e when $b > 0$, then the firms decrease their reliability investments.

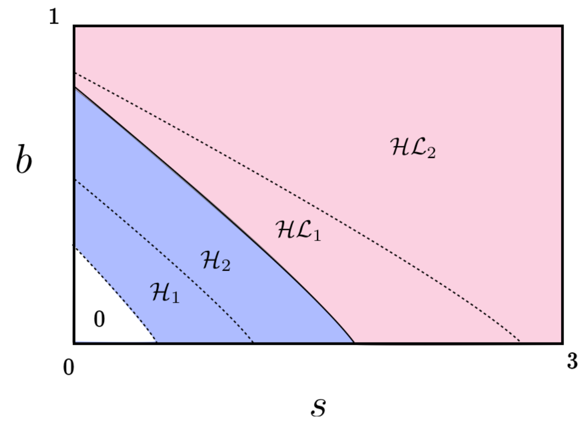


Figure 4.3: Illustration of the model regions.

4.4.2 Stage 2: Optimal Entry Decision

In stage 2, the firms play a simultaneous game where they make a decision regarding q_1 and q_2. This decision is based on their expected profit level, which depends on the realized market structure. The stage 2 profits for the firms are given by

$$\tilde{\Pi}_i^{\tau_i}\left(q_i, q_j|s, b\right) \equiv q_i\left[q_j \Pi_i^{\tau_i}\left(\hat{r}_i^{\tau_i}\left(b\right)|D, s, b\right) + \left(1 - q_j\right)\Pi_i^{\tau_i}\left(\hat{r}_i^{\tau_i}\left(b\right)|M, s, b\right)\right] \tag{4.3}$$

Firm i chooses an entry strategy q_i that maximizes its expected payoff, given the entry strategy q_j of the rival firm. Observe that because the payoff to each firm is contingent on its own type τ_i, it is also contingent on the type of the rival firm (τ_j), which is unknown to i. Recall, however, that firm i has a prior p=$Pr\left(\tau_j = H\right)$.

Let $B_i^{\tau_i}$ be the best response function for i. We have,

$$B_i^{\tau_i}\left(q_j^H, q_i^L|s, b, p\right) \quad = \quad \arg\max_{q_i \in [0,1]} \Pi_i^*(q_i^{\tau_i}, q_j^H, q_j^L|s, b, p) \tag{4.4}$$

where $\Pi_i^*(q_i^{\tau_i}, q_j^H, q_j^L|s, b, p) \equiv p\tilde{\Pi}_i^{\tau_i}\left(q_i, q_j^H|s, b\right) + (1-p)\,\tilde{\Pi}_i^{\tau_i}\left(q_i, q_j^L|s, b\right)$. In this game, a Bayesian Nash Equilibrium (BNE) is defined as a vector of entry strategies that is a fixed point of the best response correspondences. Lemma 11 presents this equilibrium.

Lemma 11. *Let firm 1 and firm 2 have ability type τ_1 and τ_2, respectively, and let $m_\tau \equiv d\left(1-b\right)\left(1-\hat{r}^\tau\right) + \frac{\hat{r}_i^2}{2\alpha_\tau}$ represent the expected firm liability* *plus reliability investments. The Bayesian Nash Equilibrium for the entry strategies of firm 1 and 2 is* $\left(q_1^{BNE}, q_2^{BNE}\right) = \left(\hat{q}^{\tau_1}, \hat{q}^{\tau_2}\right)$ *where,*

- $\hat{q}^H = \begin{cases} 1 & \text{if } s+\rho_M-m_H > p\left(\rho_M-\rho_D\right) \\ \frac{s+\rho_M-m_H}{p(\rho_M-\rho_D)} & \text{if } s+\rho_M-m_H \in \left[0, p\left(\rho_M-\rho_D\right)\right] \\ 0 & \text{if } s+\rho_M-m_H < 0 \end{cases}$

- $\hat{q}^L = \begin{cases} 1 & \text{if } \frac{s+p\rho_D+(1-p)\rho_M-m_L}{1-p} > \rho_M-\rho_D \\ \frac{s+p\rho_D+(1-p)\rho_M-m_L}{(1-p)(\rho_M-\rho_D)} & \text{if } \frac{s+p\rho_D+(1-p)\rho_M-m_L}{1-p} \in \left[0, \rho_M-\rho_D\right] \\ 0 & \text{otherwise} \end{cases}$

Lemma 11 presents three regions of parameter values; these regions are illustrated in Figure 4.3. In this example, we have that $\rho_M = \frac{10}{4}$ and $c_M = \frac{10}{8}$, $\rho_D = \frac{10}{9}$, $c_D = \frac{20}{9}$, $d = 6$,$\alpha_H = \frac{1}{200}$, $\alpha_L = \frac{1}{210}$ and $\lambda = 0.1$. In the figure above, region 0 represents a scenario in which both firms exit the market, i.e.$q_H = 0$ and $q_L = 0$. In regions $\mathcal{H}_1$ and $\mathcal{H}_2$, high-ability firms will choose a positive entry probability, while the low-ability firms will choose to exit. Specifically, in region $\mathcal{H}_1$ the high ability firms set $q_H = \frac{1}{p}\frac{\rho_M+s-m_H}{\rho_M-\rho_D}$ while low-ability firms set $q_L = 0$ and, in region $\mathcal{H}_2$, the high-ability firms set $q_H = 1$ and low-ability firms set $q_L = 0$. Finally, regions $\mathcal{HL}_1$ and $\mathcal{HL}_2$ represent those scenarios where both high- and low-ability firms choose a positive value for q_i. Specifically, region $\mathcal{HL}_1$ represents those scenarios in which high-ability firms set $q_H = 1$ and low-ability firms set $q_L{=}\frac{1}{1-p}\frac{s-m_L+p\rho_D+\rho_M(1-p)}{(\rho_M-\rho_D)}$ and, region $\mathcal{HL}_2$ represents those scenarios where the high-ability firms set $q_H = 1$ and the low-ability firms set $q_L = 1$. Finally, note that the probability that high- and low-ability firms enter the market is a decreasing function of the prior p.

4.4.3 Stage 1: The Optimal Policy

In stage 1, the social planner seeks to maximize social welfare ($\mathcal{W}$), which is defined as firm profits plus consumer surplus minus the opportunity cost of public funds. To characterize social welfare, let us first look at stage 3. Suppose that, at the beginning of this stage, a monopoly structure has been realized (i.e. $X = M$) and that the monopolist (firm i) has ability type τ_i. In this case, the stage 3 social welfare will be equal to

$$\begin{aligned} \mathcal{W}^{\tau_i}\left(M,s,b\right) &= \underbrace{\Pi_i^{\tau_i}\left(\hat{r}_i^{\tau_i}\left(b\right),M,s,b\right)}_{\text{Firm profits}} + \underbrace{c_M}_{\text{Consumer welfare}} \\ &- \underbrace{\left(1+\lambda\right)\left[db\left(1-\hat{r}_i^{\tau_i}\left(b\right)\right)+s\right]}_{\text{Cost of public funds}} \end{aligned} \tag{4.5}$$

Similarly, suppose that at the beginning of stage 3, a duopoly structure has been realized, and that the duopolists have ability τ_1 and τ_2. Noting that $\mathcal{W}^{H,L}(D,s,b) = \mathcal{W}^{L,H}(D,s,b)$, we have

$$\begin{array}{rcll} \mathcal{W}^{\tau_1,\tau_2}(D,s,b) & = & \sum_{i=1}^{2} \Pi_i^{\tau_i}\left(\hat{r}_i^{\tau_i}(b), D, s, b\right) & + \quad c_D \\ & - & (1+\lambda)\left[db\sum_{i=1}^{2}\left(1-\hat{r}_i^{\tau_i}(b)\right)+2s\right] & \end{array} \tag{4.6}$$

Finally, if the market fails completely, we have that $\mathcal{W}(0,s,b) = 0$.

As a result, we can write the expected social welfare (in stage 1) as

$$\begin{aligned} E\left[\mathcal{W}(s,b)\,|p\right] = \; & 2\left(p\left(1-\hat{q}^H\right)+(1-p)\left(1-\hat{q}^L\right)\right)\left(p\hat{q}^H\mathcal{W}^H(M,s,b)\right) \\ & 2\left(p\left(1-\hat{q}^H\right)+(1-p)\left(1-\hat{q}^L\right)\right)(1-p)\hat{q}^L\mathcal{W}^L(M,s,b) \\ & +p^2\left(\hat{q}^H\right)^2\mathcal{W}^{H,H}(D,s,b)+2p(1-p)\,\hat{q}^H\hat{q}^L\mathcal{W}^{H,L}(D,s,b) \\ & +(1-p)^2\left(\hat{q}^L\right)^2\mathcal{W}^{L,L}(D,s,b) \end{aligned}$$

The social planner's problem is to find

$$\max_{(s,b)} E\left[\mathcal{W}(s,b)\,|p\right]$$

To characterize the optimal policy, we first demonstrate the existence of three different regions. Consider Lemma 12 below (which is illustrated in Figure 4.4).

Lemma 12. *Let $(s^*,b^*) = \arg\max_{(s,b)} E\left[\mathcal{W}(s,b)\,|p\right]$. There exists $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$ such that:*

1. ***(Region 0)*** *If $\alpha_H \leq \alpha_1$, then (s^*,b^*) will induce both high-ability and low-ability firms to exit the market, i.e. to choose $q_i = 0$.*

2. ***(Region $\mathcal{H}$)*** *If $\alpha_H > \alpha_1$ but $\alpha_L \leq \alpha_2$, then (s^*,b^*) will induce low-ability firms to exit the market, i.e. to choose $q_i = 0$. Conversely, (s^*,b^*) will induce high-ability firms to enter with positive probability, i.e. to choose $q_i > 0$.*

3. ***(Region $\mathcal{HL}$)*** *If $\alpha_L > \alpha_2$, then (s^*,b^*) will induce high-ability firms to enter, i.e. to choose $q_i > 0$, and low-ability firms to enter with positive probability i.e. to choose $q_i > 0$.*

Lemma 12 tells us that, whenever $\alpha_H \leq \alpha_1$, then the optimal policy will lie in region 0 (of figure 4.3). Conversely, if $\alpha_H > \alpha_1$ but $\alpha_L \leq \alpha_2$, then the optimal policy will lie in either region $\mathcal{H}_1$ or $\mathcal{H}_2$. However, if $\alpha_L \geq \alpha_2$, then the optimal policy will lie in either region $\mathcal{HL}_1$ or $\mathcal{HL}_2$. The intuition behind

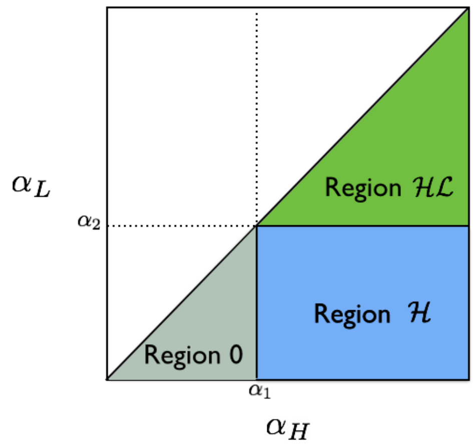


Figure 4.4: A visualization of Lemma 12

this result is straightforward. Suppose that α_H and α_L are very large. In this case, both firms are very inefficient at exerting reliability improvements and, all else equal, they will choose low values for r_i. If this happens, it is better to "shut down" the market than to incentivize unreliable firms to operate. The second region in Lemma 12 presents a scenario where the high ability firms are efficient, but the low-ability firms are inefficient. In this scenario, it is optimal to incentivize only high-ability firms to operate in the market. Finally, in region $\mathcal{HL}$, the lemma presents a scenario in which both types of firms are efficient. In this region, it is optimal to induce both firms of high- and low-ability to enter the market with positive probability.

Region 0 is trivial in the sense that the social planner will find it optimal to offer no incentives. For this reason, the rest of this essay focuses on those scenarios where $\alpha_H \geq \alpha_1$, i.e. on those scenarios where it is socially optimal to provide incentives so that, at least, the high-ability firms enter the market. In the lemma below, we characterize the optimal policy for these regions. Without loss of generality we present results for the case where, at equilibrium, $\hat{r}_i(s^*, b^*)$ is less than one. In other words, we focus on interior solutions.

Proposition 13. *Suppose $\alpha_H \geq \alpha_1$. The optimal policy, (s^*, b^*), can be characterized as follows:*

1. ***(Region $\mathcal{H}$)*** *If $\alpha_L \leq \alpha_2$, then $(s^*, b^*) = \left(s^{\mathcal{H}}, 0\right)$, where*

$$s^{\mathcal{H}} = d - \frac{\alpha_H d^2}{2} - \rho_M \tag{4.7}$$

$$+ (\rho_M - \rho_D)\, min \left\{ \frac{c_M + (1+\lambda)\left(\rho_M - d + \frac{\alpha_H d^2}{2}\right)}{2(1+\lambda)(\rho_M - \rho_D(\kappa)) + (2c_M - c_D(\kappa))}, p \right\} \tag{4.8}$$

2. ***(Region $\mathcal{HL}$)*** *If $\alpha_L \geq \alpha_2$, then $(s^*, b^*) = \left(s^{\mathcal{HL}}, b^{\mathcal{HL}}\right)$, where*

$$s^{\mathcal{HL}} = \left(1 - b^{\mathcal{HL}}\right) d - \frac{\alpha_L \left(\left(1 - b^{\mathcal{HL}}\right) d\right)^2}{2} - \rho_M \tag{4.9}$$

$$+ \frac{(\rho_M - \rho_D)\left[c_M + (1+\lambda)\left(\rho_M - d + \alpha_L \left(1 - b^{\mathcal{HL}}\right) d^2 \left(1 - \frac{1 - b^{\mathcal{HL}}}{2}\right)\right)\right]}{2(1+\lambda)(\rho_M - \rho_D(\kappa)) + (2c_M - c_D(\kappa))}$$

and $b^{\mathcal{HL}} \in (0,1)$, satisfying the first order condition

$$\left(1 + \frac{\lambda\left(1 - 2\left(1 - b^{\mathcal{HL}}\right)\right)}{b^{\mathcal{HL}}}\right)\left(\frac{p(\alpha_H - \alpha_L)}{\alpha_L(1+\lambda)}\right) \tag{4.10}$$

$$= -\frac{\frac{c_M}{1+\lambda} + \left(\rho_M - d + \alpha_L\left(1 - b^{\mathcal{HL}}\right) d^2 \left(1 - \frac{1 - b^{\mathcal{HL}}}{2}\right)\right)}{2(\rho_M - \rho_D(\kappa)) + \frac{(2c_M - c_D(\kappa))}{1+\lambda}} \tag{4.11}$$

The intuition behind Proposition 13 is the following. In region $\mathcal{H}$, the social planner wants to institute a policy so that only high-ability firms choose a positive entry probability. Through this policy, any entrant will reveal its type (as a high-ability firm) and, for this reason, there will be no hidden information about the entrant's ability. Furthermore ex ante subsidies do not distort a firm's incentives to invest in reliability, while ex post liability protection does. As a result, ex ante subsidies are superior to ex post liability protection.

Let us now turn to region $\mathcal{HL}$. In this region, it is optimal to provide incentives so that both high-ability and low-ability firms choose to enter with positive probability. If the policy were solely based on ex ante subsidies, it would provide *informational rents* to high-ability firms. That is, to encourage entry among low-ability firms, the high-ability firms will necessarily receive incentives in excess of their needs. In other words, the social planner could decrease the level of ex ante incentives, by some small quantity, without affecting the entry decision of the high-types, but affecting the entry strategy of low-ability firms. Therefore, it is optimal to substitute some level of ex ante subsidies for ex post liability protection. Specifically, the optimal policy will balance out the cost of informational rents (caused by ex ante subsidies) with the cost of hidden action (caused by ex post liability protection). In summary, the optimal policy balances out three components: the welfare gains from increased entry, the costs of hidden information, and the costs of moral hazard.

4.5 Comparative Statics

In this section, we analyze the factors that determine whether the optimal policy lies in region $\mathcal{H}$ or $\mathcal{HL}$. We consider three factors: (i) the degree of market competition κ, (ii) the magnitude of the accident damages d, and (iii) the marginal cost of public funds λ.

4.5.1 Industry Competition

The level of market competition is a key factor in the analysis of market entry and social welfare (see Cabral 2004). Specifically, the level of competition measures the social value of having multiple firms in the industry. When firms in the industry compete aggressively, the cost of firm exit is much higher to society than it is to the firm, and vice versa. This is because when there is aggressive competition, a firm that enters the market will significantly benefit consumers, while only earning modest economic rents.

To see how the level of competition affects the optimal policy, consider Lemma 14 below. Note that when duopoly competition (κ) increases, then the consumer surplus increases by $c'(\kappa)$, and firm profits decrease by $\rho'_D(\kappa)$.

Lemma 14. *If* $\frac{c'_D(\kappa)}{2|\rho'_D(\kappa)|} - 1 \leq \lambda$, *then* $\frac{\partial \hat{q}^L(s^*,b^*)}{\partial \kappa} \leq 0$. *If* $\frac{c'_D(\kappa)}{2|\rho'_D(\kappa)|} - 1 > \lambda$, *then* $\frac{\partial \hat{q}^L(s^*,b^*)}{\partial \kappa} \geq 0$.

Recall that by our assumption that increased market competition always increases market surplus (i.e. decreases deadweight loss), we get $\frac{c'_D(\kappa)}{2|\rho'_D(\kappa)|} - 1 \geq 0$. Lemma 14 tells us that when the marginal increase in market surplus (from increased competition) is less than the opportunity cost of public funds, then the social planner will be less inclined to induce firms of low ability into the market, and vice versa. Lemma 14 leads us to the following proposition.

Proposition 15. *When the surplus gains of increased competition are small, the social planner will only offer ex ante subsidies (but no ex post liability protection). Conversely, when the surplus gains of increased competition are large, the social planner will offer ex post liability protection.*

Proposition 15 yields the following intuition. Suppose that market competition is very aggressive. In this case, consumer surplus will be high in a

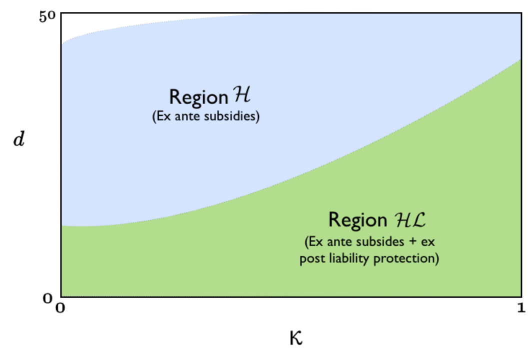


Figure 4.5: Illustrative example with inverse demand function $P(Q) = 1 - Q$.

duopoly. Conversely firm surplus will be low. In other words, consumers will greatly benefit from having two firms in the markets, but the firms will see their profits considerably reduced by the aggressiveness of competition. As a result, they will be less motivated to enter the market. On the other hand, the social planner will be more inclined to incentivize the firms to enter the market, to achieve a duopoly, and reduce the deadweight loss. As a result he will offer a more "generous" policy. The proposition above tells us that when the gains from increased competition are very large, then the generosity of this policy will be high enough so that low-ability firms are inclined to enter the market (i.e. he will offer a policy that lies in region $\mathcal{HL}$). And, as we saw in Lemma 13, when the policy is in region $\mathcal{HL}$, the social planner will offer ex post liability protection. If, conversely, the policy lies in region $\mathcal{H}$ (i.e. when the surplus gains from competition are low), the social planner will only offer ex ante subsidies.

An alternative interpretation to the proposition above is the following. When the duopolists compete in such a way that it increases welfare significantly, then the social planner will be willing to trade reliability (by inducing low-ability firms to operate) for the benefits of competition. As a result, the social planner will offer ex post liability protection. See Figure 4.5 for an illustration.

4.5.2 Accident Damages

In this subsection, we analyze the impact that the accident damages have on the optimal policy. We begin with the following proposition.

Proposition 16. *When the accident damages, d, are small, the optimal policy will offer a combination of ex ante subsidies and ex post liability protection (i.e. (s^*, b^*) will lie in region $\mathcal{HL}$). Conversely, when the accident damages are large, the optimal policy will only offer ex ante incentives (i.e. (s^*, b^*) will lie in region $\mathcal{H}$).*

The intuition behind Proposition 16 is the following. When a firm decides to enter the market, not only does it bring surplus, it also introduces a burden to society. In other words, the entry of a firm raises expected accident costs. When the social planner induces entry by low-ability firms, he is willing to trade reliability in exchange for market efficiency. However, when the accident damages are large, then the social planner is less willing to make this tradeoff. When this happens, the social planner will find it optimal to induce entry only among high-ability firms, by offering ex ante subsidies and no ex post liability protection. See Figure 4.5 for an illustration; note that the y axis represents the accident damages d.

4.5.3 Marginal Cost of Public Funds

The marginal cost of public funds λ refers to the opportunity cost of transferring funds to the firms. In this subsection, we analyze the impact that this parameter has on the optimal policy. Consider Proposition 17 below.

Proposition 17. *Let $\lambda \in [0, 1]$ be given. We have that:*

1. *When the marginal cost of public funds λ is small, the optimal policy will offer high levels of ex ante subsidies, but low levels of ex post liability protection.*

2. *When λ is intermediate, the optimal policy will offer intermediate levels of ex ante subsidies and high levels of ex post liability protection.*

3. *When λ is large, then the optimal policy will offer low levels of ex ante subsidies, and no ex post liability protection.*

If the marginal cost of public funds is low, the social planner will offer large amounts of ex ante subsidies, but low amounts of ex post liability protection. To understand why, let us consider an extreme scenario, where $\lambda = 0$. Here, there is no welfare loss attributable to transferring funds to the firms (i.e. a $1 loss from the public budget is offset by a $1 gain by the firms). In this case, the rent extraction problem that arises with ex ante subsidies disappears. In other words, the social planner will be indifferent

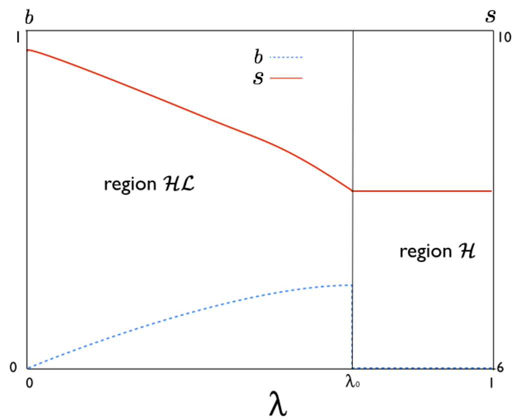


Figure 4.6: Optimal policy as a function of λ.

to the fact that high-ability firms extract rents from public funds. As λ increases to intermediate amounts, the rents extraction problem increases, and the social planner will find it optimal to decrease the level of ex ante subsidies, in exchange for some level of ex post liability protection.

When λ is large, the opportunity cost of public funds becomes significant. Here, the social planner will find it very costly to provide incentives to the firms and, as a result, it will become too costly to incentivize low-ability firms. When this happens, the optimal policy will shift to region $\mathcal{H}$, and the social planner only offer ex ante subsidies (but no ex post liability protection). See figure 4.6 for an illustration.

4.6 Extensions, Discussion and Conclusion

This essay bridges the operational risk management and the public policy literatures by studying the impact of potential operational accidents on market entry, and the role of government intervention in correcting any resulting market failure. It also contributes to a large debate in the public policy literature: are ex ante incentives always superior to ex post incentives, or vice versa? Our results show that when there is information asymmetry about the firms' ability to prevent accidents, the provision of ex post subsidies may be efficient. This is because a social planner cares not only about investments to make operations safer, but also about having the socially optimal number of firms enter the industry. Because firms only care about their private profits, and not social welfare, forcing firms to fully internalize the costs of accidents may lead to insufficient entry. By appropriately choosing ex ante and ex post subsidies the social planner can balance the impact of two externalities: the externality created by an accident and the externality created by insufficient entry.

The above results, however, are obtained by analyzing a highly stylized model. For example, our model does not consider multi-period interactions, demand-contingent or stochastic accident damages or the role of insurance markets. Below, we discuss the robustness of our model to the introduction of these elements.

Stochastic Accident Damages

Our model studies a setting where the accident damages are deterministic and known ex ante. However, our results are also robust to scenarios where the accident damages are stochastic. To understand why, note that our model focuses on high-impact and low-probability accidents, e.g. oil spills and nuclear meltdowns. The probability distribution of these accidents is typically characterized by a very long tail, which includes devastating accidents (i.e. black-swan events). In these scenarios, firms typically assess their decisions, not on the extent of the accidents, but rather on the Value at Risk (VaR), i.e. on the probability that an accident realization will be sufficiently high to jeopardize the firm. For example, suppose that a firm has a \$10 billion net worth. An oil spill that results in clean-up costs above this amount will jeopardize a firm's existence, whether the net cost of this spill is \$15 or \$20 billion. Under this argument, we could re-interpret our model by assuming that d measures the value that would place the firm at financial risk, and not the accident damages.

Demand-Contingent Accident Damages

Our model also assumes that the accident damages are independent of demand. For example, in the childhood vaccine industry the extent of the accident damages is highly (or even perfectly) correlated to the market demand. In this case accident damages will depend on industry output, which in turn depends on whether the industry is monopolized or is a duopoly. To model this setting we should assume that the magnitude of the accident damages is dependent on industry structure. Under this assumption, the monopoly outcome becomes "less desirable" to the firms, as more demand also implies higher potential accident costs. Conversely, the monopoly outcome may become more desirable to the social planner because less demand implies less accident costs. However, the fundamental tradeoff between market entry and operational reliability is still present, which leads us to conjecture that the insights will be unaltered by this extension.

Menu of Policies In our model, we consider a context where the social planner's intervention is characterized by a single policy. An alternative would be to take a mechanism design approach, which would yield a menu of policies. This approach, however, would have not affected the main insights of our model. First, if the social planner wishes to incentivize entry by high-ability firms only, then it is optimal to have a single policy targeted at the high-ability firms (which would not be attractive to low-ability firms). If, on the other hand, the social planner wishes to induce entry by firms with high- and low-ability, then he would have offered a menu containing two policies. The first policy, which would be incentive compatible with the high-ability firms, would offer ex ante subsidies, but not ex post incentives. This policy would provide positive rents to the high-ability firms. The second policy, which is incentive compatible with the low-ability firms, would offer some degree of ex post incentives, and a lower level of ex ante subsidies. This policy, however, would provide no rents for the low-ability firms. While this will be a more efficient intervention, implementing a menu of contracts is often infeasible in public policy. The key point, however, is that the trade-off between rent extraction and efficiency would also exist in the mechanism design approach.

Multi-Period Dynamics

The scope of our analysis is restricted to a single period setting. If we were to consider a dynamic model with entry and exit decisions in each period, then the optimal policy would need to address issues such as bankruptcy, which are absent in a single period model. That is, if the firms have limited wealth, the possibility of (costly) operational accidents does not only affect their entry decision, but also their ability to operate after the occurrence of a potential accident. For example, if the cost of an operational accident (d) is stochastic, and large realizations of d are possible, then the firms will be forced to exit the market after the occurrence of a such a realization. In this setting, the social planner will not only need to address issues related to ex ante entry (i.e. firms deciding whether to undertake operations), but also ex post entry and exit (i.e. firms deciding whether to continue in the market after causing an operational accident).

The introduction of a dynamic model also introduces the possibility of learning-by-doing effects. In the presence of these effects, the social planner may be more willing to tolerate the entrance of low-ability firms, especially during the initial periods. This implies that the social planner may be initially willing to offer higher levels of ex post liability protection (to induce entry by low-ability firms) and, as these firms converge towards a high ability level, the social planner may progressively switch to a policy that is purely characterized by ex ante subsidies.

Liability Insurance

We did not explicitly highlight the role of insurance markets. In practice, there are two possible roles that an insurer would play. First, the insurer could provide coverage for any "black-swan" events that impose costs above the accident damages d. This feature can easily be accommodated in our model, by simply adding an (ex ante) insurance premium to the firm's profit function. Under this interpretation, d would represent the insurance deductible.

A second role for insurance could be to provide liquidity to the firm to cover ex post accident costs. In our current model we assume that the firm self-insures, i.e. it has access to enough cash reserves to pay its share of ex post costs in the event of an accident. If there exists a perfectly competitive insurance market, and the two parties are able to write a complete contract, then the firm has access to sufficient liquidity in the event of an accident. In this case, we can interpret the firm not as a single entity, but rather a "firm-insurer" pair. This interpretation will fail, however, if there are any frictions in the insurance market: for example, if the insurance market is not competitive or if transaction costs are non-trivial. In these circumstances, the social planner would need to consider the role of the insurance market and will need to respond to any inefficiencies in this market.

Chapter 5

Conclusions

The first essay bridges the firm-level productivity and supply chain literatures. This essay provides new evidence about the link between Total Factor Productivity and supply chains. It does this by identifying the various mechanisms through which productivity can spillover across firms. Our results can be useful for practitioners at the time of managing their supply chain relationships. To arrive at this result, however, our essay first had to deal with several identification issues, some which are considered extremely challenging in the econometrics literature. To this end, we used a novel econometric approach that allowed us to overcome these challenges. This methodology could be adopted in future research, in order to identify the mechanisms through which different types of spillover effects propagate across vertical relationships.

The second essay (Chapter 3) expands our view on the roles of business insurance, by showing that business insurance may allow the supply chain to operate more efficiently. These results are particularly relevant to settings where large firms contract with small suppliers, and vice versa. The uneven distribution of wealth would lead to firms being more likely to buy insurance as a strategic tool. This role of insurance has not been previously highlighted in the literature. Our results thus imply that the availability of insurance has non-trivial implications for supply chain contracting (at least when these relationships are asymmetric). In general, our results imply that insurance coverage and contractual incentives are not necessarily substitutes, but may rather complement each other. These results contribute to bridging the supply chain contracting and risk management literatures.

The third essay (Chapter 4) contributes to a long-standing debate about government intervention in industries that engage in "dangerous" operations. Coglianese (2010) argues that "ex post liability, while useful, does not always by itself provide socially optimal level of risk control. As such, preventative risk regulation will be needed, and the core questions remain about how stringent should such regulation be, and what form such regulation take. Risk regulation research will continue to be needed to provide conceptual clarity to the normative basis for risk standards." Our paper is of interest for policy makers, as it demonstrates the optimal combination of ex ante and ex post mechanisms when market exit decisions are relevant. We also show that the optimal intervention depends on the level of industry competition and on the extent of the accident damages. For the operations management community, this paper is a first step to understanding how policy tools affect not only the incentives to exert operational safety but also the structure of the industry, i.e. the number and type of competitors that firms will face. Second, while previous papers in this area have studied the interplay between public policy and operational risk management, we bring a new perspective to this literature. The extant literature focuses on how should governments correct negative externalities (i.e. on policies that discourage the provision of dangerous activities). Our paper looks at a setting where the social planner intervenes to encourage the provision of hazardous activities. To our knowledge, our paper is the first one to consider this perspective within the operations management risk literature.

Bibliography

Aakvik, Arild, Frank Hansen, Gaute Torsvik. 2013. Dynamic Peer Effects in Sales Teams. *Working Paper* .

Anand, Krishnan, Francois Charles Giraud-Carrier. 2013. Pollution Regulation and Production.

Angrist, Joshua David, Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Antle, John M. 1996. Efficient Food Safety Regulation in the Food Manufacturing Sector. *American Journal of Agricultural Economics* **78**(5) 1242–1247.

Armstrong, Julie. 2003. Suppliers Wary of Increasing Warranty Charges. *Automotive News* .

Arora, A, R Telang, H Xu. 2008. Optimal Policy for Software Vulnerability Disclosure. *Management Science* **54**(4) 642–656.

Bakshi, Nitin, Noah Gans. 2010. Securing the Containerized Supply Chain: Analysis of Government Incentives for Private Investment. *Management Science* **56**(2) 219–233.

Bell, Tom. 2013. Maine Economy Tied to Rail line's Fate. *Portland Press Herald* .

Bhattacharyya, Sugato, Francine Lafontaine. 1995. Double-sided Moral Hazard and the Nature of Share Contracts. *The RAND Journal of Economics* **26** 36.

Bloom, Nicholas, Mark Schankerman, John Van Reenen. 2013. Identifying Technology Spillovers and Product Market Rivalry. *Econometrica* **81**(4) 1347–1393.

Bloomberg. 2011. India Woos GE, Westinghouse With Cap on Atomic Equipment Supply Liability. URL `http://www.bloomberg.com/news/2011-11-17/`.

Blume, Lawrence E, William A Brock, Steven N Durlauf, Yannis M Ioannides. 2010. Identification of Social Interactions. Handbook of Social Economics.

Bramoullé, Yann. 2013. Comment on "Social Networks and the Identification of Peer Effects". *Jorunal of Business, Economics and Statistics (forthcoming)* .

Bramoullé, Yann, Habiba Djebbari, Bernard Fortin. 2009. Identification of peer effects through social networks. *Journal of Econometrics* **150**(1) 41–55.

Brynjolfsson, Erik. 2003. The IT Productivity Gap. *Optimize Magazine* (21).

Brynjolfsson, Erik, Lorin M Hitt. 2003. Computing productivity: Firm-level evidence. *Review of Economics and Statistics* **85**(4) 793–808.

Burk, D L, B A Boczar. 1993. Biotechnology and tort liability: A strategic industry at risk. *University of Pittsburgh, Literature Review* **55**(791).

Business Development Bank of Canada. 2014. How Small Firms can Tap Into Global Supply Chains . *Business Development Bank of Canada* .

Cabral, Luis M B. 2004. Simultaneous entry and welfare. *European Economic Review* **48**(5) 943–957.

Cachon, Gerard P, Santiago Gallino, Marcelo Olivares. 2013. Severe Weather and Automobile Assembly Productivity . *Working Paper* .

Calabresi, Guido. 1970. *The Cost of Accidents*. A Legal and Economic Analysis, Yale University Press.

Cardella, Laura. 2013. Credit Ratings and Inter-firm Dependence. Working Paper, University of Arizona .

Chandra, Amitabh, Amy Finkelstein, Sacarny Adam, Chad Syverson. 2012. Healthcare Exceptionalism? Productivity and Allocation in the U.S. Healthcare Sector. Working paper, University of Chicago.

Chao, Gary H, Seyed M R Iravani, R Canan Savaskan. 2009. Quality Improvement Incentives and Product Recall Cost Sharing Contracts. *Management Science* **55**(7) 1122–1138.

Cheung, May, Jun Zhuang. 2012. Regulation Games Between Government and Competing Companies: Oil Spills and Other Disasters. *Decision Analysis* **9**(2) 156–164.

Chu, Leon Yang, David E M Sappington. 2010. Contracting with private knowledge of signal quality. *The RAND Journal of Economics* **41**(2) 244–269.

Coglianese, Cary. 2010. *The Law and Economics of Risk Regulation*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Dahlby, Bev. 2008. *The marginal cost of public funds*. Theory and applications, The MIT Press.

Danzon, Patricia M, Nuno Sousa Pereira. 2011. Vaccine Supply: Effects of Regulation and Competition. *International Journal of the Economics of Business* **18**(2) 239–271.

de Véricourt, Francis, Denis Gromb. 2014. Financing capacity investment under demand uncertainty. *ESMT Working Paper* .

Desiraju, Ramarao. 2004. Costs and Benefits of Inducing Intrabrand Competition: The Role of Limited Liability. *Marketing Science* **23**(3) 429–450.

Dong, Lingxiu, Brian Tomlin. 2012. Managing Disruption Risk: The Interplay Between Operations and Insurance. *Management Science* **58**(10) 1898–1915.

Ericson, Clifton A. 2005. *Hazard Analysis Techniques for System Safety*. John Wiley & Sons.

Etemad, Hamid, Richard W Wright, Leo Paul Dana. 2001. Symbiotic International Business Networks: Collaboration between Small and Large Firms. *Thunderbird International Business Review* **43**(4) 481–499.

Fernandes, Ana M. 2008. Firm Productivity in Bangladesh Manufacturing Industries. *World Devevelopment* **36**(10) 1725–1744.

Fillard, Brian, Scott Frahm, Amy Mercer, Kendyle Scott. 2011. Developing Collaborative Supplier Partnerships . *University of South Carolina Resource Cooperative Library* .

Food and Drug Administration (FDA). 2010. Outbreaks associated with Fresh and Fresh-Cut Produce. Tech. rep.

Gallagher, Ryan. 2012. Apportioning Potential Liability for the Rail Transportation of Hazardous Commodities. *47th Annual Proceedings of the Canadian Transportation Research Forum* .

Garvey, Jack I, State farm Fire and Casualty Companyx. 1989. Garvey v. State Farm Fire & Casualty Co. *Supreme Court of California. 48 Cal. 3d 395* .

Goldsmith-Pinkham, P, G Imbens. 2013. Social Networks and the Identification of Peer Effects. *Jorunal of Business, Economics and Statistics (forthcoming)* .

Greenwald, B C, J E Stiglitz. 1990. Asymmetric information and the new theory of the firm: Financial constraints and risk behavior. *The American Economic Review* **80**(2) 160–165.

Griliches, Zvi, Jacques Mairesse. 1997. Production Functions: The Search For Identification. Econometrics and Economic Theory in the 20th Century: Ragnar Frisch Centennial Symposium.

Gromb, Denis, David Martimort. 2007. Collusion and the organization of delegated expertise. *Journal of Economic Theory* **137**(1) 271–299.

Hendricks, Kevin B, Vinod R Singhal. 2005. 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.

Hewitt, T. 2008. Who is to Blame-Allocating Liability in Upstream Project Contracts. *Journal of Energy and Natural Resources Law* **26**(2) 177.

Holmstrom, B, J Tirole. 1997. Financial intermediation, loanable funds, and the real sector. *The Quarterly Journal of Economics* **112** 663–691.

Imrohoroglu, Ayse, Selale Tuzel. 2014. Firm Level Productivity, Risk, and Return. *Management Science (forthcoming)* .

Innes, R D. 1990. Limited liability and incentive contracting with ex-ante action choices. *Journal of Economic Theory* **52**(1) 45–67.

Ivanhoe Energy (USA) Inc., Union Oil Company of California. 2001. Joint Venture Agreement .

Jackson, Matthew O. 2010. *Social and Economic Networks*. Princeton University Press.

Jain, Nitish, Karan Girotra, Serguei Netessine. 2013a. Managing Global Sourcing: Inventory Performance. *Management Science (forthcoming)* .

Jain, Nitish, Sameer Hasija, Dana G Popescu. 2013b. Optimal Contracts for Outsourcing of Repair and Restoration Services. *Operations Research* **61**(6) 1295–1311.

Jensen, J Bradford, Robert H McGuckin, Kevin J Stiroh. 2001. The Impact of Vintage and Survival on Productivity: Evidence from Cohorts of U.S. Manufacturing Plants. *Review of Economics and Statistics* **83**(2) 323–332.

Jones, Chris. 2005. *Applied welfare economics*. 1st ed. Oxford University Press, USA.

Kalwani, Manohar U, Narakesari Narayandas. 1995. Long-Term Manufacturer-Supplier Relationships: Do They Pay off for Supplier Firms? *Journal of Marketing* **59**(1) 1–16.

Keller, Wolfgang, Stephen R Yeaple. 2009. Multinational Enterprises, International Trade, and Productivity Growth: Firm-Level Evidence from the United States. *Review of Economics and Statistics* **91**(4) 821–831.

Kellogg, Ryan. 2011. Learning by Drilling: Inter-Firm Learning and Relationship Persistence in the Texas Oilpatch. *The Quarterly Journal of Economics* **126** 1961–2004.

Kim, Sang-Hyun, Morris Cohen, Sergei Netessine. 2007. Performance contracting in after-sales service supply chains. *Management Science* **53**(12) 1843–1858.

Kim, Sang-Hyun, Sergei Netessine. 2013. Collaborative cost reduction and component procurement under information asymmetry. *Management Science* **59**(1) 189–206.

Kim, Sang-Hyun, Brian Tomlin. 2013. Guilt by Association: Strategic Failure Prevention and Recovery Capacity Investments. *Management Science* **59**(7) 1631–1649.

Kogan, Kostantin, Charles Tapiero. 2007. *Supply Chain Games: Operations Management and Risk Valuation: Operations Management and Risk Evaluation*. Springer Science.

Kolstad, Charles D, Thomas S Ulen, Gary V Johnson. 1990. Ex Post Liability for Harm vs. Ex Ante Safety Regulation: Substitutes or Complements? *The American Economic Review* **80**(4) 888–901.

Kraft, T, Y Zheng, F Erhun. 2013. The NGO's Dilemma: How to Influence Firms to Replace a Potentially Hazardous Substance. *Manufacturing and Services Operations Management* **4** 649–669.

Krugman, Paul R. 1997. *The Age of Diminished Expectations*. U.S. Economic Policy in the 1990s, MIT Press.

Laffont, Jean-Jacques, Jean Tirole. 1996. Pollution permits and environmental innovation. *Journal of Public Economics* **62**(1-2) 127–140.

Lee, Lung-fei. 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* **140**(2) 333–374.

Levinsohn, James, Amil Petrin. 2003. Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies* **70**(2) 317–341.

Lieberman, Marvin B, Lieven Demeester. 1999. Inventory Reduction and Productivity Growth: Linkages in the Japanese Automotive Industry. *Management Science* **45**(4) 466–485.

Lilleholt, Kare, Albrecht Zeuner, Harald Koch. 2012. *International Encyclopedia of Comparative Law*, *Instalment 41: Volume VIII Specific Contracts, Chapter 9*, vol. VIII. Martinus Nijhoff Publishing.

Limited, AIG Israel Insurance. 2010. Product Recall Insurance: Policy Underwritings .

Lucent Technologies Inc., Agere Systems Inc. 2001. Separation and Distribution Agreement .

Manning, Richard L. 1994. Changing Rules in Tort Law and the Market for Childhood Vaccines. *Journal of Law and Economics* **37**(1) 247–275.

Manski, Charles F. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* **60**(3) 531–542.

McNish, Jacquie, Grant Robertson. 2013. Probe of Lac-Mégantic train disaster turns to composition of oil . *The Globe and Mail* .

Mosleh, A, D K Rasmuson, F M Marsal. 1998. Guidelines on Modeling Common-Cause Failures in Probabilistic Risk Assessment. Tech. Rep. NUREG/CR-5485.

Moss, Michael. 2009. Food Companies Are Placing the Onus for Safety on Consumers . *The New York Times* .

Okes, Duke. 2009. *Root Cause Analysis: The Core of Problem Solving and Corrective Action*. 1st ed. Q Quality Press.

Olley, G Steven, Ariel Pakes. 1996. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* **64**(6) 1263–1297.

Osborne, Martin J, Ariel Rubinstein. 1994. *A Course in Game Theory*. The MIT Press.

Oyer, Paul. 2000. A Theory of Sales Quotas with Limited Liability and Rent Sharing. *Journal of Labor Economics* **18**(3) 405–426.

Patatoukas, Panos N. 2011. Customer-Base Concentration: Implications for Firm Performance and Capital Markets. *The Accounting Review* **87**(2) 363–392.

Patel, Chandu C, Patricia Grossi, Howard Kunreuther. 2005. *Catastrophe modeling: A new approach to managing risk*. Springer USA.

Plambeck, Erica, Qiong Wang. 2009. Effects of E-Waste Regulation on New Product Introduction. *Management Science* **55**(3) 333–347.

Poblete, Joaquín, Daniel Spulber. 2012. The form of incentive contracts: agency with moral hazard, risk neutrality, and limited liability. *The RAND Journal of Economics* **43**(2) 215–234.

Roels, Guillaume, Urday Karmarkar, Scott Carr. 2010. Contracting for Collaborative Services. *Management Science* **56**(5) 849–863.

RSA Group. 2011. Equipment Breakdown Insurance. Tech. rep.

Sacerdote, Bruce. 2001. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics* **116**(2) 681–704.

Saouma, Richard. 2008. Optimal Second-Stage Outsourcing. *Management Science* **54**(6) 1147–1159.

Sappington, David E M. 1983. Limited liability contracts between principal and agent. *Journal of Economic Theory* **29**(1) 1–21.

Schmitz, Patrick W. 2000. On the joint use of liability and safety regulation. *International review of law and economics* **20**(3) 371–382.

Shavell, Steven. 2009. *Economic Analysis of Accident Law*. 1st ed. Harvard University Press.

Swinney, Robert, Gerard P Cachon, Serguei Netessine. 2011. Capacity Investment Timing by Start-ups and Established Firms in New Markets. *Management Science* **57**(4) 763–777.

Syverson, Chad. 2004. Product Substitutability and Productivity Dispersion. *The Review of Economics and Statistics* **86**(2) 534–550.

Syverson, Chad. 2011. What Determines Productivity? *Journal of Economic Literature* **2** 58.

Tirole, J. 1988. The theory of industrial organization, 1988. *MIT press* .

Tommasi, Mariano, Federico Weinschelbaum. 2007. Principal-Agent Contracts under the Threat of Insurance. *Journal of Institutional and Theoretical Economics JITE* **163**(3) 379–393.

Trebilcock, Michael, Ralph A Winter. 1997. The economics of nuclear accident law. *International review of law and economics* **17**(2) 215–243.

Van Biesebroeck, Johannes. 2005. Firm Size Matters: Growth and Productivity Growth in African Manufacturing. *Economic Development and Cultural Change* **53**(3) 545–583.

Viscusi, W Kip, Michael J Moore. 1993. Product Liability, Research and Development, and Innovation. *Journal of Political Economy* **101**(1) 161–184.

Winter, Ralph A. 2000. Optimal Insurance under Moral Hazard . J David Cummins, Georges Dionne, eds., *Handbook of Insurance*. Springer Netherlands, Dordrecht, 155–183.

Appendix A

Technical Results and Variable Definitions of Chapter 2

A.1 The Olley-Pakes Approach

When we estimate TFP through OLS, the resulting estimates are likely to suffer from simultaneity and selection biases. A simultaneity bias arises because there is systematic correlation between the input factors and the error term. A selection bias arises because a firm's profitability is correlated to its level of capital stock, which is fixed in the short term.

To deal with these problems, Olley and Pakes (1996) introduced a semi-parametric specification that controls for both biases. This approach uses capital investments, I_{it}, as a proxy variable, and makes the following assumptions (which are grounded on empirical results; see Olley and Pakes, 1996) :

1. Labour (l_t) is a variable factor at time t.

   (a) Capital (k_{it}) is a fixed factor at time t, and a function of the productivity level at $t-1$, i.e. $k_{it} = k\left(\rho_{i,t-1}\right)$.

   (b) I_{it} is a function of ρ_{it} and k_{it}, i.e. $I_{it} = I\left(\rho_{it}, k_{it}\right)$. This function satisfies $\frac{\partial I_{it}}{\partial \rho_{it}} > 0$ for any $I_{it} > 0$.

   By assumption 3, we have that $\rho_{it} = h\left(I_{it}, k_{it}\right)$, where h is the inversion of I. Therefore, we can re-write equation (2.1) as

   $$y_{it} = \beta_l l_{it} + \phi_{it}\left(I_{it}, k_{it}\right) + \varepsilon_{it}$$

   where $\phi_{it}\left(I_{it}, k_{it}\right) \equiv h\left(I_{it}, k_{it}\right) + \beta_k k_{it}$.[39] Note that ϕ isolates ρ, which is the source of the simultaneity bias. For this reason, we can estimate y_{it} and obtain consistent estimates for β_l. Although ϕ is unobservable, we

[39] For simplicity of exposition, we ignore industry fixed effects. In the actual calculation, we include these terms.

estimate this function by using a third-order polynomial expansion on I and k, i.e. $\phi_{it}\left(I_{it}, k_{it}\right) \approx c_0 + \sum_{m=0}^{3}\sum_{n=0}^{3} c_{mn} k_{it}^m I_{it}^n$. This estimation yields estimates β_l^{OP} and $\hat{\phi}_{it}$.

Second, we estimate β_k. To this end, let $\Delta\rho_{it} \equiv \rho_{it} - \rho_{i,t-1}$ and assume that $cov(k_{it}, \Delta\rho_{it}) = 0$. Thus,

$$\begin{aligned} y_{it} - \beta_l^{OP} l_{it} &= \beta_k k_{it} + \rho_{it} + \varepsilon_{it} \\ &= \beta_k k_{it} + \rho_{i,t-1} + \Delta\rho_{it} + \varepsilon_{it} \\ &= \beta_k k_{it} + \hat{\phi}_{i,t-1} - \beta_k k_{i,t-1} + \Delta\rho_{it} + \varepsilon_{it} \end{aligned}$$

where $\hat{\phi}_{i,t-1} - \beta_k k_{i,t-1}$ is an unbiased estimate of $\rho_{i,t-1}$.

Now, define $\underline{\rho}_{it}(I_{i,t-1}, k_{it})$ as the productivity threshold for which a company is indifferent between exiting the market and running operations. Also, assume that probability of survival of firm i, P_{it}, is a function of $\rho_{i,t-1}$ and $\underline{\rho}_{i,t-1}$. To estimate P_{it} we run a probit regression on a third order polynomial expansion on $I_{i,t-1}$, $k_{i,t}$ and the firm's age.

We use the estimated survival probability, $\hat{P}_{it}$, to estimate $y_{it} - \beta_l^{OP} l_{it} = \beta_k k_{it} + g\left(\hat{\phi}_{i,t-1} - \beta_k k_{i,t-1}, \hat{P}_{it}\right) + \Delta\rho_{it} + \varepsilon_{it}$, where $g(\cdot)$ is an unknown function. To approximate this function, we use a third order polynomial expansion on its parameters, where $g\left(\hat{\phi}_{i,t-1} - \beta_k k_{i,t-1}, \hat{P}_t\right) \approx c_0 + \sum_{m=0}^{3}\sum_{n=0}^{3} c_{mn}\left(\hat{\phi}_{i,t-1} - \beta_k k_{i,t-1}\right)^m \hat{P}_{it}^n$.

This allows us to obtain β_k^{OP}. We use β_l^{OP} and β_k^{OP} to derive the firm's productivity, where,

$$\rho_{it} = y_{it} - \beta_k^{OP} k_{it} - \beta_l^{OP} l_{it}$$

A.2 The Formation of Supply Chain Networks

Identification Strategy

To control the selection of supply chain partners, we build an approach that jointly estimates two types of processes: (1) the selection process and; (2) the determinants of productivity (i.e. the firm effects and the spillover effects), *conditional* on the selection process. We explain these two processes below and, thereafter, our estimation technique.

Process 1: Selection of Supply Chain Partners

A link between a supplier and a customer is formed as a result of two choices, i.e. the customer selecting a potential supplier, and the supplier agreeing to trade with the customer. To simplify our model we assume that suppliers are *always* willing to trade with any customer. Although this assumption does not always hold,[40] it is often the case that customers are the dominant decision makers at the time of forming links.

Consider the network matrix $\mathbf{C}$, and recall that this matrix has typical element $c_{ij} = 1$ if j is a customer of i, and $c_{ij} = 0$ otherwise. At the beginning of period t, firm j evaluates the expected profitability of forming a link with supplier i, $U_j(i)$. We assume that j selects i as a supplier if and only if $U_j(i) > 0$. In other words,

$$c_{ij} = \mathbf{1}_{U_j(i)>0}$$

The term $U_j(i)$ depends primarily on the compatibility between the firms' industries. For example, manufacturers of home furniture are likely to require inputs from firms in the wood and lumber industry, but less likely to source from, say, firms in the air travel industry. The term $U_j(i)$ also depends on the observed characteristics of both the customer and the supplier, $\mathbf{X}_i$ and $\mathbf{X}_j$, and on the state of their relationship in the previous period. Specifically, let

$$U_j(i) = \alpha_0 + \alpha_c c_{0ij} + \alpha_V V_{0ij} + \alpha_{geo} geo_{ij} + \sum_k \alpha_k x_{ki} + \eta_{ij} \tag{A.1}$$

The variable c_{0ij} is equal to 1 if j was a customer of supplier i in the *previous* period $(t-1)$, and $c_{0ij} = 0$ otherwise. Let $\mathbf{C}_0$ be the matrix with typical element c_{0ij}.

V_{0ij} is a proxy to determine the compatibility between i's industry and j's industry. Specifically V_{0ij} measures the proportion of firms in j's industry that were customers of a supplier in i's industry, at time $t-1$.[41]

The function geo_{ij} is a categorical variable representing the firms' geographic proximity. Specifically, geo_{ij} is equal to one if the firms are in the same geographic region (midwest, northeast, etc.), and zero otherwise.

We also include a subset of idiosyncratic characteristics about the supplier, including the level of inventory turnover, financial leverage, size and

[40] For example, the supplier may have inventory constraints, exclusivity agreements, or reputational concerns that prevent him from selling to a given customer.

[41] When constructing this variable, we exclude the relationship between i and j. This allows us to avoid collinearity issues between $c_{0,ij}$ and $V_{0,ij}$.

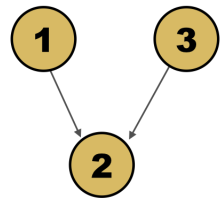


Figure A.1: A sample network, where $c_{12} = c_{32} = 1$ and $c_{ij} = 0$ everywhere else.

age. All of these characteristics are measured at the beginning of the period. Finally, we assume that all the error terms, η_{ij}, are iid and follow a logistic distribution.

Let α be the vector of coefficients from equation (A.1). If we define p_{ij} as the probability of observing a link between firms i and j, we thus have that

$$\begin{aligned} p_{ij} &\equiv pr\left(c_{ij} = 1\right) \\ &= pr\left(U_j\left(i\right) > 0\right) \\ &= \frac{\exp\left(\alpha_0 + \alpha_c c_{0,ij} + \alpha_V V_{0,ij} + \alpha_{geo} geo_{ij} + \sum_k \alpha_k x_{ki}\right)}{1 + \exp\left(\alpha_0 + \alpha_c c_{0,ij} + \alpha_V V_{0,ij} + \alpha_{geo} geo_{ij} + \sum_k \alpha_k x_{ki}\right)} \end{aligned}$$

where the last equality follows from the fact that the error terms, η_{ij}, follow a logistic distribution. Observe that p_{ij} is a function of α, and is conditional on: (i) the state of the network at $t-1$, $\mathbf{C}_0$, and; (ii) on the characteristics of both firms, $\mathbf{X}_i$ and $\mathbf{X}_j$. In other words, p_{ij}=$p_{ij}(\alpha|\mathbf{X_i}, \mathbf{X_j}, \mathbf{C_0})$

Therefore, the likelihood of observing network matrix $\mathbf{C}$, conditional on $\mathbf{X}$ and $\mathbf{C}_0$, is equal to

$$\begin{aligned} &\mathcal{L}_{network}\left(\alpha|\mathbf{C};\mathbf{X},\mathbf{C}_0\right) \\ =\prod_{i\neq j}&\left[p_{ij}\left(\alpha|\mathbf{X_i},\mathbf{X_j},\mathbf{C_0}\right)^{c_{ij}} \times \left(1 - p_{ij}\left(\left(\alpha|\mathbf{X_i},\mathbf{X_j},\mathbf{C_0}\right)\right)\right)^{1-c_{ij}}\right] \end{aligned}$$

For example, consider the network from Figure A.1 and suppose that, given weights α, we have $p_{12} = p_{31} = 0.3$, $p_{21} = p_{13} = 0.4$ and $p_{23} = p_{32} = 0.8$. We can verify that the likelihood of observing this network is $\mathcal{L}_{network} = \left[\left(p_{12}\right)^{c_{12}}\left(1-p_{12}\right)^{1-c_{12}}\right] \times \left[\left(p_{13}\right)^{c_{13}}\left(1-p_{13}\right)^{1-c_{13}}\right]\ldots. \; =\left[\left(0.3\right)^{1}\left(0.7\right)^{0}\right] \times \left[\left(0.4\right)^{0}\left(0.6\right)^{1}\right]\ldots = 0.0121$.

Process 2: Determinants of Productivity

The second process estimates the determinants of productivity: (i) the firm effects, γ; (ii) the exogenous spillover effects, θ, and; (iii) the endogenous effect, ω.

Consider equation (2.2) and let $\mathcal{L}_{TFP}\left(\gamma, \theta, \omega | \mathbf{TFP}; \mathbf{C}, \mathbf{X}\right)$ be the likelihood of observing the productivity vector, **TFP,** given coefficient vectors (γ, θ, ω). This likelihood is conditional on: (i) network matrix **C** (which is formed in Process 1) and; (ii) the characteristics of the firms, **X.**

Also, assume that the error terms from equation (2.2) are all normally distributed with mean 0 and variance σ^2. Under this assumption, we can verify that $\mathcal{L}_{TFP}\left(\gamma, \theta, \omega | \mathbf{TFP}; \mathbf{C}, \mathbf{X}\right) = \phi\left(\mu_{TFP}, \Sigma_{TFP}\right)$, where ϕ is the Probability Density Function of the Normal distribution, and where $\mu_{TFP} = \left(\mathbf{I} - \omega\mathbf{W}\right)^{-1}\left(\gamma + \theta\mathbf{W}\right)$ and $\Sigma_{TFP} = \sigma^2\left(\mathbf{I} - \omega\mathbf{W}\right)^{-1}\left(\mathbf{I} - \omega\mathbf{W}'\right)^{-1}$.

Estimation Procedure

As explained above, we will jointly estimate: (i) the determinants of the selection process (α), and (ii) the determinants of productivity(γ, θ, ω), conditional on the selection process. Our goal here is to find the vector $(\alpha, \gamma, \theta, \omega)$ that maximizes

$$\begin{aligned} \mathcal{L}\left(\alpha, \gamma, \theta, \omega | \mathbf{TFP}, \mathbf{C}, \mathbf{X}, \mathbf{C_0}\right) &= \underbrace{\mathcal{L}_{network}\left(\alpha | \mathbf{C}; \mathbf{C_0}, \mathbf{X}\right)}_{\text{selection process}} \times \\ &\times \underbrace{\mathcal{L}_{TFP}\left(\gamma, \theta, \omega | \mathbf{TFP}; \mathbf{X}, \mathbf{C}\right)}_{\text{TFP determinants}} \end{aligned}$$

This above estimation will allow us to control for partner selection biases when estimating spillover effects. To estimate this function we use the same approach that can be found in the Appendix of Goldsmith-Pinkham and Imbens (2013). This approach uses Monte-Carlo-Markov-Chain (MCMC) algorithms to estimate the distribution of the likelihood function.

Data Description

Given the size of our sample, firms are able to choose (literally) billions of combinations at the time of designing a supply chain structure. For this reason, it is impossible to obtain results arising from a large sample. Therefore, we limit our estimation to a subsample of firms. In this subsample we only include the most recent observations, i.e. those recorded between the years 2006 and 2009. This subsample contains a panel of 1,697 firms and approximately 5,600 linkages.

Variable	***Coeff.***	***t-stat***
Previous relationship (c_{oij})	7.136***	66.25
Industry synergy (V_{oij})	8.825 ***	52.39
Geographic Proximity (geo_{ij})	0.50***	10.29
Supplier's size	0.022***	2.62
Supplier's age	0.009***	6.63
Supplier's inventory turnover	0.112	1.57
Supplier's financial leverage	0.150*	1.82
constant	-4.95***	-97.92
Linkages	5,605	

*** p<0.01, **$p < 0.05$,*p <0.1

Table A.1: Model estimates for the selection equation.

Estimates

In Table A.1, we report estimates for the factors that influence a firm's selection of partners (α). According to these estimates, there are two key predictors in the formation of networks: the synergy between the industries, and the existence of a linkage in the previous year. From here we can see that supply chains are largely determined by factors that are exogenous.

We also find that a supplier is more likely to be selected by a firm if it is large, aged, or if the supplier is located in the same geographic region as the customer. A supplier's financial leverage and inventory control do seem to have a positive impact on the selection process, albeit this impact is not very robust (i.e. it has a relatively small $t-$value).

In Table A.2, we report estimates for the determinants of productivity. In Column 1, we report the Lee estimates without controlling for selection process; in Column 2, we report the estimates with these controls.

When we fail to control for the selection process, our estimates overreport the magnitude of the endogenous effects. Specifically, we have that ω decreases from 0.4826 to 0.3232.

The exogenous spillover effects are not highly affected by the partner selection biases. Note, however, that the spillover effect of inventory turnover is negative in Column 2. Although this effect contradicts the estimates from Column 1, the effect is statistically insignificant.

Effect	Variable	(1) No Controls *Coeff.*	(1) No Controls *t-stat*	(2) Selection controls *Coeff.*	(2) Selection controls *t-stat*
Endogenous effect (ω)	Cust. TFP	0.4826**	(2.16)	0.3232**	(1.98)
Exogenous spillover effects (θ)	Age	0.0185	(1.50)	0.2316	(0.60)
	Age^2	-0.0003*	(-1.72)	-0.0004	(-0.69)
	Size	0.0259	(1.01)	0.06195	(0.87)
	Leverage	-0.0082	(-1.35)	-0.1122	(-1.14)
	Inv. Turnover	1.8121***	(2.65)	-1.4732	(-1.44)
	Inv. Turnover^2	-0.1561	(-0.55)	-0.1040	(-0.71)
	Region: West	0.3879*	(1.74)	0.2292	(1.45)
	Region: Midwest	0.4222***	(2.66)	0.1032	(0.8)
	Region: South	0.1966*	(1.70)	01040	(1.04)
	Region Northeast	-0.0162	(-0.11)	0.10273	(0.86)
Firm effects (γ)	Age	0.0044	(0.49)	0.02118	(1.14)
	Age^2	-0.0001	(-0.63)	-0.0036	(-1.44)
	Size	0.0595***	(3.81)	0.03923**	(2.05)
	Leverage	-0.1097	(-1.11)	-0.0105	(-0.22)
	Inv. Turnover	-1.6752*	(-1.95)	-1.8446	(-1.61)
	Inv. Turnover^2	2.1655	(0.99)	1.3447	(1.61)
	Region: West	0.1900*	(1.65)	0.5122	(0.87)
	Region: Midwest	0.0811	(0.71)	0.2357	(1.18)
	Region: South	0.0906	(0.97)	0.2234***	(2.70)
	Region Northeast	0.0786	(0.80)	0.0546	(0.49)
Number of Observations		5,605		5,605	

*** p<0.01, **p < 0.05,*p <0.1

Table A.2: Model estimates for the linear-in-means equation.

A.3 Variables Definition and Construction

- **Sales:** Net sales [SALE in Compustat], deflated by the 4-digit industry-level deflator. We obtain the deflators from the NBER-CES Manufacturing Industry Database.

- **Capital:** Gross property, plant and equipment [PPEGT], deflated by the price deflator for investment. Since investment is made at various times in the past, it would be imprecise to use the current year's capital deflator. To deflate capital stock, we use method adopted by Brynjolfsson and Hitt (2003). In this method, we deflate capital stock at the calculated average age of capital. To calculate the average age of capital stock, we divide accumulated depreciation [DPACT] by current depreciation [DP].

- **Materials:** Total Expenses-labour expenses. Total expenses is equal to Sales minus operating income before depreciation and amortization. We obtain Operating Income from item 13 in compustat [OIBDP]. The resulting value is deflated by the 4-digit deflator for materials

- **Labour expense:** To proxy for labour expense we use sector-average labour cost per employee, and multiply it to the total number of employees [EMP]. To determine the average sector labour cost, we use the annual sector-level wage data (salary plus benefits) from the Bureau of Labor Statistics. Labour expense is deflated by the price index for total labour compensation.

- **Value added:** Sales (deflated) - Materials (deflated).

- **Capital investments (I):** Capital expenditures, from Compustat [CAPX], deflated by the 4-digit industry-level deflator.

- **Firm age:** Proxied by the year the firm first appeared in Compustat.

- **Firm size:** Measured by the natural logarithm of total assets [AT], from Compustat.

- **Financial leverage:** Measured as the ratio of total debt (short-term debt [DLC] + long-term debt [DLTT]) to the book value of total assets.

- **Inventory turnover:** Measured as the ratio of net sales to the level of inventory [INVT].

- **Geographic regions:** Constructed by using the headquarters' location of each firm. We looked at their home state [STATE] in Compustat. If this variable was not reported, we looked at their home city [CITY] to verify if the company is located in the U.S. or abroad.

- **Customer weight** (w_{ij}) : Constructed as the ratio of the customer sales [CSALE] to the sum of all customer sales in that year. This variable was obtained from the Compustat business segments.

Appendix B

Proofs for Chapter 3

Proof of Lemma 1: In stage 2, the integrated supply chain solves $\max_{e_S,e_O} E_X\left[\Pi\left(e_S,e_O,v,X\right)\right] = \pi - \left(\left(1-v\right)xF\left(e_S,e_O\right)+P+e_S+e_O\right)$. By taking the first order conditions with respect to e_O and e_S we find that any critical point must satisfy $\frac{x\beta(1-v)}{\left(e_O^*+1\right)^{\beta+1}\left(e_S^*+1\right)^{1-\beta}} = 1$ and $\frac{x(1-\beta)(1-v)}{\left(e_O^*+1\right)^{\beta}\left(e_S^*+1\right)^{2-\beta}} = 1$. By solving for e_O^* and e_S^*, we find that $e_O^* = \sqrt{\frac{x(1-v)\beta^{2-\beta}}{(1-\beta)^{1-\beta}}} - 1$ and $e_S^* = \sqrt{\frac{x(1-v)(1-\beta)^{1+\beta}}{\beta^{\beta}}} - 1$.

To verify that this critical point is a maximum, we use the second order partial derivative test. We have that $\frac{\partial E[\Pi]}{\partial e_S^2} = -\frac{x(1-v)(1-\beta)(2-\beta)}{(e_O+1)^{\beta}(e_S+1)^{3-\beta}} < 0$, $\frac{\partial E[\Pi]}{\partial e_O^2} = -\frac{x(1-v)(\beta+1)\beta}{(e_O+1)^{\beta+2}(e_S+1)^{1-\beta}} < 0$ and $\frac{\partial E[\Pi]}{\partial e_O \partial e_S} = -\frac{x(1-v)\beta(1-\beta)}{(e_O+1)^{\beta+1}(e_S+1)^{2-\beta}} < 0$, and $\frac{\partial E[\Pi]}{\partial e_S^2}\frac{\partial E[\Pi]}{\partial e_O^2} - \left(\frac{\partial E[\Pi]}{\partial e_O \partial e_S}\right)^2 = \left[\left(\frac{(x(1-v))}{(e_O+1)^{\beta+1}(e_S+1)^{2-\beta}}\right)^2\left(2\beta\left(1-\beta\right)\right)\right] > 0$. Hence, $\left(e_S^*, e_O^*\right)$ is a maximum. ■

Proof of Proposition 2: In stage 1, the supply chain solves $\max_{v\in[0,1]} E_X\left[\Pi^{\mathcal{FB}}\left(v,X\right)\right]$ subject to $\Pi^{\mathcal{FB}}\left(v,x\right) + W \geq 0$, where $\Pi^{\mathcal{FB}}\left(v,X\right) \equiv \Pi\left(e_S^{\mathcal{FB}}\left(v\right), e_O^{\mathcal{FB}}\left(v\right), v, P\left(v|e_S^{\mathcal{FB}}\left(v\right), e_O^{\mathcal{FB}}\left(v\right)\right), X\right)$. By Lemma 1 we can write $E_X\left[\Pi^{\mathcal{FB}}\left(v,X\right)\right] = \pi - \left(\sqrt{\frac{x}{\beta^{\beta}(1-\beta)^{1-\beta}}}\right)\left(\sqrt{\frac{(2-v)^2}{(1-v)}}\right) + 2$. Because $\frac{\partial}{\partial v}\sqrt{\frac{(2-v)^2}{1-v}} = \sqrt{\frac{(2-v)^2}{1-v}}\frac{v}{2(2-v)(1-v)}$ is positive for any $v \in [0,1]$, it follows that $\frac{\partial}{\partial v}E\left[\Pi^{\mathcal{FB}}\left(v,x\right)\right] < 0$ for any $v \in [0,1]$. Therefore, $\arg\max_{v\in[0,1]}\left\{E_X\left[\Pi^{\mathcal{FB}}\left(v,X\right)\right]\right\} = 0$. In other words, $v = 0$ maximizes the (unconstrained) profits for the centralized supply chain and, hence, the supply chain chooses $v = 0$ whenever $\Pi^{\mathcal{FB}}\left(0,x\right) + W \geq 0$.

Now suppose that $\Pi^{\mathcal{FB}}\left(0,x\right) + W < 0$. Here, $v = 0$ is not in the feasible region. Furthermore, for any $x \geq 1$,[42] it is true that $\frac{\partial}{\partial v}\Pi^{\mathcal{FB}}\left(v,x\right)|_{v=0} = \sqrt{x}\left(\sqrt{x} - \sqrt{\frac{1}{4(1-\beta)^{1-\beta}\beta^{\beta}}}\right) > 0$, which implies that the wealth constraint is

[42]Note that because we assume that x must be large to guarantee an interior solution for the effort levels, it can be shown that $x \geq 1$.

relaxed when v is increased beyond 0. Using this fact in conjunction with the fact that $\frac{\partial}{\partial v}E_X\left[\Pi^{\mathcal{FB}}(v,X)\right]<0$, we find that the optimum is given by $v=v^{\mathcal{FB}}$ where $\Pi^{\mathcal{FB}}\left(v^{\mathcal{FB}},x\right)+W=0$. ■

Proof of Lemma 3: In stage 2, the best response level of effort for the supplier i→s equal to $\tilde{e_S}|_{e_O}=\arg\max_{e_S\geq 0}E_X\left[\Pi_S\left(e_S,w,y,P,X|e_O\right)\right]$ $=\left(\frac{xy(1-\beta)}{(e_O+1)^\beta}\right)^{\frac{1}{2-\beta}}-1$, and the best response function for the operator is $\tilde{e_O}|_{e_S}=E_X\left[\Pi_O\left(e_O,w,y,v,P,X|e_S\right)\right]=\left(\frac{x\beta(1-y-v)}{(e_S+1)^{1-\beta}}\right)^{\frac{1}{\beta+1}}-1$. The Nash equilibrium effort levels is given by the fixed point correspondences of the best response functions, i.e. $e_S^*\equiv\tilde{e_S}|_{\tilde{e_O}}$ and $e_O^*\equiv\tilde{e_O}|_{\tilde{e_S}}$, where $e_O^*(y,v)=\sqrt{\frac{x(\beta(1-y-v))^{2-\beta}}{((1-\beta)y)^{1-\beta}}}-1$ and $e_S^*(y,v)=\sqrt{\frac{x(y(1-\beta))^{1+\beta}}{(\beta(1-y-v))^\beta}}-1$.■

Proof of Lemma 4: This Lemma follows directly from Lemma 3, by solving explicitly for $F\left(e_S^*(y,v),e_O^*(y,v)\right)=\left[\left(1+e_O^*(y,v)\right)^\beta\left(1+e_S^*(y,v)\right)^{1-\beta}\right]^{-1}=\left[\sqrt{x\left(\beta\left(1-y-v\right)^\beta\right)\left((1-\beta)\,y\right)^{1-\beta}}\right]^{-1}$ ■

Proof of Proposition 5: We prove our results for the case where $\beta\neq 0.5$, as the proof for the case where $\beta=0.5$ is identical. In region $\mathcal{UC}$, the operator seeks to find $max_{w,y,v}\ E_X\left[\Pi_O\left(e_O^*,w,y,v,P\left(v|e_S^*,e_O^*\right),X\right)|e_S^*\right]$ subject to $(\mathcal{IR})$, $y\in[0,1]$ and $v\in[0,1-y]$,$e_S^*=\arg\max_{e_S\geq 0}\{\Pi_S\}$. First, note that if $\mathcal{IR}$ is non-binding and (w^*,y^*,v^*) are maximizers, then $w^*>xy^*\Phi(y^*,v^*)+e_S^*$. But if this were true, the operator could decrease w^* by $\varepsilon=\frac{w-xy^*\Phi(y^*,v^*)-e_S^*}{2}$, which increases $E_X[\Pi_O]$ without violating any constraint. Therefore, $w^*=xy^*\Phi(y^*,v^*)+e_S^*$, which implies that $\mathcal{IR}$ is binding.

If we plug $w^*=xy^*\Phi(y^*,v^*)+e_S^*$ into the objective function, we can re-write the operator's problem as $E\left[\Pi_O^*(y,v)\right]=\frac{x(y+\beta-v\beta-2y\beta+1)}{\Phi(y,v)}+2$ subject to $y\in[0,1]$ and $v\in[0,1-y]$. But note that $\lim_{y\to 0}\Pi_O^*(y,v)=\lim_{y\to 1}\Pi_O^*(y,v)=-\infty$ and $\lim_{v\to 1-v}\Pi_O^*(y,v)=-\infty$. Hence the optimal solution must be such that $y\in(0,1)$ and $v\in[0,1-y)$. Under this assumption, we write the Lagrangian function of the operator's problem as $L(y,v)\equiv x\Phi(y,v)+e_S^*(y,v)+e_O^*(y,v)-\lambda v$, where λ is the Lagrangian multiplier for the non-negativity constraint, $v\geq 0$. In this program, the KKT conditions are equal to $x\Phi_y+e_{0,y}^*+e_{S,y}^*=0$,$x\Phi_v+e_{0,v}^*+e_{S,v}^*=\lambda$,$\lambda\geq 0$,$v\geq 0$ and $v\lambda=0$.[43] Through algebraic arrangements, we can show that if $v\neq 0$,

[43]Where, to simplify notation, we let $\Phi_y\equiv\frac{\partial\Phi(y,v)}{\partial y}$,$e_{S,y}^*\equiv\frac{\partial e_S^*(y,v)}{\partial y}$,etc.

the KKT conditions are only solved when $v = -\beta - 1$ and $y = \beta + 1$, which violates the non-negativity of v. Therefore, the KKT conditions can only be solved when $v^* = 0$ and $y^* = \frac{(1+\beta)\left(\beta(1+\beta)-2+\sqrt{\beta(8-7\beta-2\beta^2+\beta^3)}\right)}{2\left((1+\beta)^2-2\right)}$.

To verify that y^* is a maximum over the region where $v = 0$, we check the second order condition: $\frac{\partial^2 E[\Pi_O(y,0)]}{\partial y \partial y} = \frac{-\Phi(y,0)xh(y,\beta)}{4y^2(1-y)^2}$, where $h(y,\beta) \equiv \beta(2-\beta)y^3 - \left(\beta^3 + 2\beta^2 - 2\beta - 3\right)y^2 + \left(\beta^3 - 6\beta^2 - \beta + 6\right)(1-y) - 3\left(1-\beta^2\right) - 3\left(1-\beta^2\right)$. Note that $h(y,\beta)$ is positive because: (i) $\arg\min_{y\in[0,1],\beta\in[0,1]} h(y,\beta) = \{(1,0),(0,1)\}$ and; (ii) $h(0,1) = h(1,0) = 0$. Hence, $\frac{\partial^2 E[\Pi_O(y,0)]}{\partial y \partial y} = \frac{-\Phi(y,0)xh(y,\beta)}{4y^2(1-y)^2} < 0.$■

Proof of Proposition 6: Consider the operator's problem in region $\mathcal{SC}$. Here, the operator must find $\max_{w,y,v} E_X\left[\Pi_O\left(e_O^*, w, y, v, P\left(v|e_S^*, e_O^*\right), X\right)|e_S^*\right]$ subject to constraints e_S^*, $v \in [0, 1-y]$ and $y \in [0,1]$. By the same argument as the one made in the Proof of Lemma 5, we can show that $y^* \in (0,1)$ and $v^* \in [0, 1-y)$.

By Lemma 5, we can show that the ex post profits for the supplier in case $\mathcal{UC}$, given an operational failure, are $\Pi_S\left(e_O^*\left(y^{\mathcal{UC}}, 0\right), w^{\mathcal{UC}}, y^{\mathcal{UC}}, 0, x|e_S\left(y^{\mathcal{UC}}, 0\right)\right) = y^{\mathcal{UC}}x\left(1-\Phi^{\mathcal{UC}}\right)$. Now, by definition of $\mathcal{WC_S}$, we know that if $W_S + \Pi_S = W_S + y^{\mathcal{UC}}x\left(1-\Phi^{\mathcal{UC}}\right) \equiv W_S^I > 0$, then $\mathcal{WC_S}$ is not binding and, therefore, the optimal solution must be given by $v^{\mathcal{SC}} = 0$,$w^{\mathcal{SC}} = w^{\mathcal{UC}}$ and $y^{\mathcal{SC}} = y^{\mathcal{UC}}$.

Now, assume that $W_S < W_S^I$ or, alternatively, that $\mathcal{WC_S}$ binds at optimum. Therefore, by this constraint we know that $w^* = yx + e_S^*(y,v) - W_S$. We can plug w^* to the objective function and re-write the Lagrangian of operator's problem as $L(w,y,v,\lambda,\mu) \equiv \pi - (x\Phi(y,v)(1-y) + xy + e_O^*(y,v) + e_S^*(y,v) + W_S) - v\lambda - \mu\left(W_S - xy\left(1-\Phi(y,v)\right)\right)$. Here μ and λ are the shadow price for $\mathcal{IR}$ and for the non-negativity constraint for v. The KKT conditions give us $x\Phi - x\Phi_y(1-y) - x - e_{O,y}^* - e_{S,y}^* - \mu x\left((1-\Phi) - y\Phi_y\right) = 0$, $x\Phi_v(1-y) + e_{O,v}^* + e_{S,v}^* - \lambda + \mu\left(xy\Phi_v\right) = 0$, $\mu\left(xy(1-\Phi) - W_M\right) = 0$, $v\lambda = 0$, $\mu \geq 0$, $v \geq 0$ and $\lambda \geq 0$.

We can re-write these conditions to show that $\lambda = \frac{x\beta((1-\Phi)(y+v)-(1-\beta)(1-v-2y))}{\left(\frac{2}{\Phi}-(1+\beta)\right)(1-y-v)-y\beta}$ and $\mu = \left(\frac{2x\beta((1-\Phi)(y+v)-(1-\beta)(1-v-2y))}{\left(\frac{2}{\Phi}-(1+\beta)\right)(1-y-v)-y\beta}\right)\left(\frac{(1-y-v)-\frac{x\Phi\beta}{2}(1-y\beta-(2-\beta)(1-y-v))}{yx\Phi\beta}\right)$. By rearranging these conditions, we can see that when $W_S > W_S^{II} \equiv x(1-\beta)\left(1-2y^{\mathcal{UC}}\right)$ but $W_S \leq W_S^I = y^{\mathcal{UC}}x\left(1-\Phi^{\mathcal{UC}}\right)$, the KKT solutions can only be solved at a point where $\lambda > 0$ and $\mu > 0$, i.e. when both the $\mathcal{IR}$

and non-negativity constraints are binding. The optimal solution is thus given by $v = 0$ and $y^{\mathcal{SC}}$, where $y^{\mathcal{SC}}$solves $(1 - \Phi(y, 0)) = \frac{W_S}{xy}$.

Now, we can also check that when $W_S \leq W_S^{II}$, the non-negativity constraint is non-binding, i.e. $\lambda = 0$ and $v > 0$. Here, if $W_S \geq xm\left(1 - \frac{m(2-\beta)}{1-\beta(1-m)}\right)$, where m solves $x(1-\beta)m\left(\frac{(1-m\beta)\beta}{(2-\beta)m(1-\beta)}\right)^{\beta} = \left(\frac{(1+m\beta-\beta)}{(2-\beta)m}\right)^2$, the individual rationality constraint is binding and, at optimum $v^{\mathcal{SC}} = y^{\mathcal{SC}}\left(\frac{x(1-y^{\mathcal{SC}})(1-\beta)}{W_S+xy^{\mathcal{SC}}(1-\beta)} - 1\right) > 0$ where $y^{\mathcal{SC}}$ satisfies $y^{\mathcal{SC}}\left(1 - \Phi^{\mathcal{SC}}\right) = \frac{W_S}{x}$. Finally, if $W_S < xm\left(1 - \frac{m(2-\beta)}{1-\beta(1-m)}\right)$, the $\mathcal{IR}$ constraint is non-binding (i.e. $\mu = 0$). Therefore, $v^{\mathcal{SC}} = 2\frac{(1-\beta)\left(1-2y^{\mathcal{SC}}\right)}{2-\beta} > 0$ and $y^{\mathcal{SC}}$ satisfying $\Phi^{\mathcal{SC}} = \frac{y^{\mathcal{SC}}(2-\beta)}{1-\beta(1-y^{\mathcal{SC}})}$. ■

Proof of Proposition 7: In region $\mathcal{OC}$ the operator seeks to find max $E_X\left[\Pi_O\left(e_O^*, w, y, v, P\left(v|e_S^*, e_O^*\right), X\right)|e_S^*\right]$ subject to $\mathcal{IR}$, $\mathcal{WC}_{\mathcal{O}}$, $y \in [0,1]$ and $v \in [0, 1-y]$. By Lemma 5, the ex post profits for the operator (in region $\mathcal{UC}$), given the realization of an operational failure, are equal to $\Pi_O = 2 + \pi - x\Phi^{\mathcal{UC}}\left(2y^{\mathcal{UC}}(1-\beta) + \beta\right) - x\left(1 - y^{\mathcal{UC}}\right)$. Hence, constraint $\mathcal{WC}_{\mathcal{O}}$ is non-binding when $W_O > W_O^I \equiv \left(2 + \pi - x\Phi^{\mathcal{UC}}\left(2y^{\mathcal{UC}}(1-\beta) + \beta\right) - x\left(1 - y^{\mathcal{UC}}\right)\right)$, which implies that the operator's problem is identical to the one in case $\mathcal{UC}$, i.e. that $v^{\mathcal{OC}} = 0$ and $y^{\mathcal{OC}} = y^{\mathcal{UC}}$.

Now, assume that $W_O \leq W_O^I$, i.e. that $\mathcal{WC}_{\mathcal{O}}$ binds at optimum. As a result, we can re-write $\Pi_O(e_O, w, y, v, P, x) + W_O = 0$. Moreover, by an argument similar to the one made in the Proof of Lemma 5, we have that y^* $\in (0,1)$, $v^* \in [0, 1-y)$ and $\mathcal{IR}$ is binding. As a result, we can re-write the operator's problem as $\max_{y,v} \pi - x\Phi(y,v) - e_O^*(y,v) - e_S^*(y,v)$ subject to $\Phi(y,v)\,x(y+v) + x(1-y-v) + e_O^*(y,v) + e_S^*(y,v) - \pi = W_O$ and $v \geq 0$.

If we let μ and λ be the shadow prices for constraint $\mathcal{WC}_{\mathcal{O}}$ and for the non-negativity constraint of v, we can re-arrange the KKT conditions to show that $\lambda = \frac{\left(\frac{2(1-\beta)}{\beta}(1-y-v)^2\right)((1-y)(1-\Phi)+\beta(1-v-2y))}{((v-1)(v+2y-1)\beta^2+(v(3-y)-2(v^2+y^2)+y-1)\beta+(v^2-v(1+y)+2y(1-y)))\Phi-2y(1-v-y)}$. By using this condition, jointly with the conditions that $v \geq 0$, $\lambda \geq 0$, $\lambda v = 0$ and with the fact that $\mathcal{WC}_{\mathcal{O}}$ is binding, we find that when W_O is larger than $W_O^{II} \equiv x\Phi(m,0)(2m(1-\beta)+\beta) + x(1-m) - 2 - \pi$, where m solves $\frac{x((1-m)\beta)^{\beta}\left(1+\beta\left(\frac{1-2m}{1-m}\right)\right)^2}{(m(1-\beta))^{\beta-1}} = 1$, then $v^{\mathcal{OC}} = 0$ and $y^{\mathcal{OC}}$ satisfies $2 + W_O = x\Phi^{\mathcal{OC}}(2(1-\beta)+\beta) + x\left(1 - y^{\mathcal{OC}}\right)$. However, we can also verify that when $W_O \leq W_O^{II}$, then the KKT solutions are solved by setting $\lambda = 0$, $v^{\mathcal{OC}} = \left(\frac{1}{1-\beta}\right)\left(\frac{y^2}{2} - \frac{(1-y)(W_O+2-xy)(1-\beta)-x}{x\beta}\right)^{\frac{1}{2}} - \beta + 2y\left(\frac{3}{4}+\beta\right)$ and $y^{\mathcal{OC}}$ satisfying $\Phi^{\mathcal{OC}} = \frac{\beta\left(1-v^{\mathcal{OC}}-y^{\mathcal{OC}}\right)}{1-y^{\mathcal{OC}}} + 1.$■

Proof of Proposition 8: Suppose that $W_S \leq \infty$ and $W_O \leq \infty$. In this case, the operator's problem is to find $\max_{w,y,v} E_X\left[\Pi_O\left(e_O^*, w, y, v, P\left(v|e_S^*, e_O^*\right), X\right)|e_S^*\right]$ subject to $\mathcal{IR}$, $\mathcal{WC_S}$,$\mathcal{WC_O}$, $y \in [0,1]$ and $v \in [0, 1-y]$. By Propositions 6 and 7, we know that if $W_S > W_S^I$ and $W_O > W_O^I$ then the $\mathcal{WC_S}$ and $\mathcal{WC_O}$ constraints are non-binding. As a result, the operator's problem reduces to the one analyzed in case $\mathcal{UC}$.

Suppose that $W_S \leq W_S^I$. By Proposition 6, we know that $\mathcal{WC_S}$ constraint is binding. Moreover, we can verify that, in this sub-region, the $\mathcal{WC_O}$ constraint is binding if and only if $W_O < \tilde{W}\left(y^{\mathcal{SC}}, v^{\mathcal{SC}}\right) - W_S$. If this constraint is non-binding, the operator's problem is identical to the one analyzed in case $\mathcal{SC}$. If, however, $W_O \leq \tilde{W}\left(y^{\mathcal{SC}}, v^{\mathcal{SC}}\right) - W_S$, the solution to the operator's problem can be found by solving for the wealth constraints of the operator and the supplier. This means that the optimal solution must ensure that $W_O + W_S = \tilde{W}\left(y^*, v^*\right)$ and $2 + \pi + W_O = x\left(1-y^*-v^*\right) + x\Phi\left(y^*, v^*\right)\left(2y^* + v^* + \beta\left(1-2y^*-v^*\right)\right)$. We can make the same analysis for the case where $W_O \leq W_O^I$ to derive the remaining results in the proposition. ■

Proof of Lemma 9: Suppose the effort of the parties is sequential, and that the supplier leads. Under these assumptions, the best response level of effort for the operator, given e_S, is equal to $\tilde{e}_{O,seq}\left(e_S\right) = \arg\max_{e_{O\geq 0}} E_X\left[\Pi_O\left(e_O, w, y, v, P, X|e_S\right)\right] = \left(\frac{x\beta(1-y-v)}{(e_S+1)^{1-\beta}}\right)^{\frac{1}{\beta+1}} - 1$. As such, the optimal level of effort for the supplier is $e_{S,seq}^* = \arg\max_{e_S\geq 0} E_X\left[\Pi_S\left(e_S, w, y, P, X|\tilde{e}_{O,seq}\right)\right] = \sqrt{\frac{x(\beta(1-y-v))^{2-\beta}}{\left(\left(\frac{1-\beta}{1+\beta}\right)y\right)^{1-\beta}}} - 1$. By plugging $e_{S,seq}^*$ into the operator's effort function, we can verify that $e_{O,seq}^* = \tilde{e}_{O,seq}\left(e_{S,seq}^*\right) = \sqrt{\frac{x(\beta(1-y-v))^{2-\beta}}{\left(\left(\frac{1-\beta}{1+\beta}\right)y\right)^{1-\beta}}} - 1.$■

Proof of Proposition 10: Suppose that both the supplier and the operator have the option of purchasing insurance coverage, v_S and v_O. In this model, the supplier's and the operator's ex post profits are $\Pi_S\left(e_S, w, y, v_S, P_S, X\right) = T\left(w, y, X\right) - I_S\left(v_S, P_S, X\right) - e_S$ and $\Pi_O\left(e_O, w, y, v_O, P_O, X\right) = \pi - X - T\left(w, y, X\right) - I_O\left(v_O, P_O, X\right) - e_O$, where $I_S\left(v_S, P_S, X\right)$, $I_O\left(v_O, P_O, X\right)$ and $T\left(w, y, X\right)$ are defined as in Section 3. In this case, the best response level of effort for the supplier and operator are equal to $\tilde{e_S}|_{e_O} = \arg\max_{e_S \geq 0}\left\{w - \frac{(y-v_S)x}{(e_S+1)^{1-\beta}(e_O+1)^{\beta}} - e_S\right\} = \left(\frac{x(y-v_S)(1-\beta)}{(e_O+1)^{\beta}}\right)^{\frac{1}{2-\beta}} - 1$, and $\tilde{e_O}|_{e_S} = \arg\max_{e_O \geq 0}\left\{\pi - w - \frac{x(1-y-v_O)}{(e_S+1)^{1-\beta}(e_O+1)^{\beta}} - P - e_O\right\} = \left(\frac{x\beta(1-y-v_O)}{(e_S+1)^{1-\beta}}\right)^{\frac{1}{\beta+1}} - 1$. By solving for the correspondences of these functions, we get that $e_O^*\left(y, v_S, v_O\right) = \sqrt{\frac{x(\beta(1-y-v_O))^{2-\beta}}{((1-\beta)(y-v_S))^{1-\beta}}} - 1$ and $e_S^*\left(y, v_S, v_O\right) = \sqrt{\frac{x((y-v_S)(1-\beta))^{1+\beta}}{(\beta(1-y-v_O))^{\beta}}} - 1$. Now, note that the optimal level of coverage for the supplier are given by $v_S^* = \arg\max_{v_S \geq 0} E_X\left[\Pi_S\left(e_S^*, w, y, v_S, P_S, X\right)|e_O^*\right] = w - \frac{(y-v_S)x}{\left(e_O^*+1\right)^{\beta}\left(e_S^*+1\right)^{1-\beta}} - P\left(v_S|e_O^*, e_S^*\right) - e_S^*$. By plugging the expression of e_S^* and e_O^* in the expected profits of the supplier, and by performing basic optimization techniques, we can verify that $v_S^* = \frac{\beta y}{(1+\beta)}$. To prove part 2 of the proposition, we first need to write $e_S^*\left(y, \frac{\beta y}{1+\beta}, v_O\right) = \sqrt{\frac{x\left(\frac{\beta(1-\beta)}{1+\beta}y\right)^{1+\beta}}{(\beta(1-y-v_O))^{\beta}}}$ and $e_O^*\left(y, \frac{\beta y}{1+\beta}, v_O\right) = \sqrt{\frac{x(\beta(1-y-v_O))^{2-\beta}}{\left(\frac{\beta(1-\beta)}{1+\beta}y\right)^{1-\beta}}} - 1$. Using these results, we can directly verify that the ratio $\frac{e_O^*+1}{e_S^*+1}$ is identical to the ratio given in equation (3.3) ■

Appendix C

Proofs for Chapter 4

Proof of Lemma 11

First, note that, from equation (4.2) the stage-3 profits for type-τ firm are equal to $\Pi_i^{\tau_i}\left(\hat{r}_i^{\tau_i}|X,s,b\right) = \rho_X - s - m_\tau$. This allows us to re-write the stage-2 profits, which are given in equation (4.3), as $\tilde{\Pi}_i^{\tau_i}\left(q_i,q_j|s,b\right)$ $= q_i\left[q_j\rho_D + \left(1-q_j\right)\rho_M - m_{\tau_i} + s\right]$ and therefore, to re-write

$$\begin{aligned}\Pi_i^*\left(q_i^{\tau_i},q_j^H,q_j^L|s,b,p\right) &= q_i^{\tau_i}\left[\rho_D\left(pq_j^H+\left(1-p\right)q_j^L\right)\right]+\\ &\quad q_i^{\tau_i}\left[\rho_M\left(p\left(1-q_j^H\right)+\left(1-p\right)\left(1-q_j^L\right)\right)\right]+s-m_{\tau_i}\end{aligned}$$

Recall that the best response function for firm i, with type τ, is given by $B_i^\tau\left(q_j^H,q_j^L|s,b,p\right) = \arg\max_{q_i^{\tau_i}\in[0,1]}\left\{\Pi_i^*\left(q_i^{\tau_i},q_j^H,q_j^L|s,b,p\right)\right\}$.

A Bayesian Nash Equilibrium exists if and only if there is a type-contingent strategy $\hat{q}^\tau$ for $\tau = H, L$, such that $B_i^\tau\left(\hat{q}^H,\hat{q}^L|s,b,p\right) = \hat{q}^\tau$. In other words, to find a BNE, we need to find a set of fixed points $\hat{q}^H$ and $\hat{q}^L$ for the best response functions of high- and low-ability firms. By writing these two conditions explicitly, one can show that a BNE equilibrium exists if we can find $\hat{q}^H$ and $\hat{q}^L$, such that the following system is solved:

$$\hat{q}^H = \max\left\{\min\left\{\frac{\left(s+\rho_M-m_H\right)}{p\left(\rho_M-\rho_D\right)} - \frac{1-p}{p}\hat{q}^L, 1\right\}, 0\right\}$$

$$\hat{q}^L = \max\left\{\min\left\{\frac{s+\rho_M-m_L}{\left(1-p\right)\left(\rho_M-\rho_D\right)} - \frac{p}{1-p}\hat{q}^H, 1\right\}, 0\right\}$$

To find the BNE for all regions, we need to solve for these equations. We can find a solution by first finding the conditions under which $\left(\hat{q}^H,\hat{q}^L\right) = (1,1)$ and $\left(\hat{q}^H,\hat{q}^L\right) = (0,0)$. That is, we can verify that $\left(\hat{q}^L,\hat{q}^H\right) = (0,0)$ if and only if $s+\rho_M-m_H<0$. We can also verify that $\left(\hat{q}^L,\hat{q}^H\right) = (1,1)$ if and only if $s+\rho_D-m_L\geq 0$. We can solve the remaining regions by moving away from these extreme thresholds. ■

Proof of Lemma 12

First, let α_H and α_L be arbitrarily close to 0. By looking at equation (4.2), jointly with equations (4.5) and (4.6), we can see that (for any (s,b)) the following two conditions are satisfied: (i) $\mathcal{W}^{\tau_i}(M,s,b) < \rho_M + c_M - d$ and (ii) $\mathcal{W}^{\tau_1,\tau_2}(D,s,b) < 2\rho_D + c_D - 2d$. Moreover, recall that by the assumptions stated at the end of §3, we have that $\rho_M + c_M - d < 0$ and $2\rho_D + c_D - 2d < 0$. This implies that, whenever α_H and α_L are very close to 0, both the monopoly and the duopoly welfare will be negative, regardless of the type of firm(s) that enter the market. In other words, it is socially optimal to have no firms operating in the market. As a result, the policy $(0,0)$ will maximize the expected social welfare. Moreover, we can also see that because $\rho_D < \rho_M < d$, then the firms will never profit from entering the market and, at equilibrium, both high- and low-ability firms will choose $q_i = 0$.

Second, let α_H and α_L be arbitrarily large. This implies that for *any* policy (s,b), all firms will choose $q_i = 1$. This is true because if $\alpha_\tau \to \infty$, then $m_\tau \to 0$ and, therefore, $\Pi_i^{\tau_i}(\hat{r}^\tau | M,s,b) > \Pi_i^{\tau_i}(\hat{r}^\tau | D,s,b) = \rho_D + s - m_\tau > 0$. In other words, the firms will always profit from entering to the market. The result follows immediately.

Third, let α_H be arbitrarily large, but α_L be arbitrarily close to 0. By the argument above, the high ability firms will always choose $q_i = 1$. Also, because α_L is arbitrarily close to 0, then a low-ability (if it enters) will cause an accident with probability arbitrarily close to 1. By the assumptions stated at the end of §3, then the social planner will want to induce market exit by low-ability firms. Hence, by Lemma 1, the optimal policy (s^*, b^*) must be such that

$$s^* + p\rho_D + (1-p)\rho_D - (1-b^*)d + \frac{\alpha_L\left((1-b^*)d\right)^2}{2} \leq 0 \qquad \text{(L1.1)}$$

i.e., such that low-ability firms will choose $q_i = 0$. Otherwise, the social planner could always decrease the level of incentives to satisfy equality in (L1.1). By doing this, he would decrease the entry probability of low-ability firms, without decreasing the entry decision of high-ability firms.

By the above arguments, there must exist α_1 and α_2 satisfying the properties of Lemma 2. ■

Proof of Proposition 13

To decrease the notational burden in the proof, we will simplify the notation. We let, for $\tau = H, L$

$$\begin{aligned}
\Phi_{\tau}(b) &= d\left(1-\alpha_{\tau}(1-b)d\right)+\frac{\left(\alpha_{\tau}(1-b)d\right)^{2}}{2\alpha_{\tau}} \\
\phi_{\tau}(b) &= \lambda\left(1-\alpha_{\tau}(1-b)d\right)b \\
\gamma(\kappa) &= 2\left(c_{M}-\frac{c_{D}(\kappa)}{2}+(1+\lambda)\left(\rho_{M}-\rho_{D}(\kappa)\right)\right)
\end{aligned}$$

Here, $\Phi_{\tau}(b) > 0$ refers to the expected primary accident costs, as a function of ex post liability protection. $\phi_{\tau}(b) > 0$ refers to the opportunity cost of providing ex post subsidies b. $\gamma(\kappa) > 0$ is a measure of the expected welfare gains as a function of the competition parameter, κ.

1. Suppose that $\alpha_H \geq \alpha_1$ and $\alpha_L \leq \alpha_2$. By Lemma 2, we know that the optimal policy will be such that (s^*, b^*) will induce entry by (only) high-ability firms. By Lemma 1, therefore, we know that (s^*, b^*) is such that

$$s^{*}+p\rho_{D}+(1-p^{*})\rho_{D}-m_{L}\leq 0 \tag{P1.1}$$

where $m_L = (1-b^*)d + \frac{\alpha_L((1-b^*)d)^2}{2}$. Now, if we take the first order condition of $E[\mathcal{W}(s,b)]$ with respect to s, we will find that whenever (s^*, b^*) satisfies condition $(P1,1)$, then

$$p\hat{q}^{H}(s,b)=-\frac{\Phi_{H}-c_{M}-\rho_{M}+\lambda(s+\phi_{H})}{\gamma}$$

must be satisfied. Similarly, if we take the first order condition with respect to b, we will find that

$$p\hat{q}^{H}(s,b)=\left(\frac{\Phi_{H}+\lambda(s+\phi_{H})-p(c_{M}+\rho_{M})}{2c_{M}+2\rho_{M}-c_{D}-2\rho_{D}-(\rho_{M}-\rho_{D})\frac{\Phi'_{H}+\lambda\phi'}{-\Phi'_{H}+\phi'}}\right)$$

must also be satisfied. By combining these conditions, we find that the policy that satisfies the first order conditions must also satisfy $\lambda = \frac{\Phi'_H+\lambda\phi'_H}{-\Phi'_H+\phi'_H}$ or $\Phi'_H(\lambda+1) = 0$, where $\Phi'_H(b) = bd^2\alpha_H$. Hence, at the first order condition, it must be true that $b^* = 0$. Furthermore, if we solve for s (after setting $b = 0$) we find that the optimal level of ex ante subsidies satisfies $s^{\mathcal{H}} =$

$2\sqrt{d\alpha} - \rho_M + (\rho_M - \rho_D)\min\left\{\frac{\left(c_M+(\lambda+1)\left(\rho_M-d+\frac{\alpha_H d^2}{2}\right)\right)}{(2c_M-c_D+2(1+\lambda)(\rho_M-\rho_D))}, p\right\}$, where $\frac{\left(c_M+(\lambda+1)\left(\rho_M-d+\frac{\alpha_H d^2}{2}\right)\right)}{(2c_M-c_D+2(1+\lambda)(\rho_M-\rho_D))} = p$ whenever $\hat{q}^H(s^*, b^*) = 1$.

Now, to show that the first order condition is sufficient, we use the second order partial derivative test. Note that $\frac{\partial^2 E\left[\mathcal{W}\left(s^{\mathcal{H}},0\right)\right]}{\partial s^2} = \frac{2\gamma}{(\rho_M-\rho_D)^2} < 0$, $\frac{\partial^2 E\left[\mathcal{W}\left(s^{\mathcal{H}},0\right)\right]}{\partial s \partial b} = \frac{2\gamma(1-d\alpha_H)}{(\rho_M-\rho_D)^2}$ and $\frac{\partial^2 E\left[\mathcal{W}\left(s^{\mathcal{H}},0\right)\right]}{\partial b^2} = \frac{(1-d\alpha_H)\left(-(\gamma+(\rho_M-\rho_D)(1+\lambda))s^{\mathcal{H}}+(\rho_M-\rho_D)c_M-\rho_M\gamma\right)}{(\rho_M-\rho_D)^2}$. Hence, the determinant of the Hessian matrix will be equal to $\frac{2\gamma d(1-d\alpha_H)(1+\lambda)}{(\rho_M-\rho_D)^2}\left(\frac{(c_M-(\lambda+1)\rho_M-d)}{\gamma}\right) > 0$. By the second order partial derivative test, we conclude that (s^*, b^*) globally maximizes $E\left[\mathcal{W}(s,b)\right]$.

2. Suppose that $\alpha_H \geq \alpha_1$ and $\alpha_L \geq \alpha_2$. By Lemma 2, we have that (s^*, b^*) induces entry by high- and low-ability firms. By Lemma 1, therefore, we have that the optimal policy will be one in have that

$$s^* + p\rho_D + (1-\rho_D)\rho_M - d(1-b^*) + \frac{\alpha_L(d(1-b^*))^2}{2} > 0 \quad \text{(P1.2)}$$

Now, if we take the first order condition with respect to s, we will be able to see that if $(P1.2)$ is satisfied, the optimal solution will also satisfy $\Phi_L + \lambda(s+\phi_L) - ((c_M+\rho_M)(1-2p) + 2(c_D+2\rho_D)) + (\rho_M-\rho_D)\lambda = \hat{q}^L(s,b)(1-p)\left[c_D + 2\rho_D - 2c_M - 2\rho_M - (\rho_M-\rho_D)\lambda\right]$. By using the explicit form of $\hat{q}^L(s,b)$, we will be able to see that the optimal level of ex ante subsidies is given by $s^* = s^{\mathcal{HL}} \equiv \left(1-b^{\mathcal{HL}}\right)d - \frac{\alpha_L\left(\left(1-b^{\mathcal{HL}}\right)d\right)^2}{2} - \rho_M + \frac{(\rho_M-\rho_D)\left[c_M+(1+\lambda)\left(\rho_M-d+\alpha_L\left(1-b^{\mathcal{HL}}\right)d^2\left(1-\frac{1-b^{\mathcal{HL}}}{2}\right)\right)\right]}{2(1+\lambda)(\rho_M-\rho_D(\kappa))+(2c_M-c_D(\kappa))}$. To verify that this first order condition is a maximum, note that $\frac{\partial^2 E[\mathcal{W}(s,b)|p]}{(\partial s)^2} = \left(\frac{-2\gamma}{(\rho_M-\rho_D)^2}\right) < 0$.

Now that we have solved for s^*, we proceed to solve for the optimal level of b. To do this, we can plug in $s^{\mathcal{HL}}$ in the welfare function, and see that $\mathcal{W}\left(s^{\mathcal{HL}}(b), b\right) \equiv \mathcal{W}^{\mathcal{HL}}(b) = \frac{\left(\left(c_M+(\lambda+1)\left(\rho_M-d+\alpha_L(1-b)d^2\left(1-\frac{1-b}{2}\right)\right)\right)^2\right)}{\gamma} + \frac{+bd^2(\alpha_H-\alpha_L)(b+2(\lambda+1)b)p\gamma}{\gamma}$. By taking the first order condition with respect to b, we will get that $\frac{\partial}{\partial b}W^{\mathcal{HL}}(b) = 0$ whenever b satisfies $\frac{p(\alpha_H-\alpha_L)}{\alpha_L(\lambda+1)}\left(1+\frac{\lambda(1-2(1-b))}{b}\right)=$ $-\frac{c_M+(\lambda+1)\left(\rho_M-d+\alpha_L(1-b)d^2\left(1-\frac{1-b}{2}\right)\right)}{\gamma}$We can re-write this equation as $\frac{A}{2}\left(1-b^{\mathcal{HL}}\right)^3-\frac{3A}{2}\left(1-b^{\mathcal{HL}}\right)^2+(A-M-T)\left(1-b^{\mathcal{HL}}\right)+M=0$, where $A = d^2(1+\lambda)\alpha_L > 0$, $M = c_M+(1+\lambda)(\rho_M-d)+\frac{p\gamma(\alpha_H-\alpha_L)}{\alpha_L}$ and $T = \frac{p(\alpha_H-\alpha_L)\gamma\lambda}{\alpha_L(\lambda+1)} > 0$.Through the cubic formula, we find that the only real root that solves this equation is given at $b^{\mathcal{HL}} = \frac{1}{A}\sum_{i=1}^{2}\left[T+(-1)^i\left(T^2-\frac{\left(2(M+T)+A^3\right)^3}{27A}\right)^{\frac{1}{2}}\right]^{\frac{1}{3}} > 0$. Note that the inequality follows directly from the fact that $T>0$. To verify that this point is a maximum, we evaluate $\frac{\partial^2\mathcal{W}^{\mathcal{HL}}(b)}{\partial b^2}|_{b=b^{\mathcal{HL}}}$, and find that

$$
\begin{aligned}
&\frac{\partial^2\mathcal{W}^{\mathcal{HL}}\left(b^{\mathcal{HL}}\right)}{\partial b^2}\\
&=\frac{2d^2\alpha_L}{\gamma}\left(T-\left(\sum_{i=1,2}\left[T+(-1)^i\left(T^2-\frac{\left(2(M+T)+A^3\right)^3}{27A}\right)^{\frac{1}{2}}\right]^{\frac{1}{3}}\right)^3\right)\\
&<\frac{-2d^2\alpha_L T}{\gamma}\\
&<0
\end{aligned}
$$

where the first inequality follows from the triangle inequality for the p-norm for $p=\frac{1}{3}$.■

Proof of Lemma 14

One can verify, by plugging (s^*,b^*) into the optimal entry strategies, that

$$
\hat{q}^L(s^*,b^*) = \max\left\{\frac{\left(\frac{\left(c_M+(\lambda+1)\left(\rho_M-d+\alpha_L(1-b^*)d^2\left(1-\frac{(1-b^*)}{2}\right)\right)\right)}{(2c_M-c_D(\kappa)+2(1+\lambda)(\rho_M-\rho_D(\kappa)))}\right)-p}{1-p},0\right\}
$$

where $\hat{q}^L(s^*,b^*)>0$ iff the optimal policy lies in region $\mathcal{HL}$, and $\hat{q}^L(s^*,b^*)=0$ iff the optimal policy lies in region $\mathcal{H}$.

Note that in region $\mathcal{HL}$, the optimal level of ex post subsidies, $b^* = b^{\mathcal{HL}}$, is a function of κ, where

$$\frac{\partial b^*(\kappa)}{\partial \kappa} = -\frac{\frac{\partial \gamma(\kappa)}{\partial \kappa}\left(\alpha_L d^2\left(1-b^{*2}\right)+2\left(c_M+(\lambda+1)\left(\rho_M-d\right)\right)\right)}{2\gamma(\kappa)^2\left(\frac{d^2\alpha_L b^*}{\gamma(\kappa)}-\frac{p\lambda(\alpha_H-a_L)}{\lambda_1\alpha_L b^{*2}}\right)}$$

where γ is defined as in the beginning of Proposition 1. Hence, if we use the chain rule to differentiate $\hat{q}^L$ with respect to κ , we will find out that

$$\frac{\partial \hat{q}^L}{\partial \kappa} = -\frac{3d^2(1+\lambda)\alpha_L\frac{\partial\gamma(\kappa)}{\partial\kappa}}{2\gamma^2(1-p)}$$

where $\gamma'(\kappa) = -c_D'(\kappa) - 2(1+\lambda)(\rho_D'(\kappa))$. The result follows directly, by observing that $\frac{\partial \hat{q}^L}{\partial \kappa} \geq 0$ iff $\frac{\partial\gamma(\kappa)}{\partial\kappa} \leq 0$ and vice versa. ■

Proof of Proposition 15

This result follows directly from Lemma 3, by noting that the welfare gains, given more aggressive competition, are given by $\frac{c'(\kappa)}{2\rho_D'(\kappa)} - 1$. Hence, when $\frac{c'(\kappa)}{2\rho_D'(\kappa)} - 1$ is smaller than 0, $\hat{q}^L$ will decrease (and vice versa). Moreover, if $\hat{q}^L$ decreases to 0, then the optimal policy will lie in region $\mathcal{H}$, and the social planner will not offer ex ante subsidies. ■

Proof of Proposition 16

To show this result, we use an argument similar to the one we used to prove Lemma 2. First, we show that, in region $\mathcal{HL}$, $\frac{\partial b^*}{\partial d} = \frac{d\alpha_L\left(1-b^2\right)}{\gamma\left(\frac{d^2\alpha_L b}{\gamma}-\frac{p\lambda\left(\alpha_H-\alpha_L\right)}{\lambda_1\alpha_L b^2}\right)}$. Hence, we can use the chain rule (by noting that b^* is a function of d) to differentiate $\hat{q}^L$ with respect to d. By Lemma 1 and Proposition 1, we obtain that

$$\frac{\partial \hat{q}^L}{\partial d} = \max\left\{-\frac{d(\lambda+1)^2\alpha_L\left(1-b^2\right)\left(c_M+(\lambda+1)\rho_M\right)}{\gamma(1-p)\left(\alpha_H-\alpha_L\right)}, 0\right\}$$

From the above result, we will be able to see that $\hat{q}^L$ is decreasing in d. Hence, there must exist a d large enough so that, evaluated at such d, it must be true that $\hat{q}^L = 0$. By Lemma 2 and Proposition 1, the result will follow. ■

Proof of Proposition 17

To show this result, first note that (in region $\mathcal{HL}$) we have that $\frac{\partial b^*}{\partial \lambda} = \left(\frac{p(\alpha_H-\alpha_L)\left(1-b^2\right)\gamma(c_M+(1+\lambda)d)}{\alpha_L(\lambda+1)^2\gamma^2}\right) > 0$, $\frac{\partial s^*}{\partial \lambda} = -\frac{\left(c_M+(1+\lambda)\left(\rho_M+\frac{\alpha_H d^2}{2}\right)\right)^2(\alpha_H-\alpha_L)}{\gamma\alpha_L} < 0$. Now, if plug the optimal policy into the equilibrium entry strategies, we find that $\frac{\partial \hat{q}^L(s^*(\lambda),b^*(\lambda))}{\partial \lambda} = -\frac{c_D\alpha_L d^2+\gamma+(\rho_M-\rho_D)d}{2(1-p)\gamma^2} < 0$. This means that the level of ex ante subsidies will be decreasing as λ increases. Conversely, we see that the level of ex post subsidies will increase as λ increases. However, when λ moves beyond some threshold, then $\hat{q}^L$ will decrease to the point where the low-ability firms will exit the market. Here, the level of ex post subsidies will decrease to 0, and s^* will eventually converge to $s^{\mathcal{H}}$. ■