University of Alberta

Decision Making for Information Security Investments

by

Melanie Lisa Yeo

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Operations and Information Systems

Faculty of Business

Fall 2012
Edmonton, Alberta

Dedication

Thanks to my many supporters; friends, family and colleagues. But especially to Jonathan.

Abstract

Enterprises must manage their information risk as part of their larger operational risk management program. Traditionally, IT security investment decisions are made in isolation. However, as firms that compete for customers in an industry are closely interlinked, a macro perspective is needed in analyzing these decisions. Using the notions of direct- and cross-risk elasticity to describe the customer response to adverse IT security events in the firm and competitor, respectively, we analyze optimal security investment decisions. The continuous-time Markov chain (CTMC) is a natural way to examine how the combination of the expected adverse event arrival rate and the expected duration of customer reactions to these adverse events impacts security spending and expected profits, given different types of customer reaction. Expanding this work, customer utility is modelled using a Hotelling setting in order to examine how the introduction of minimum security spending requirements affect total social welfare. Optimal IT security spending, expected firm profits, total social welfare, and the willingness of firms to cooperate on security improvements are highly dependent on the nature of customer response to adverse events.

Once firms have made a decision regarding their security investment level, managers must consider how to implement information security controls. The effectiveness of three different control placement methods is examined by defining a flow risk reduction problem and presenting a formal model using a workflow framework. One year of simulated attacks is used to validate the quality of the solutions, finding that the math programming control placement method yields substantial improvements in terms of risk reduction and risk reduction on investment measures compared to heuristics that would typically be used by managers to solve the problem. By using a workflow approach to control placement, guiding the manager to examine the entire infrastructure in a holistic manner, this research is unique in that it enables information risk to be examined strategically.

The contribution of this body of work is to provide managers with methods for deciding on the level and selection of information security investments, obtaining significantly better returns on these security investments.

Acknowledgements

The work presented in Chapters 2 and 3 have been undertaken with Dr. Raymond Patterson and Dr. Bora Kolfal, both at the University of Alberta School of Business. The work presented in Chapter 4 has been undertaken with Dr. Raymond Patterson (University of Alberta), Dr. Jackie Rees-Ulmer (Purdue University), and Dr. Erik Rolland (University of California, Merced). I am grateful for their contributions to this work. Use of the term "we" in this thesis anticipates joint submission to target journals.

I would like the thank my co-supervisors, Dr. Raymond Patterson and Dr. Kenneth L. Schultz, for their guidance.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis presents three separate essays on information risk management decision making. They include: (a) an examination of optimal security spending in a duopoloy setting based on demand changes in reaction to security breaches, (b) exploration of the impact of minimum security spending on firm profits, consumer surplus and total social welfare, and (c) the potential risk reduction afforded by optimal placement of security controls within a workflow.

The average cost of an information security breach was estimated to be $7.2 US million in 2010 (Ponemon and Symantec 2011). This cost, combined with increased threats from highly-motivated attackers (Schwartz 2012) and pressure to disclose breaches (Lardner 2012), suggest that firms need to find the most effective ways to manage the risk associated with information systems. The ISO 31000 risk management standard talks of seven key elements in the risk management process: establishing the context of risks, identifying risks, assessing risks, evaluating potential treatments, creating a risk management plan, implementing the risk management plan, and evaluating the plan. Rolland et al. (2011), in placing information risk management research into context, use the ISO 31000 standard combined with identifying the focus as either on markets, policy and frameworks or decision making and control. They suggest that much of the information security research falls into evaluating potential risk treatments (that is, examines ways to avoid, mitigate, transfer or accept risks). In particular, there is much market, policy and framework research on the role of insurance for risk transfer with the general findings that while insurance would be beneficial to both firms and society (Srinidhi et al. 2008, Kesan et al. 2005), although it is under-utilized for a variety of reasons (Gordon et al. 2003, Bandyopadhyay et al. 2009, Bohme 2005, Ogut et al. 2005). On the decision-making and control side, Bai et al (2007) discuss the need to incorporate risk mitigation strategies when designing business processes, a theme picked up in Chapter 4. Chapters 2 and 3 help establish the context of risk by examining market forces with an eye to policy decisions, an area in the literature that appears to be less well covered. Gordon and Loeb (2002) start down this path with their work on the economics of information security, finding that there is a limit to investment in security.

This collection of essays explores formal methods for determining information security investments and implementations in order to help businesses make more

informed choices about information risk management. Chapter 2 uses a continuous-time Markov chain (CTMC) to examine how customer reactions to adverse events such as information security breaches impact security spending in a duopoly. This work demonstrates that spending is affected by both the riskiness of the environment (a function of security breach arrival rates and the duration of demand changes as a result of these breaches) as well as the nature of customer reactions. In essence, market forces can be used to help firms make better decisions about information security investment levels. Chapter 3 expands on this initial model by including customer utility and examinating the effect on total social welfare of a minimum security investment above the market equilibrium identified in Chapter 2.

Chapter 4 moves away from the investment decision and looks at how to best allocate that investment. This chapter examines the placement of controls to minimize the damage of successful attacks by defining the flow risk reduction problem and presenting a formal model using a workflow framework. Three different control placement methods are introduced to solve the problem, and a comparative analysis is presented using a robust test set of 162 simulations. One year of simulated attacks is used to validate the quality of the solutions. The math programming control placement method yields substantial improvements in terms of risk reduction and risk reduction on investment when compared to heuristics that would typically be used by managers to solve the problem.

Bibliography

Bandyopadhyay, Tridib, Vijay S. Mookerjee, Ram C. Rao. 2009. Why IT managers don't go for cyber-insurance products. *Commun. ACM* **52**(11) 68–73.

Bohme, Rainer. 2005. Cyber-Insurance revisited. *Workshop on the Economics of Information Security (WEIS)* .

Gordon, L. A., M. P. Loeb. 2002. The economics of information security investment. *ACM Transactions on Information and System Security* **5** 438–457.

Gordon, Lawrence A., Martin P. Loeb, Tashfeen Sohail. 2003. A framework for using insurance for cyber-risk management. *Commun. ACM* **46**(3) 81–85.

Kesan, J, R.P. Majuca, W.J. Yurcik. 2005. Cyberinsurance as a Market-Based solution to the problem of cybersecurity - a case study. *Workshop on the Economics of Information Security (WEIS)* .

Lardner, Richard. 2012. U.S. pressures companies to report cybercrime. *Associated Press* URL `http://www.usatoday.com/money/media/story/2012-06-29/reporting-cybercrime/55921858/1`.

Ogut, H., M. Menon, S. Raghunathan. 2005. Cyber insurance and IT security investment: Impact of interdependent risk. *Workshop on the Economics of Information Security (WEIS)* .

Ponemon, Symantec. 2011. 2010 annual study: U.S. cost of data breach .

Rolland, E., R. Patterson, J. Rees Ulmer, M.L. Yeo. 2011. Risk mitigation decisions for it security. Presentation.

Schwartz, Matthew J. 2012. Advanced persistent threats get more respect. *InformationWeek* URL `http://www.informationweek.com/news/security/cybercrime/232600562`.

Srinidhi, Bin, Jia Yan, Giri Kumar Tayi. 2008. Firm-level resource allocation to information security in the presence of financial distress. Working Papers 2008-17, School of Economic Sciences, Washington State University. URL `http://ideas.repec.org/p/wsu/wpaper/yan-1.html`.

Chapter 2

Market Impact on IT Security Spending

A version of this chapter has been submitted for review.

2.1. Introduction

IT security spending and risk at one firm impact other firms through demand changes as customers react to security incidents. Lost business as a result of security breaches is estimated to account for 62% of the $7.2 US million average cost of an IT security breach (Ponemon and Symantec 2011), and a study by Javelin (2007) found that 77% of respondents intend to stop doing business with firms who have experienced data breaches. With laws such as California's SB 1386, HIPAA, and DPA in the UK requiring that consumers and authorities be notified of data breaches in a timely manner (SB-1386 2002, HIPAA 1996, DPA 1998), security breaches become public knowledge. Such an adverse IT security event for the firm results in both direct effects on the firm and indirect effects on the firm's competitors which last for some duration. We discuss the concepts of direct- and cross-risk elasticities of demand to capture these direct and indirect demand effects as we examine how the firm's optimal spending is affected by: (i) spending by competitors to prevent adverse events in their respective firms, (ii) the internalization of adverse events into the customer demand, both for the firm and its competitors, (iii) the arrival rate and duration of adverse events, and (iv) industry standardization or regulation.

Firms can be viewed as being substitutes in loss or complements in loss to indicate the effects of an adverse event on the unaffected firm. Examples of such customer reactions can be found in many industries; toy and food safety issues can impact related products (Freedman et al. 2009, Smed and Jensen 2005), recent news that location data is stored on both the Apple iPhone and 3GS iPad (Allan and Warden 2011, Arthur 2011a) may lead privacy-sensitive consumers to consider other alternatives, or security failures in the commercial banking (Acohido 2009, Krebs 2009a,b) and online gaming industries (Baker and Finkle 2011, Arthur 2011b) have the potential to result in customer demand changes. In light of customer reactions, our results would allow managers to answer questions such as "How are our IT security spending and profits affected by customer reaction to events in our own firm and at competitors?" and "Under what conditions should we work within our industry to increase security spending beyond market equilibrium?"

The continuous-time Markov chain (CTMC) model is a natural way to examine how the combination of the expected adverse event arrival rate and the expected du-

ration of customer reactions to these adverse events impacts security spending and expected profits, given different types of customer reaction. We call this combination of expected arrival rate and expected duration the *riskiness of the environment*. Using a CTMC, we demonstrate that security spending is affected by both the riskiness of the environment and the customer reactions to these adverse events. We begin with an analytical examination of the symmetric case before extending our work with numerical analyses for both correlated arrivals of adverse events and the asymmetric case where customer reactions may differ.

Customer demand is affected by concerns for the safety of a product (Conz 2008, Crawford 2008, Smed and Jensen 2005). Duh et al. (2002) discuss the need for control (which they define as "expectational equilibrium between what participants do and what others expect of them") in eCommerce as a necessary component for building customer trust in order to increase business activity. They also examine how third party assurance services such as TRUSTe and BBB Online can help with risk mitigation. Thus, customer demand is affected by industry security standards. The topic of cyber security regulation is a significant issue, but most of the legislation currently under consideration in the United States has to do with security of federal government systems and homeland security (Coyle 2011). What is not under consideration is legislation over security of commercial systems. ISO 27001 specifies best practices for an information security management system (IT Governance 2012). Some organizations are required to comply with industry standards relating to data protection; Sarbanes Oxley, PCI DSS, and ISO 27001 ensure that most of the necessary processes for compliance are in place (Sysnet 2012a,b). We are able to show where firms are able to benefit from industry standards or government regulation.

In one stream of literature, actual risks of an adverse event happening to a firm are dependent on the security at another firm. Examples of such interdependencies between firms, where the success of one firm is dependent on the effort of another firm, include the airline industry, fire protection, vaccinations and even bankruptcy in financial firms (Kunreuther and Heal 2003, Varian 2004). In another stream of literature, the perceived risk differs from the actual risk. Yu and Lester (2008) describe how adverse events can negatively impact the reputation of not only the affected firm, but may spill over onto other industry participants when they are

perceived to be "using the same or similar technologies (p. 95/96)." Yu and Lester (2008) identify "proximity and structural equivalence as the influential drivers of social contagion process (p. 98)." Social contagion, as described by Yu and Lester (2008), is this spill-over effect and lost reputation can last for several years (Burke 2011). Incidents that harm a firm's reputation include data theft and cyber-attacks Burke (2011), with lost reputation one very clear way in which an adverse IT security event can change customer demand (Burke 2011). From a managerial standpoint, clearly reputation influences customer demand and IT security breaches impact the firm's reputation. We do not address reputation specifically except to say that, for some reason, demand changes when an IT security breach occurs. However, Yu and Lester (2008) explore the reputational impacts of adverse events of any kind (not just IT security), and thus our work can reasonably be extended.

Researchers have considered investment decisions for a single firm (Gordon and Loeb 2002), finding that it is not always the case that a firm should increase security investment as the vulnerability of the information being protected increases. When considering the interaction between competitors, each firm's choice to share information and invest in security interact in interesting ways. Several researchers show that there is a clear benefit to firms in sharing their security strategy as well as information about the success of that strategy in warding off attacks (Cavusoglu et al. 2008, Gal-Or and Ghose 2005, Gordon et al. 2003). Costs change when information sharing occurs between firms (Gal-Or and Ghose 2005, Gordon et al. 2003), but the nature of those changes depends on how the cost of a breach is viewed. Gordon et al. (2003) consider a fixed loss for the breach of a specific information set and show that when two firms coordinate by sharing information, it is possible that they can achieve the same level of security they had prior to information sharing, but at a reduced cost. We incorporate costs of information sharing into the per-unit cost of security.

Prior literature has examined cases where no firm interaction was considered (single firm) (Gordon and Loeb 2002), where firm interaction, but not spill-over effects of customer demand changes, was considered (Gordon et al. 2003), and where firm interaction with spill-over effects with some customers switching to the unaffected firm (Gal-Or and Ghose 2005, Cezar et al. 2010) were considered. Cezar et al. (2010) consider correlated breach events in a periodic model to explore outsourcing

the security function. The model presented in Cezar et al. (2010) captures the effects of cross-risk elasticity of demand in a way that is different than our model by incorporating demand spill-over and pricing changes, which also change demand, in reaction to the security breach. In their model, the security spending budget is fixed. In contrast, our model focuses on how security spending has an effect on the security quality and eventual changes in demand.

Another important difference between our paper and previous work is that we are using a CTMC setting which enables us to analyze the effects of the duration of successful attacks on important outcomes such as demand, equilibrium spending, and firm profits in a simple way. In a related work, Yue and Çakanyildirim (2007) explore incident response strategies to manage the expected duration of adverse events. We also extend our initial CTMC model to take correlated arrivals into consideration before examining the asymmetric case where customer reactions may be different for the firms.

The remainder of this paper examines this cross-risk elasticity of demand and how it may be used to explain several phenomena that we observe. The paper proceeds as follows. Section 2.2 presents a model for the two-firm case; analytical results are presented in section 2.3; Section 2.4 presents discussion and analysis of model extensions; and finally, conclusions are presented in Section 2.5.

2.2. Model

In this section, we present the details of our model. We examine how the firm's optimal spending is affected by: (i) spending by competitors to prevent adverse events in their respective firms, (ii) the internalization of adverse events into the customer demand, both for the firm and its competitors, (iii) the arrival rate and duration of adverse events, and (iv) industry standardization or regulation. We use a CTMC process to model the evolution of the state in which firms operate. It should be noted that the state change process of our CTMC is a discrete-time Markov chain (DTMC), which is a multi-period model.

We consider a game theoretical approach with two profit maximizing firms (duopoly) – firm 1 and firm 2. In our model, firms decide their own IT security spending level and with this they can alter the frequency of experiencing an adverse event, to a

point. In this context, let:

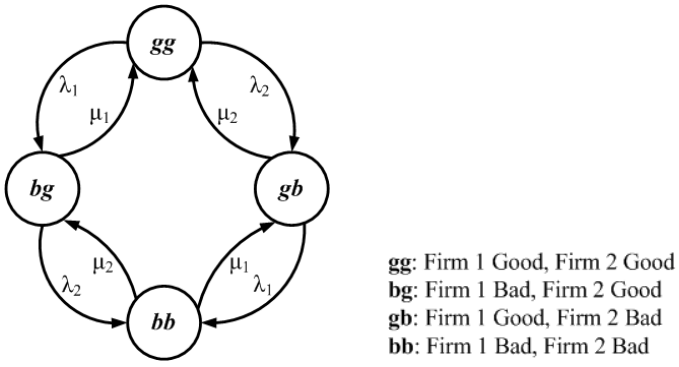
$$c_i = \text{Firm } i\text{'s per product spending on IT security, where } i = 1, 2.$$

In our model, security related attacks follow a Poisson process and they happen with rate Λ_i for firm $i = 1, 2$, independent of the attacks against the other firm. We call this arrival rate, Λ_i, the *base arrival rate*. However, not all attacks are "successful"; in order for an attack to be deemed "successful" the attack must both compromise security and become public knowledge. That is, if customers are not aware of the security breach, then it is not a successful attack for our modeling purposes. We let λ_i for firm $i = 1, 2$ be the arrival rate of successful attacks against firm i. A successful attack against firm i results in an adverse event where the demand for that firm's product is reduced. For the remainder of the paper, we use the terms *successful security related attack* and *adverse event* interchangeably. To a point, each firm can alter the arrival rate of adverse events, λ_i, by adjusting its security spending level, c_i; that is, $\lambda_i = f(c_i)$ where $\partial f/\partial c_i \leq 0$ and $\partial^2 f/\partial c_i^2 \geq 0$. We define a specific form for this relationship in section 2.2.1. Every time firm $i = 1, 2$ experiences an adverse event, the effects last for a random duration of time that has an exponential distribution with expected length $1/\mu_i$. In Section 2.2.2, we detail the effects of adverse events to firms.

Let $\{S(t), t \geq 0\}$ be the state process for the firms with $S(t) = (S_1(t), S_2(t))$ denoting the state of each firm at time t. At any point in time, each firm can either be in a "bad" state, where the firm is still under the effects of an adverse event it has experienced, or in a "good" state, where the firm is no longer under the negative effects of its most recent adverse event. We adopt the notation $S_i(t) \in \{g, b\}$ for each firm $i = 1, 2$, where g denotes a *good* state and b denotes a *bad* state. As we have two firms, we have four possible states: $S(t) = \{gg, gb, bg, bb\}$ (Note that we simplify notation by using gg instead of (g, g), etc.). The possible states are shown in Figure 2.1. Besides the states, Figure 2.1 also displays all the possible transitions between states and corresponding transition rates.

As shown in Figure 2.1, in state gg none of the firms are experiencing an adverse event. From the state gg, an adverse event happening for firm 1 (firm 2) moves the system to the state bg (gb), where firm 1 (firm 2) is in a bad state and the other firm is in a good state. If the system is at either of these two bad event states, bg

Figure 2.1: State Diagram.



or *gb*, and an adverse event occurs for the other firm, then the system will move to the state *bb*, where both firms are simultaneously in a bad state.

2.2.1 State Probabilities

To find the equilibrium spending level that maximizes the long-run expected firm profits, we obtain the steady-state probabilities of being in each state shown in Figure 2.1. Let P_s denote the steady-state probability of being in state $s \in S = \{gg, gb, bg, bb\}$. With this notation, P_{gg} is the probability that both firms are in a *good* state, P_{bg} (P_{gb}) is the probability that firm 1 (firm 2) is in a *bad* state and the other firm is in a *good* state, and lastly P_{bb} is the probability is that both firms are in a *bad* state.

As we have modeled the system as a CTMC, the steady-state probabilities, P_s, for the system in Figure 2.1 can be obtained from the following equations:

$$P_{gg}(\lambda_1 + \lambda_2) = P_{bg}\mu_1 + P_{gb}\mu_2 \tag{2.1}$$

$$P_{bg}(\lambda_2 + \mu_1) = P_{bb}\mu_2 + P_{gg}\lambda_1 \tag{2.2}$$

$$P_{bb}(\mu_1 + \mu_2) = P_{bg}\lambda_2 + P_{gb}\lambda_1 \tag{2.3}$$

$$P_{gb}(\mu_2 + \lambda_1) = P_{gg}\lambda_2 + P_{bb}\mu_1 \tag{2.4}$$

$$\sum_{P_i, i \in S} P_i = P_{gg} + P_{bg} + P_{gb} + P_{bb} = 1 \tag{2.5}$$

where (2.1)-(2.4) are state balance equations and (2.5) normalizes the probabilities to 1.

In our model, the arrival rate of successful adverse events, λ_i, for the firm $i = 1, 2$ is a function of the base arrival rate for the firm, Λ_i, and the security spending

c_i; in particular, we let $\lambda_i = \Lambda_i/c_i$. As our focus in this paper is not risk-free environments, we are only interested in cases where $\Lambda_i > 0$. This functional form provides two important properties: (i) the arrival rate of the successful adverse events is decreasing in security spending, and (ii) the returns to security spending are decreasing. In section 2.4.1 we show that our main results are not dependent on this specific functional form, $\lambda_i = \Lambda_i/c_i$, by repeating the analysis for alternative functions including the ones used in Gordon and Loeb (2002).

From (2.1)-(2.5), we obtain the steady-state probabilities as follows:

$$\begin{aligned} P_{gg} &= \frac{\mu_1\mu_2 c_1 c_2}{(\mu_1 c_1 + \Lambda_1)(\mu_2 c_2 + \Lambda_2)}, \qquad P_{bb} = \frac{\Lambda_1\Lambda_2}{(\mu_1 c_1 + \Lambda_1)(\mu_2 c_2 + \Lambda_2)} \\ P_{bg} &= \frac{\mu_2 c_2 \Lambda_1}{(\mu_1 c_1 + \Lambda_1)(\mu_2 c_2 + \Lambda_2)}, \qquad P_{gb} = \frac{\mu_1 c_1 \Lambda_2}{(\mu_1 c_1 + \Lambda_1)(\mu_2 c_2 + \Lambda_2)} \end{aligned} \tag{2.6}$$

Let ρ_i, which we call the *riskiness of the environment*, be the base arrival rate times the expected duration of an adverse event for the firm:

$$\rho_i = \frac{\Lambda_i}{\mu_i}, \qquad \text{for firm } i = 1, 2. \tag{2.7}$$

In essence, the base arrival rate can be thought of as representing the attractiveness of the firm, or industry, to attackers. A defence contractor (or the defence industry as a whole) would conceivably have a higher base arrival rate than, say, a firm that manufactures paper. How long firms feel the effects of demand changes due to successful adverse events depends greatly on how their customers view the firm or industry. In all, though, if either the base arrival rate is higher or the expected duration of an event is higher, then the firms are operating in an environment with higher "risk." It is this fact that leads us to refer to ρ as the riskiness of the environment.

It can be seen from (2.7) that for a higher base arrival rate, Λ_i, the environment is more risky and ρ_i will be higher. Likewise, for a longer expected duration the environment is also riskier and thus ρ_i is higher. To summarize, ρ provides an indication of how dangerous the firms' operating environment is.

Using the relationship (2.7), the probabilities in (2.6) simplify to

$$\begin{aligned} P_{gg} &= \frac{c_1 c_2}{(c_1 + \rho_1)(c_2 + \rho_2)}, \qquad P_{bb} = \frac{\rho_1\rho_2}{(c_1 + \rho_1)(c_2 + \rho_2)} \\ P_{bg} &= \frac{\rho_1 c_2}{(c_1 + \rho_1)(c_2 + \rho_2)}, \qquad P_{gb} = \frac{\rho_2 c_1}{(c_1 + \rho_1)(c_2 + \rho_2)} \end{aligned} \tag{2.8}$$

2.2.2 Modeling Demand

When a firm experiences an adverse event, consumers may change their purchasing decisions for a period of time. In our model, the expected duration of this changed consumer behavior, $1/\mu$, is not the same as the duration of the attack itself. A firm may recover from the technical aspects of an attack before (or even after) the consumers return to their original purchasing behavior.

Let $D_{i,s}$ denote the demand rate for the product of firm $i = 1, 2$ in state $s \in \{gg, bg, bb, gb\}$. Also, let $Q_i > 0$ denote the demand rate for the product of firm i in the absence of an adverse event, i.e., when both firms are in a good state, $D_{i,gg} = Q_i$. When firm i experiences an adverse event, the demand for its product will decrease by β_{i1}, e.g., $D_{1,bg} = Q_1 - \beta_{11}$. When firm j experiences an adverse event, the demand for the product of firm i ($i \neq j$) will change by β_{i2}, e.g., $D_{1,gb} = Q_1 - \beta_{12}$. Note that, $D_{1,gb}$ ($D_{2,bg}$) may be greater than Q_1 (Q_2), if the firm gains demand when the competitor faces an adverse event. Finally, when both firms experience adverse events, firm i's demand will change by both β_{i1} and β_{i2}, resulting in $D_{i,bb} = Q_i - \beta_{i1} - \beta_{i2}$. To summarize, we have the following:

$$\begin{aligned} D_{i,gg} &= Q_i, & D_{i,bb} &= Q_i - \beta_{i1} - \beta_{i2}, \quad \text{for } i = 1, 2 \\ D_{1,bg} &= Q_1 - \beta_{11}, & D_{1,gb} &= Q_1 - \beta_{12}, \\ D_{2,gb} &= Q_2 - \beta_{21}, & D_{2,bg} &= Q_2 - \beta_{22}. \end{aligned} \tag{2.9}$$

Let $X_i(t)$, be a binary variable indicating an adverse event in firm $i = 1, 2$ at time t:

$$X_i(t) = \begin{cases} 1 & \text{Firm } i = 1, 2 \text{ is under the effects of an adverse event at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

Then, the demand for firm i's product at time t can be expressed as follows:

$$D_i(t) = Q_i - \beta_{i1} X_i(t) - \beta_{i2} X_j(t) \qquad \text{for } i = 1, 2 \text{ and } j \neq i. \tag{2.10}$$

The cross-effect term, β_{i2}, allows the model to capture the alternative ways that firm $i = 1, 2$ may be affected by an adverse event in the other firm j ($j \neq i$). During the time firm j is under the effects of an adverse event, firm i may gain market share by picking up demand, if some of the customers switch from firm j to firm i. This case happens when $\beta_{i2} < 0$. Alternatively, when $\beta_{i2} > 0$, firm i loses market share temporarily as customers either find substitutes for the product offered by both firm

i and j or choose not to buy the product due to the adverse event. Lastly, when $\beta_{i2} = 0$, firm i's demand is unaffected by adverse events at firm j.

We now focus on the symmetric duopoly case. Results for asymmetric cases are discussed in section 2.4.3. In the symmetric setting, we can drop the index i from Q_i, β_{i1}, β_{i2} and ρ_i. Next, we normalize the demand functions by dividing (2.9) by Q. The normalized demand functions for firm i in the symmetric setting are provided below:

$$\begin{aligned} &D_{i,gg} = 1, \qquad D_{i,bb} = 1 - Z_1 - Z_2, \qquad \text{for } i = 1, 2 \\ &D_{1,bg} = D_{2,gb} = 1 - Z_1, \qquad D_{1,gb} = D_{2,bg} = 1 - Z_2. \end{aligned} \tag{2.11}$$

where $Z_1 = \frac{\beta_1}{Q}$ is the percentage change in demand due to an adverse event in own firm and $Z_2 = \frac{\beta_2}{Q}$ is the percentage change in demand due to an adverse event in the other firm. In this setting, Z_1 is the *direct-risk elasticity of demand* and Z_2 is the *cross-risk elasticity of demand*. The normalized version of the demand function in (2.10) is:

$$D_i(t) = 1 - Z_1 X_i(t) - Z_2 X_j(t) \qquad \text{for } i = 1, 2 \text{ and } j \neq i.$$

In our model, demand cannot be negative and a firm cannot gain demand from its own adverse event. Therefore, we are interested in cases where $Z_1 \in [0, 1]$ (corresponding to $\beta_1 \in [0, Q]$). Likewise, the firm cannot gain more than the other firm's demand or lose more than its own demand when other firm has an adverse event and thus we are interested in cases where $Z_2 \in [-1, 1]$ (corresponding to $\beta_2 \in [-Q, Q]$). We only consider the cases with $Z_1 + Z_2 \leq 1$, which ensures that $D_{i,bb} \geq 0$ holds. Also, we focus on the parameter region where $D_{i,bb} \leq 1$, since we do not find it realistic to think that demand for firm i will be greater after it has experienced an adverse event, even if it takes demand from the other firm. Hence, in this paper, we analyze the cases with $Z_1 + Z_2 \geq 0$, which ensures that $D_{i,bb} \leq D_{i,gg} = 1$ holds. Lastly, we are interested in the cases where the impact of events at a firm is greater than the impact from events at the competitor. Therefore, we consider the cases with $Z_1 \geq Z_2$. Our assumptions are listed below:

Assumption 2.1. $0 \leq Z_1 + Z_2 \leq 1$

Assumption 2.2. $Z_2 \leq Z_1$

2.3. Analytical Results

We continue our analysis by deriving the best response functions and the equilibrium security spending for the firms. In our model, each firm maximizes its long-run average profit. All firms are price takers; price and gross profit margin per unit excluding security spending, are fixed. We hereafter refer to gross profit margin per unit excluding security spending simply as per unit profit. The long-run average profit for firm i, $E[\Pi_i]$, is given as:

$$E[\Pi_i] \;=\; lim_{t\to\infty}\, E\Big[\frac{1}{T}\int_0^T D_i(t)(\pi - c_i)Q dt\Big] \tag{2.12}$$

where $\pi > 0$ denotes the per unit profit, excluding the security spending, c_i. In this paper we require $\pi - c_i > 0$ as a long-run participation constraint, as otherwise the firm would not be willing to produce the product. By using $D_i(t) = 1 - Z_1 X_i(t) - Z_2 X_j(t)$, we can write (2.12) in its steady-state form as follows,

$$\begin{aligned} E[\Pi_i] &= E[D_i](\pi - c_i)Q \;=\; \sum_{s\in S} P_s D_{i,s}(\pi - c_i)Q \\ E[\Pi_i] &= (P_{gg}D_{i,gg} + P_{bg}D_{i,bg} + P_{gb}D_{i,gb} + P_{bb}D_{i,bb})(\pi - c_i)Q \end{aligned} \tag{2.13}$$

where, $E[D_i]$ is the expected demand rate for firm i. By substituting the probabilities in (2.8) and demand functions in (2.11) into (2.13), we obtain the firm's profit as:

$$E[\Pi_i] = \left(1 - \frac{Z_1\rho}{c_i + \rho} - \frac{Z_2\rho}{c_j + \rho}\right)(\pi - c_i)Q \qquad \text{for } i = 1, 2, \text{ and } j \neq i. \tag{2.14}$$

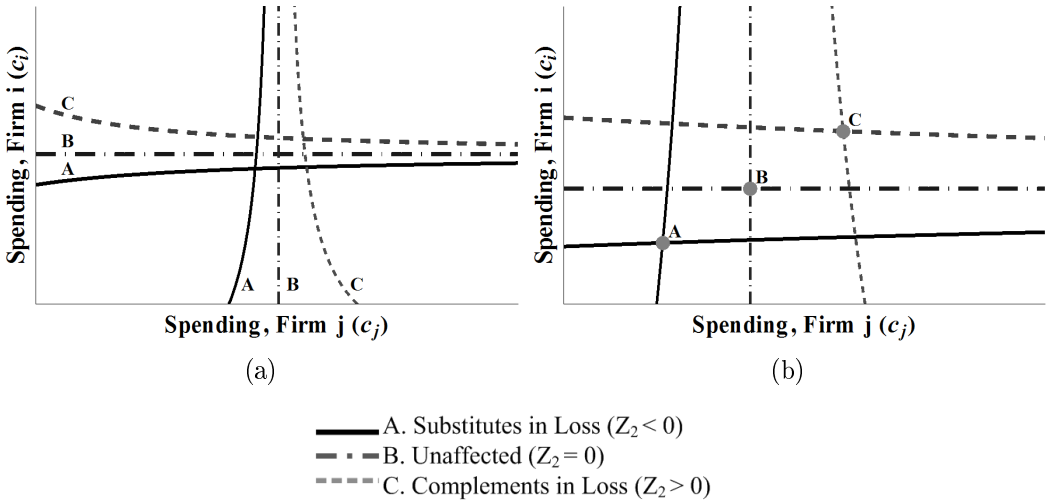
From the first-order conditions, we find the best response of firm i, c_i^*, below:

$$c_i^* = \rho\left(\sqrt{\frac{Z_1(\pi + \rho)(c_j + \rho)}{\rho(c_j + \rho - Z_2\rho)}} - 1\right) \qquad \text{for } i = 1, 2, \text{ and } j \neq i. \tag{2.15}$$

2.3.1 Best Response Spending Curves

An intuitive way to discuss the effects of competition on the security spending levels is to examine the different possible response curves. In this section, i denotes the firm under consideration and j denotes the other firm. Figure 2.2a shows the response curves for optimal spending for firm i under different effects of adverse events on firm j. Different curves in Figure 2.2a correspond to different values of Z_2 in the best response function provided in (2.15).

Figure 2.2: Best response curves



Response curves are generated with the parameters $Z_1 = .3, \pi = 100, \rho = 5$ and $Z_2 = \{-0.3, 0, 0.3\}$.

As the partial derivative of c_i^* with respect to c_j (shown below in (2.16)) demonstrates, when the cross-risk elasticity of demand, Z_2, is less than 0, firm i should increase spending as firm j increases spending. This is because Z_1 is non-negative and π and ρ are positive in our model. Conversely, when $Z_2 > 0$, firm i should decrease spending as firm j increases spending.

$$\frac{\partial c_i^*}{\partial c_j} = -\frac{Z_1 Z_2 \rho(\pi+\rho)}{2\sqrt{\frac{Z_1(c_j+\rho)(\pi+\rho)}{\rho(c_j+\rho-Z_2\rho)}}(c_j+\rho-Z_2\rho)^2} \qquad \text{for } i = 1,2, \text{ and } j \neq i. \tag{2.16}$$

When the cross-risk elasticity, Z_2, is such that there is no change in consumer demand for firm i's products when firm j experiences an adverse event, we call this "unaffected" ($Z_2 = 0$). In this case, as firm j spends more on security, firm i's optimal security spending does not change (see horizontal line B in Figure 2.2a and equation (2.16)). Similarly, the vertical line B in Figure 2.2a illustrates firm j's optimal security spending when the cross-risk elasticity, Z_2, is zero. Hence, when $Z_2 = 0$, regardless of the other firm's spending level, optimal spending for the firm is unchanging. As observed in (2.15), when $Z_2 = 0$, the optimal spending levels for each firm are independent of each other and $c_i^* = \rho(\sqrt{\frac{(\pi+\rho)}{\rho}Z_1} - 1)$ for $i = 1, 2$.

When the cross-risk elasticity is such that consumers increase demand for firm i's product when firm j experiences an adverse event, we call this is a "substitute in loss" ($Z_2 < 0$). As firm j spends more on security, firm i's optimal security spending also increases (see horizontal response curve A in Figure 2.2a and equation (2.16)).

Similarly, the vertical response curve A in Figure 2.2a illustrates firm j's optimal security spending when the cross-risk elasticity, Z_2, is smaller than zero. Optimal security spending is lowest under conditions of substitutes in loss because, all else being equal, c_i^* for both firms in (2.15) is increasing in Z_2, when $Z_1 > 0$. This can be observed from the partial derivative of c_i^* with respect to Z_2, which is provided below.

$$\frac{\partial c_i^*}{\partial Z_2} = \frac{\rho^2 \sqrt{Z_1 \frac{\pi+\rho}{\rho}(c_j+\rho)}}{2(c_j+(1-Z_2)\rho)^{3/2}} \qquad \text{for } i=1,2, \text{ and } j \neq i. \tag{2.17}$$

Hence, under conditions of substitutes in loss, optimal spending is always less than the unaffected case and eventually converges to the unaffected spending level, as the other firm's spending level goes to infinity.

When the cross-risk elasticity is such that consumers decrease demand for firm i's products when firm j experiences an adverse event, we call this is a "complement in loss" ($Z_2 > 0$). As firm j spends more on security, firm i's optimal security spending decreases (see horizontal response curve C in Figure 2.2a and equation (2.16)). Similarly, the vertical response curve C in Figure 2.2a illustrates firm j's optimal security spending when the cross-risk elasticity, Z_2, is greater than zero. As (2.15) and (2.17) suggest, under conditions of complements in loss, optimal spending is always greater than the unaffected case and eventually converges to the unaffected spending level as the other firm's spending level goes to infinity.

2.3.2 Equilibrium Spending

We can find the equilibrium spending by replacing competitor's spending, c_j, in (2.15) by c_i and solving the resulting equation for c_i. Due to symmetry, the equilibrium spending (denoted as c^e) will be the same for both firms. We are only interested in the cases where equilibrium spending, c^e, is positive.

$$c^e = \rho\left(\frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\left(\pi+\rho\right)/\rho}\right) - 1\right) \tag{2.18}$$

We now present our results regarding the effects of model parameters on the equilibrium spending, c^e.

Lemma 2.1. *When $Z_1 > 0$, equilibrium spending, c^e, is*

(i) increasing in per unit profit, π, direct-risk elasticity of demand, Z_1, and cross-risk elasticity of demand, Z_2.

(ii) increasing in ρ when $\rho < \rho_t$, and decreasing in ρ when $\rho > \rho_t$, where ρ_t is given below:

$$\rho_t = \frac{\pi Z_1}{2[1-(Z_1+Z_2)] + \sqrt{(Z_2-2)^2[1-(Z_1+Z_2)]}}, \qquad \textit{when } Z_1 + Z_2 < 1 \tag{2.19}$$

The case $Z_1 = 0$ could be included in the Lemma but then the words 'increasing' and 'decreasing' would be replaced by 'non-decreasing' and 'non-increasing' respectively, as the derivatives ($\partial c^e/\partial \pi$, $\partial c^e/\partial Z_1$, $\partial c^e/\partial Z_2$, and $\partial c^e/\partial \rho$) are equal to zero.

Proof: Proof of Lemma 2.1 is presented in Appendix 2.6.1.

As stated in Lemma 2.1, the equilibrium spending, c^e, for the firm is increasing in per unit profit, π. Keeping everything else constant, higher per unit profit increases losses in an adverse event, motivating the firm to spend more to reduce the likelihood of such an event.

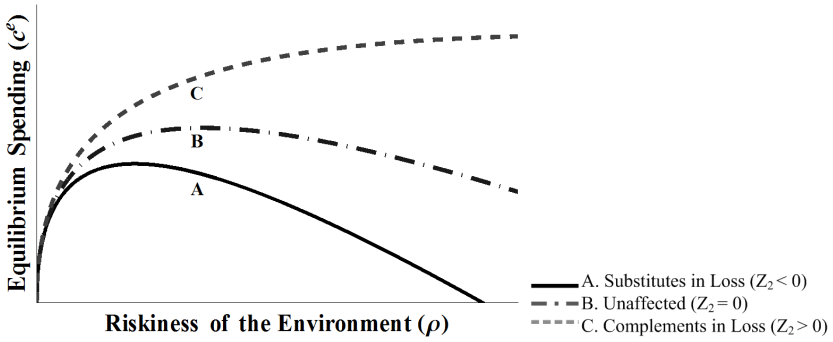
$$\frac{\partial c^e}{\partial \pi} = \frac{Z_1}{\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}} \tag{2.20}$$

Furthermore, (2.20) shows that the effect of a change in the per unit profit, π, on the equilibrium spending increases when the cross-risk elasticity of demand, Z_2, decreases in magnitude.

As stated in Lemma 2.1, the equilibrium spending, c^e, is increasing in ρ when ρ is below a certain threshold value, ρ_t provided in (2.19), and decreasing in ρ when ρ is above that threshold value. This effect is illustrated for different cross-risk elasticities of demand (Z_2) in Figure 2.3. The threshold values for curves A and B can be observed from Figure 2.3 but the threshold value for curve C ($\rho_t = 7.08$) is outside the figure range. This effect is interesting because it shows that after some point, the environment is so risky that the firm is nearly always in a bad state; decreased spending to save costs provides more benefit than increasing the security spending. Our results echo the finding for single firms that security investment is not necessarily always increasing in vulnerability (Gordon and Loeb 2002). In our model, this extreme riskiness of the environment can be due to a very high adverse event arrival rate, Λ, a very long effect duration, $1/\mu$, or both.

Lemma 2.1 shows that the equilibrium spending, c^e, for the firm is increasing in direct-risk elasticity of demand, Z_1. Increased percentage losses from adverse events

Figure 2.3: Sensitivity of equilibrium spending to changes in ρ.



force the firm to increase the level of preventative measures.

$$\frac{\partial c^e}{\partial Z_1} = \frac{\pi + \rho}{\sqrt{Z_2^2 + 4Z_1 \frac{\pi+\rho}{\rho}}} \tag{2.21}$$

As (2.21) shows, the change in the equilibrium spending with respect to a change in Z_1 increases when the cross-risk elasticity of demand, Z_2, decreases in magnitude or when the riskiness of the environment, ρ, increases.

As Lemma 2.1 shows, the equilibrium spending for the firm is increasing in cross-risk elasticity of demand, Z_2. The change in the equilibrium spending with respect to a change in Z_2 is as follows:

$$\frac{\partial c^e}{\partial Z_2} = \frac{1}{2}\left(1 + \frac{Z_2}{\sqrt{Z_2^2 + 4Z_1 \frac{\pi+\rho}{\rho}}}\right)\rho \tag{2.22}$$

(2.22) shows that the change in the equilibrium spending with respect to a change in Z_2 increases when the direct-risk elasticity of demand, Z_1, decreases or when the riskiness of the environment, ρ, increases.

Figure 2.2b presents an enlarged section of Figure 2.2a highlighting the equilibrium points.In Figure 2.2b, points A, B and C represent the equilibrium spending points. For example, point A in Figure 2.2b corresponds to the intersection (equilibrium) point of response curves A in Figure 2.2a. As Lemma 2.1 shows, equilibrium spending levels increase with Z_2. Therefore, equilibrium spending is always lowest when firms are substitutes in loss (point A) and highest when firms are complements in loss (point C).

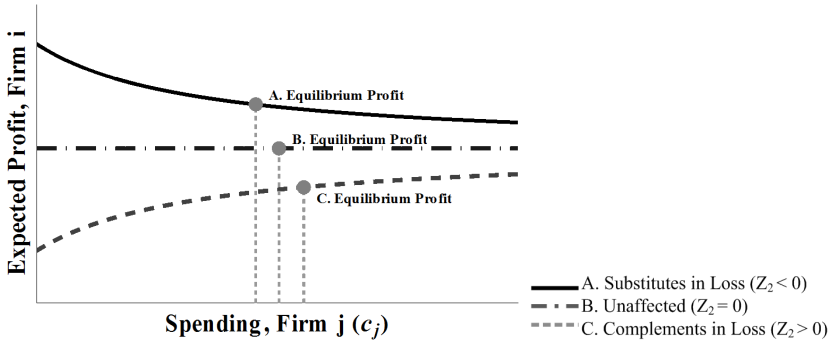
2.3.3 Expected Profit

In this section, we analyze the expected profit rate of a firm at the equilibrium security spending level, denoted as $E[\Pi^e]$. By substituting the equilibrium spending level, c^e, given in (2.18) into (2.14), we obtain expected profit rate for a firm at the equilibrium security spending level as:

$$E[\Pi^e] = \left(1 - \frac{\rho(Z_1 + Z_2)}{\rho + c^e}\right)(\pi - c^e)Q \tag{2.23}$$

Without loss of generality, we now focus on firm 1 with expected profit, $E[\Pi^e]$, at the equilibrium spending, c^e. However, let us consider what could happen if firm 2 spent an amount not at equilibrium, c_2^e. For a given spending level, c_2, of firm 2, if firm 1 follows its best response spending level, c_1^*, as given in (2.15), then the profit curves of firm 1 as a function of firm 2 spending level would look like those presented in Figure 2.4. The differences are due to the nature of the parameter Z_2, that is, they depend on how firm 1's demand is affected by events in firm 2 (just as in the optimal spending cases). The equilibrium profit obtained at equilibrium spending, c^e, is shown on each case by a dot. For both the cases where firms are either substitutes in loss or complements in loss (when $Z_2 \neq 0$), if the other firm increases its security spending, the best response of the firm is to "move to the middle" (or "regress to the mean") in the sense that the firm's security spending approaches that of the unaffected case, where $Z_2 = 0$. This regression to the mean phenomenon can also be observed in Figure 2.4, functions A and C.

Figure 2.4: Expected profit curves under best response spending.



Expected profit curves are generated with the parameters $Z_1 = .3, \pi = 100, \rho = 5$ and $Z_2 = \{-0.3, 0, 0.3\}$.

We now analyze the effects of model parameters on the expected profit at equilibrium, or equilibrium profit, $E[\Pi^e]$.

Lemma 2.2. *When $Z_1 > 0$, equilibrium profit, $E[\Pi^e]$ is*

(i) increasing in per unit profit, π,

(ii) decreasing in direct-risk elasticity of demand, Z_1, and cross-risk elasticity of demand, Z_2,

(iii) decreasing in riskiness of the environment, ρ, when $Z_2 \geq 0$, or when $Z_2 < 0$ and $\rho < \rho_T$, and increasing in ρ when $Z_2 < 0$ and $\rho > \rho_T$, where ρ_T is given below:

$$\rho_T = \frac{\pi(2Z_1 + Z_2)^2}{(4Z_1 + Z_2^2)(1 - (Z_1 + Z_2))}, \qquad \text{when } Z_1 + Z_2 < 1 \tag{2.24}$$

The case $Z_1 = 0$ could be included in the Lemma but then the words 'increasing' and 'decreasing' would be replaced by 'non-decreasing' and 'non-increasing' respectively, as the derivatives ($\partial E[\Pi^e]/\partial\pi$, $\partial E[\Pi^e]/\partial Z_1$, $\partial E[\Pi^e]/\partial Z_2$, and $\partial E[\Pi^e]/\partial\rho$) are equal to zero.

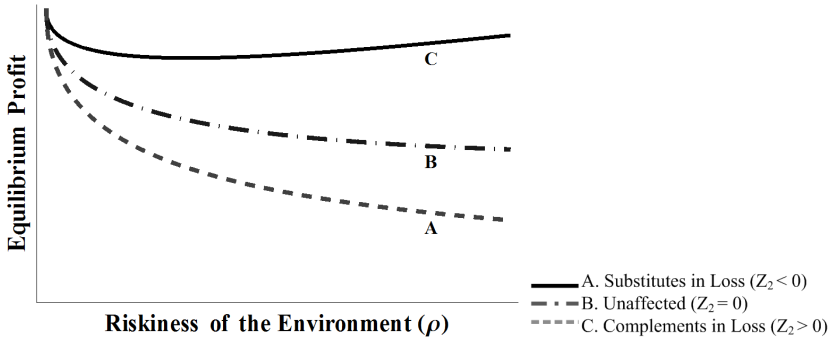
Proof: Proof of Lemma 2.2 is presented in Appendix 2.6.2.

The equilibrium profit, $E[\Pi^e]$, for the firm is increasing in per unit profit, π. Keeping everything else constant, higher per unit profit increases the overall profit rate. The equilibrium profit, $E[\Pi^e]$, for the firm is decreasing in direct-risk elasticity of demand, Z_1. Increased percentage losses lead to fewer products being sold and thus decreased total expected profit.

The equilibrium profit, $E[\Pi^e]$, for the firm is decreasing in the riskiness of the environment ρ when cross-risk elasticity is non-negative ($Z_2 \geq 0$; firms are unaffected or complements in loss). This effect is shown by curves B and C in Figure 2.5. The equilibrium profit, $E[\Pi^e]$ is also decreasing when cross-risk elasticity is negative ($Z_2 < 0$) and the riskiness is below the threshold given by (2.24). Above that threshold value, ρ_T, equilibrium profit is increasing in the riskiness for firms that are substitutes in loss. In that case, the environment becomes extremely risky and firms actually gain demand when their competitors have adverse events, therefore the profits can increase as each firm spends less on security. This effect is shown by curve A in Figure 2.5 where profits initially decrease but begin increasing when ρ is above the threshold value, ρ_T.

The equilibrium profit for the firm is decreasing in cross-risk elasticity of demand, Z_2. Figure 2.4 presents the expected profits under best response spending

Figure 2.5: Sensitivity of equilibrium profit to changes in ρ.



Curves are generated with the parameters set to $Z_1 = .5, \pi = 10, Q = 100$, and $Z_2 = \{-0.4, 0, 0.4\}$.

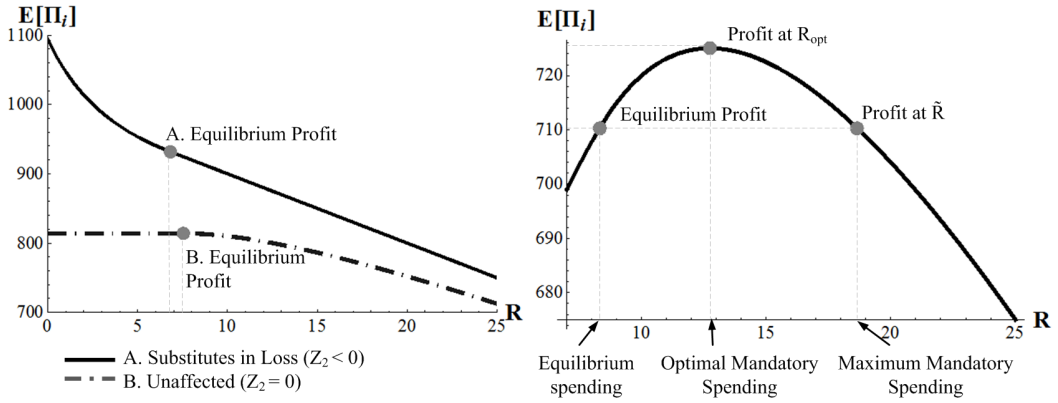
levels, highlighting the equilibrium points for the example presented in Figure 2.2 for different values of cross-risk elasticity of demand, Z_2. In Figure 2.4, points on functions A, B and C represent the equilibrium profit under equilibrium spending. As Lemma 2.2 shows, equilibrium profit levels decrease with Z_2. Therefore, equilibrium profit is always highest when firms are substitutes in loss (point on function A) and lowest when firms are complements in loss (point on function C).

2.3.4 Profits under Regulation and Industry Standardization

In this section, we analyze the effects of regulation or industry standardization on expected profits of the firm. Let us define R as the minimum spending requirement per unit product that ensures that the firm satisfies the regulation or industry standardization conditions. As long as the requirements of regulation or industry standardization dictate spending below equilibrium spending level, then each firm will spend their equilibrium amount. However, once the required spending, R, is above the equilibrium spending, then both firms will have to spend more than they would like. In this section, we are interested in cases where requirements are binding, i.e., required spending, R, is greater than or equal to the equilibrium spending, c^e. Figure 2.6 presents the expected profits of a firm as a function of security spending for different cross-demand elasticity, Z_2, values.

Proposition 2.1. *When the firms are substitutes in loss or unaffected by each other (that is, one firm's demand increases or is unaffected by an event at the other firm: $Z_2 \leq 0$), then profits decline for both firms when minimum mandatory security spending is introduced.*

Figure 2.6: Expected profit for symmetric firms with mandatory minimum security requirement.



(a) Substitutes and Unaffected ($Z_2 \leq 0$) (b) Complements ($Z_2 > 0$)

Proof: Proof of the Proposition 2.1 is presented in Appendix 2.6.3.

Proposition 2.2. *When the firms are complements in loss (that is, one firm's demand decreases due to an event at the other firm, $Z_2 > 0$), then there is a region in which profits increase under minimum mandatory security spending. Profits will remain above the equilibrium profit, $E[\Pi^e]$, while the required spending, R, is between c^e and $\tilde{R}$ and will be maximized when required spending is at R_{opt}, where $c^e \leq R_{opt} \leq \tilde{R}$. Spending levels R_{opt} and $\tilde{R}$ are given as follows:*

$$\begin{aligned} R_{opt} &= \sqrt{(Z_1+Z_2)(\pi+\rho)\rho} - \rho \geq c^e \\ \tilde{R} &= \rho\left(\frac{(Z_1+Z_2)(\pi+\rho)}{c^e+\rho} - 1\right) \geq R_{opt} \end{aligned}$$

Proof: Proof of the Proposition 2.2 is presented in Appendix 2.6.3.

Propositions 2.1 and 2.2 show that the effect of regulations or industry standards on firm profits is case dependent, varying by customer response to adverse security related events in competitor firms. Proposition 2.1 states that when firms are substitutes in loss or unaffected by an event at the other firm, regulation or industry standardization which improves industry security will decrease firm profits. Figure 2.6a depicts this result. Therefore, in this case, the benefits of improvements in the industry security should be weighed against the reduction in firm profits before a regulation or industry standard is imposed.

Proposition 2.2 states that when firms are complements in loss, there is a range of minimum mandatory spending beyond the market equilibrium over which firm

profits will increase. Therefore, in this case, there is a range under which both industry security and firm profits can be improved through a regulation or industry standard, as illustrated in Figure 2.6b. With this proposition, we provide the range over which this case becomes the well-known "prisoner's dilemma" from game theory. It can easily be shown that this range is increasing in the riskiness of the environment, ρ by considering the derivatives of $\tilde{R} - c^e$ and $R_{opt} - c^e$ w.r.t. ρ. Beyond the range where firm profits increase, the benefits of improvements in the industry security should be weighed against the reduction in firm profits before regulation or industry standards are imposed.

Our study shows that we cannot ignore the impact of the competitor's state and customer response to adverse security events. Furthermore, our results show that the interaction between the direct-risk elasticity of demand and the cross-risk elasticity of demand must be taken into account. We can use customer utility models to examine how minimum mandatory spending affects consumer surplus and total social welfare. Our findings show that consumer surplus is always increasing under minimum mandatory spending, and that it is also possible to obtain increases in total social welfare in all cases under appropriate conditions.

2.4. Discussion of Modeling Alternatives

In this section we relax certain assumptions to examine the robustness of our model. We begin by examining the functional form of security breaches, consider the impact of correlated adverse event arrivals, then look at asymmetric customer reactions to firms' adverse events.

2.4.1 Alternate Functional Forms for Security Breaches

We are able to obtain the same results for other functional forms for the arrival rate, λ. In particular, we first consider the function $\lambda_i = \Lambda_i/(c_i + 1)^\alpha$, where $\alpha > 0$. With this functional form, our model becomes analytically intractable when $\alpha \neq 1$, so we obtain our results through numerical analysis. We have tried the cases where $\alpha \in \{.5, 1, 2, 3\}$ and are able to obtain optimal spending response curves which follow the same shape as those shown in Figure 2.2. In addition, this analysis shows that the insights from Propositions 1 and 2 hold.

In order to examine additional linear and non-linear functional forms, we need to adapt our approach. By redefining the probabilities given by equations (2.8) using the relationship

$$x_i = 1 + \frac{c_i}{\rho_i}$$

we obtain the following probability functions:

$$\begin{aligned} P_{gg} &= \left(1 - \frac{1}{x_i}\right)\left(1 - \frac{1}{x_j}\right), \qquad P_{bb} = \frac{1}{x_i}\frac{1}{x_j} \\ P_{bg} &= \frac{1}{x_i}\left(1 - \frac{1}{x_j}\right), \qquad P_{gb} = \left(1 - \frac{1}{x_i}\right)\frac{1}{x_j} \end{aligned} \tag{2.25}$$

In this form, we can see that

$$\begin{aligned} P_{gg} &= P_{1g}P_{2g}, \qquad P_{bb} = P_{1b}P_{2b} \\ P_{bg} &= P_{1b}P_{2g}, \qquad P_{gb} = P_{1g}P_{2b} \end{aligned} \tag{2.26}$$

where

$$\begin{aligned} P_{ig} &= 1 - \frac{1}{x_i}, \qquad \text{probability that firm } i = 1, 2 \text{ is in a good state} \\ P_{ib} &= \frac{1}{x_i}, \qquad \text{probability that firm } i = 1, 2 \text{ is in a bad state} \end{aligned} \tag{2.27}$$

As a result, we can approximate our CTMC model with a periodic model where we denote the probability that firm i is in a good (bad) state, P_{ig} (P_{ib}), as a function of spending as given in (2.25) through (2.27).

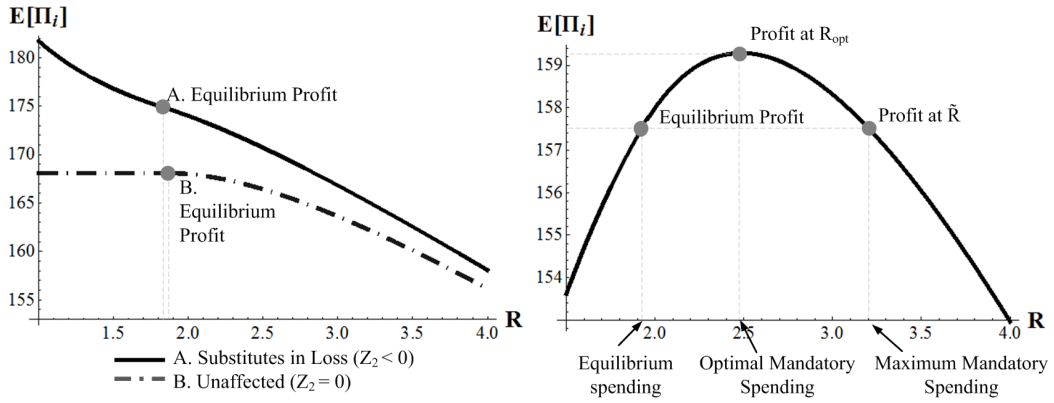
The relationship between our CTMC model and the periodic model allows us to employ the functional forms utilized in Gordon and Loeb (2002). Using a probability function derived from their "first class of security breach probability functions", we set the probability that firm i is in a bad state, $P_{ib} = 1/(c_i + 1)^{\alpha}$. The probabilities for the four possible states given by (2.26), then become:

$$\begin{aligned} P_{gg} &= \left(1 - \frac{1}{(c_1+1)^{\alpha}}\right)\left(1 - \frac{1}{(c_2+1)^{\alpha}}\right), \qquad P_{bb} = \frac{1}{(c_1+1)^{\alpha}(c_2+1)^{\alpha}} \\ P_{bg} &= \frac{1}{(c_1+1)^{\alpha}}\left(1 - \frac{1}{(c_2+1)^{\alpha}}\right), \qquad P_{gb} = \left(1 - \frac{1}{(c_1+1)^{\alpha}}\right)\frac{1}{(c_2+1)^{\alpha}} \end{aligned} \tag{2.28}$$

Numerical analysis of the model with the probabilities given by (2.28) yields results similar to our original CTMC model. Figure 2.7 illustrates that the insights from Propositions 1 and 2 hold for the probabilities given by (2.28).

Likewise, we use this method to consider a functional form derived from Gordon and Loeb (2002)'s "second class of security breach probability functions" and we set

Figure 2.7: Expected profit for symmetric firms with mandatory minimum security requirement.



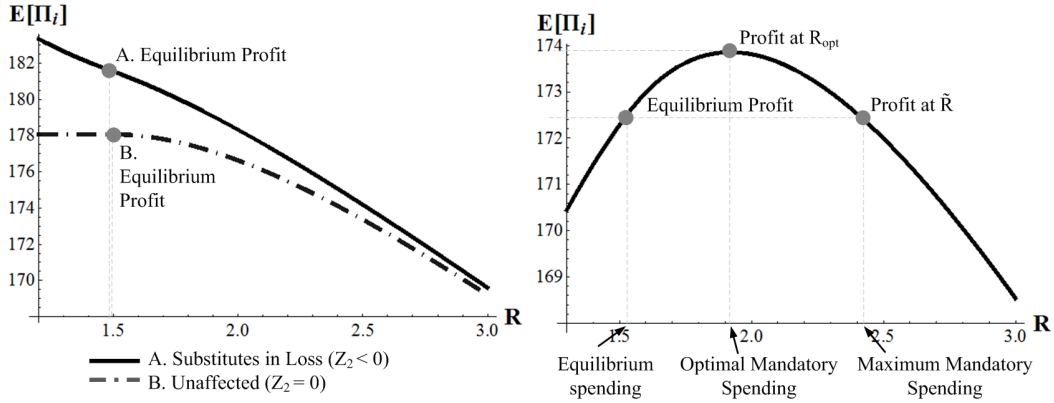
(a) Substitutes and Unaffected ($Z_2 \leq 0$) (b) Complements ($Z_2 > 0$)

the probability that firm i is in a bad state, $P_{ib} = v^{\alpha c_i + 1}$ for a constant $v \in [0, 1]$. Once again, the probabilities for the four possible states given by (2.26) become:

$$\begin{aligned} P_{gg} &= (1 - v^{\alpha c_1 + 1})(1 - v^{\alpha c_2 + 1}), \qquad P_{bb} = v^{\alpha c_1 + 1} v^{\alpha c_2 + 1} \\ P_{bg} &= v^{\alpha c_1 + 1}(1 - v^{\alpha c_2 + 1}), \qquad P_{gb} = (1 - v^{\alpha c_1 + 1}) v^{\alpha c_2 + 1} \end{aligned} \tag{2.29}$$

Once again, numerical analysis of the model that results when using the probabilities given by (2.29) yields results similar to our original CTMC model. Figure 2.8 illustrates that the insights from Propositions 1 and 2 hold for the probabilities given by (2.29).

Figure 2.8: Expected profit for symmetric firms with mandatory minimum security requirement.



(a) Substitutes and Unaffected ($Z_2 \leq 0$) (b) Complements ($Z_2 > 0$)

Our numerical analysis suggests that the insights obtained in our model are not

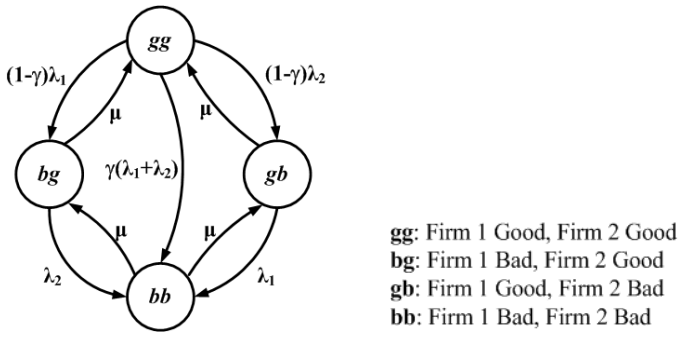
dependent on the functional form $\lambda = \Lambda/c$ but can be generalized to a broader class of security risk functions.

2.4.2 Correlated Arrivals

In some cases, a part of the overall attack volume may be from common sources leading to correlated arrivals of successful security attacks. Here we discuss how such a correlation affects the steady state probabilities, and thus optimal spending and expected profits, for the two firms. In this section we present a modified version of our original model from Section 2.2 which incorporates correlated arrivals, as illustrated in Figure 2.9. Here, γ is a correlation parameter that is an increasing function of the actual correlation between the attack arrival processes of the firms. When $\gamma = 0$, we obtain the model in Section 2.2.

As shown in Figure 2.9, when both firms are in a *good* state (gg), the system may move to any of the other three states. An event that moves both firms directly from the gg state to the bb state arrives at a rate of $\gamma(\lambda_1 + \lambda_2)$. An event that moves the firms from gg to bg (or gb) arrives at a rate of $(1-\gamma)\lambda_1$ (or $(1-\gamma)\lambda_2$). From the bg (gb), an event may move the system into state bb with an arrival rate of λ_1 (λ_2). Again, an event for the firm lasts for an expected duration of $\frac{1}{\mu}$, independently of the other firm.

Figure 2.9: Correlated Arrivals State Diagram.



Each firm can either be in a *bad* state, where the firm is still under the effects of an adverse event it has experienced, or in a *good* state, where the firm is no longer under the negative effects of its most recent adverse event.

After obtaining the steady state probabilities for this system, we are able to show that the probabilities for the states gg, bg, and gb are all decreasing and state bb is increasing in γ. Further, the expected profit at equilibrium, $E[\Pi^e]$ is also decreasing

in γ. The equations and proofs of these results are given in Appendix 2.6.4.

Finding equilibrium spending analytically in the correlated arrivals case is intractable. Thus, we perform a numerical analysis to learn how our propositions regarding minimum mandatory spending would be affected by correlated arrivals of adverse events. We first set the direct-risk elasticity at low, medium, and high values $(0.25, 0.5, 0.9)$ with various cross-risk elasticity values ($Z_2 < 0$). We then set the riskiness of the environment, ρ, at low and high values $(1, 5)$, and finally we set the correlation parameter, γ at values ranging from low to high $(0.2, 0.4, 0.6, 0.75, 0.9)$. We hold all other parameters constant throughout the test suite.

For each set of parameter combinations, we find equilibrium spending, c^e, and then calculate the expected profit at equilibrium, $E[\Pi^e]$, and expected profit under minimum mandatory spending, $E[\Pi^R]$, for $R \geq c^e$. We then calculate the difference between the two expected profits, $E[\Pi^R] - E[\Pi^e]$, and identify the situations where this value is positive. When firms are complements in loss ($Z_2 > 0$), we observe that Proposition 2 results continue to hold for the correlated arrivals case. However, when firms are substitutes in loss ($Z2 < 0$), our numerical analysis provides some interesting observations regarding Proposition 1.

OBSERVATION 1. *When firms are substitutes in loss and adverse event arrivals are correlated, there is opportunity window for firms to increase profits under minimum mandatory spending.*

OBSERVATION 2. *When the correlation parameter, γ, is larger, then the opportunity window for both firms to increase profits under minimum mandatory spending increases, everything else being constant.*

OBSERVATION 3. *When the riskiness of the environment, ρ, is smaller, then the opportunity window for both firms to increase profits under minimum mandatory spending increases, everything else being constant.*

These observations lead us to the conclusion that when the arrival of adverse events is correlated, government regulations or industry standardization is beneficial to firms for a wider range of parameters compared to the original model. In our original model, regulation or standardization is only beneficial when firms are complements in loss (Proposition 2); however, these observations show us that regulation or standardization can be quite beneficial even when firms are substitutes in loss (Proposition 1) when attack arrivals are correlated. The size of this window for

increased profits at both firms increases even further when γ is larger or the riskiness of the environment, ρ, is smaller. Overall, we conclude that considering correlation in arrivals extends the justification for regulation or standardization.

2.4.3 Asymmetric Cases

In the asymmetric case, we allow the direct- and cross-risk elasticity of demand parameters (Z_1 and Z_2) to be different for each firm. Whereas the expected profit equation in the symmetric case was given by (2.14), the expected profit equation for the asymmetric case is:

$$E[\Pi_i] = \left(1 - \frac{Z_{i1}\rho}{c_i + \rho} - \frac{Z_{i2}\rho}{c_j + \rho}\right)(\pi - c_i)Q \qquad \text{for } i = 1, 2, \text{ and } j \neq i.$$

Likewise, while the optimal spending equation in the symmetric case was given by (2.15), the optimal spending equation for the asymmetric case is:

$$c_i^*(c_j) = \rho\left(\sqrt{\frac{Z_{i1}(\pi + \rho)(c_j + \rho)}{\rho(c_j + \rho - Z_{i2}\rho)}} - 1\right) \qquad \text{for } i = 1, 2, \text{ and } j \neq i.$$

We use numerical analysis for the remainder of our exploration of the asymmetric case. Our numerical analysis demonstrates that, depending on its own cross-risk elasticity of demand, each firm's best response curve will look like the corresponding best response curve in the symmetric case as illustrated in Figure 2.2. That is, if firm 1 is a substitute in loss (in reaction to adverse events at firm 2), $Z_{12} < 0$, then firm 1's optimal spending response curve would look like the horizontal curve A in Figure 2.2. Likewise, if firm 2 is a complement in loss (in reaction to adverse events at firm 1), $Z_{22} > 0$, its optimal spending response curve would look like the vertical curve C in Figure 2.2. A firm which is a complement in loss will still decrease its spending in reaction to increased spending by its competitor; a firm which is a substitute in loss will still increase spending in reaction to increased spending by its competitor; and an unaffected firm will not adjust its spending in reaction to a change in spending by its competitor.

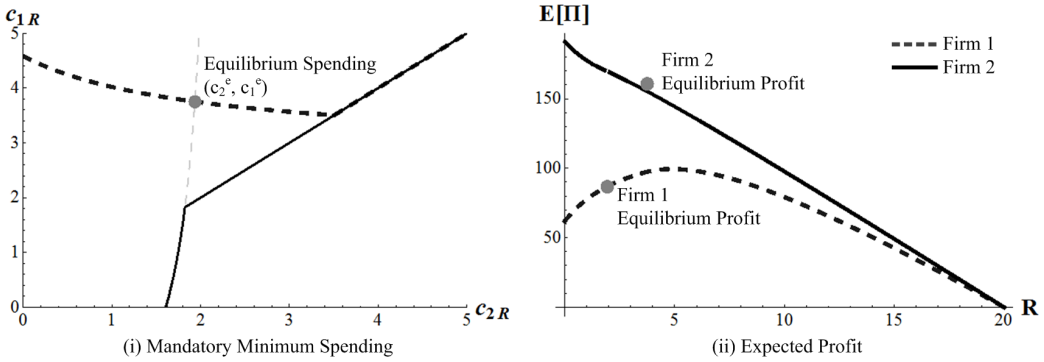
There are three possible combinations of cross-risk elasticity that are of interest: (a) one firm is a complement in loss while the other is a substitute in loss, (b) both firms are substitutes in loss, and (c) both firms are complements in loss. In all cases, without loss of generality, we focus on the instance where the equilibrium spending for firm 1, c_1^e, is greater than or equal to the equilibrium spending for firm 2, c_2^e,

as shown in Figures 2.10a-2.10c (i). Firm 2 will be affected first when a minimum mandatory spending, R, greater than market equilibrium is introduced since $c_2^e < c_1^e$. We first examine what happens when firm 1 is a complement in loss ($Z_{12} > 0$) while firm 2 is a substitute in loss ($Z_{22} < 0$), as illustrated in Figure 2.10a. Figure 2.10a (i) illustrates the effects of minimum mandatory spending, R, on each firm by plotting each firm's spending as a function of R. For firm 1, the x-axis is R, the y-axis is c_1, and the dotted curve is the maximum of c_1^* and R when firm 2 myopically spends R. For firm 2, the y-axis is R, the x-axis is c_2, and the solid curve is the maximum of c_2^* and R when firm 1 myopically spends R . It can be seen, then, that when the minimum mandatory spending, R, greater than market equilibrium is introduced, it will affect firm 2 before firm 1. As a result, while firm 2 will be required to increase its spending as soon as $R > c_2^*(R)$, firm 1 will reduce its spending until the minimum mandatory spending, R, exceeds $c_1^*(R)$. Profits for firm 2 always decrease and profits for firm 1 increase for a range of R values (Figure 2.10a (ii)). In this situation, firm 2 would oppose any move from firm 1 to increase the minimum mandatory spending level, R above c_2^e. The equilibrium profit for firm 2 in Figure 2.10a (ii) is not on firm 2's expected profit under minimum mandatory spending curve because firm 2 is required to spend R as soon as $R > c_2^*(R)$. Thus, while firm 1 is spending $c_1^*(R)$, firm 2 is spending R. Since firm 2 is a substitute in loss, profit is lower under R than at equilibrium.

We next examine what happens when both firms are substitutes in loss ($Z_{i2} < 0$, for $i = 1, 2$) as illustrated in Figure 2.10b. Here, firm 2 will be required to increase its spending as soon as $R > c_2^*(R)$ along with firm 1 who will also increase its spending according to its optimal spending response curve until it must also spend R. This results in decreased profits for both firm 1 and firm 2 (Figure 2.10b (ii)). As in Figure 2.10a (ii), the equilibrium profit for firm 2 is not on firm 2's expected profit under minimum mandatory spending curve. In this situation, both firms would oppose any move to increase the minimum mandatory spending level, R above c_2^e.

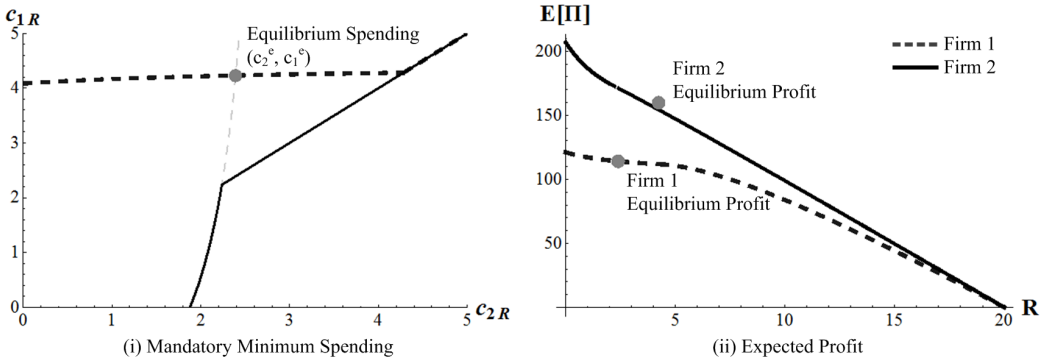
Finally, we examine what happens in the asymmetric case when both firms are complements in loss ($Z_{i2} > 0$, for $i = 1, 2$). Figure 2.10c illustrates a case where $Z_{12} > Z_{22} > 0$. The result here is that firm 1 will benefit from a regulation or industry standard and reduce its spending from market equilibrium, as long as the minimum mandatory spending, R, is less than c_1^e while firm 2 will be required to

Figure 2.10: Spending and Expected Profit Curves for Asymmetric Firms



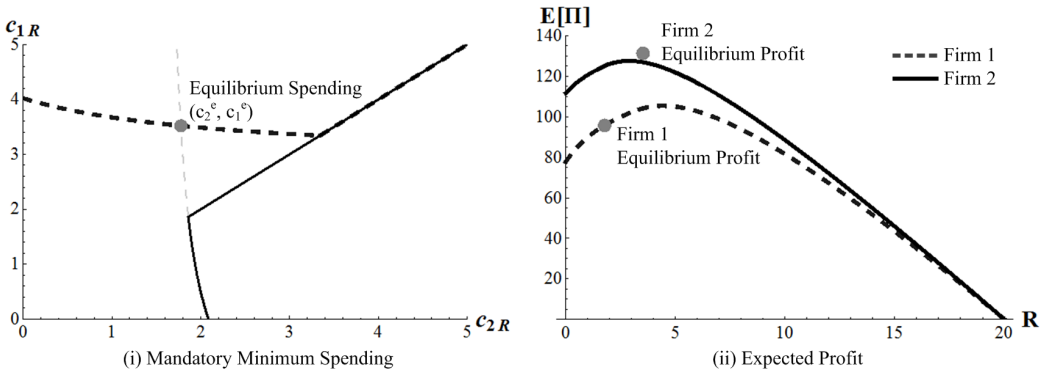
(i) Mandatory Minimum Spending (ii) Expected Profit

(a) One firm is a complement in loss and the other is a substitute in loss ($Z_{12} > 0 > Z_{22}$); generated with the parameters $Z_{11} = .5$, $Z_{12} = .4$, $Z_{21} = .4$, $Z_{22} = -.3$, $\rho = 3$, $Q = 10$, and $\pi = 20$.



(i) Mandatory Minimum Spending (ii) Expected Profit

(b) Both firms are substitutes in loss ($0 > Z_{12} > Z_{22}$); generated with the parameters $Z_{11} = .8$, $Z_{12} = -.1$, $Z_{21} = .5$, $Z_{22} = -.45$, $\rho = 3$, $Q = 10$, and $\pi = 20$.



(i) Mandatory Minimum Spending (ii) Expected Profit

(c) both firms are complements in loss ($Z_{12} > Z_{22} > 0$); generated with the parameters $Z_{11} = .5$, $Z_{12} = .3$, $Z_{21} = .3$, $Z_{22} = .2$, $\rho = 3$, $Q = 10$, and $\pi = 20$.

increase its spending as soon as $R > c_2^*(R)$. As illustrated in Figure 2.10c (ii), profits for firm 1 increase for a range of R values. However, profits for firm 2 may or may not increase depending on the situation. In the case illustrated by Figure 2.10c (ii), the equilibrium profit for firm 2 is above firm 2's expected profit under minimum mandatory spending curve; however, it is not necessary when both firms are complements in loss for equilibrium profit for firm 2 to be higher than the profit obtained when firm 2 is required to spend R.

In order to understand the characteristics that may lead to increased profits for both firms when they are complements in loss, we perform a detailed numerical analysis as described in Appendix 2.6.6. We obtain several general observations from this numerical examination of the asymmetric model when both firms are complements in loss. These observations show that there is a significant range for which the insights from Proposition 2 hold even in the asymmetric case.

OBSERVATION 1. *As direct-risk elasticity increases, the opportunity window for firm 2 to increase profit under minimum mandatory spending decreases, everything else being constant.*

For example, according to the specific parameter values we used in our numerical analysis, when direct-risk elasticity is low and cross-risk elasticity medium, direct-risk can vary by up to 20% between firms and still yield opportunity for mutually increasing profits. However, if we move to a medium direct-risk elasticity (with cross-risk elasticity still medium), then firm 2 can only increase profits when direct-risk elasticity varies by no more than 5% between firms. Therefore when direct-risk elasticity increases, firms need to be more "similar" in terms of direct-risk elasticities in order to increase profits for both firms and provide an opportunity for firms to cooperate.

OBSERVATION 2. *As cross-risk elasticity increases, the opportunity window for firm 2 to increase profit under minimum mandatory spending increases, everything else being constant.*

For example, according to the specific parameter values we used in our numerical analysis, when cross-risk elasticity is low, a 5% difference in cross-risk elasticities between firms can yield increased profits for firm 2 but when cross-risk elasticity is high they may differ by as much as 35% between firms and still increase profits for firm 2. Therefore when cross-risk elasticity increases, firms may be more "dissimilar"

in terms of cross-risk elasticities and still increase profits for both firms and provide an opportunity for firms to cooperate.

OBSERVATION 3. *The largest increases in profit for firm 2 occur when the firms are "more similar" in terms of both their direct- and cross-risk elasticity values.*

Recall that $c_1^e > c_2^e$ and firm 1 always increases its profits for our numerical study. Observation 3 states that when the differences in direct- and cross-risk elasticities between firms is small, the increase in profit for firm 2 is largest. Thus, the more similar firms are, the better the opportunity for cooperation.

2.4.4 Signalling

Signalling theory is concerned with the use of honest and dishonest signals (Spence 2002). Regulation and verifiable inter-firm cooperation (such as externally audited information sharing) can be used to enforce honest signalling regarding disclosure of IT security breaches. Indeed, transparent disclosure of IT security breaches is required by law in certain situations (SB-1386 2002, HIPAA 1996), but in many others it is not. Firms may choose to hide information regarding their IT security spending levels. However, as our model considers the Nash equilibrium spending, it is not dependent upon knowing the exact spending of the other firm in order to calculate the market equilibrium spending. Therefore, it should be noted that the transparency is not a crucial issue for our current model where we focus on cases where firms are transparently symmetric and identify the Nash equilibrium spending. On the other hand, if firms can hide information regarding the risk elasticity parameters, this indeed would have an effect on the calculation of the Nash equilibrium spending. Without loss of generality, we focus on Firm 2 as we examine what occurs when either there is not enough information for Firm 1 to accurately estimate Firm 2's risk elasticities (direct or cross) or Firm 2 can deliberately mislead the other about the nature of its elasticities. For the rest of this section, we use the phrase "dishonest signalling" to denote both of these possibilities: dishonest signalling and incorrect estimation of parameters due to lack of information.

To illustrate how dishonest signalling of the elasticity parameters affects the behaviour of our firms, we designed a test suite where we first examine symmetric direct-risk elasticity ($Z_{11} = Z_{21}$) at high, medium and low values (.75, .5, and .25) with various asymmetric cross-risk elasticity values ($Z_{12} \neq Z_{22}$). We next examine

the case where cross-risk elasticity is symmetric ($Z_{12} = Z_{22}$) at a level of high, medium or low as earlier, and direct-risk elasticity is set to various asymmetric values ($Z_{11} \neq Z_{21}$). In all, we had 956 observations in each of the complements in loss and substitutes in loss cases.

For this test suite, we use the following methodology. Starting with the assumption that firms were truly symmetric we solve for the symmetric model (referred to as Model 1) for equilibrium spending (c^e) and profit ($E[\Pi^e]$) with a focus on situations where Firm 1 receives a dishonest signal about Firm 2's risk elasticities. We then solve the asymmetric model (referred to as Model 2) that results from this inaccurate information, obtaining Firm 1's optimal response spending under the asymmetric model (c_1^a). Finally, we use Firm 1's spending from Model 2 to calculate Firm 2's optimal response spending (c_2^{a*}) in Model 1 (since Firm 2 knows the correct model) and then, using the expected profit function from Model 1, calculate expected profit for both Firm 1 and Firm 2 using $c_1 = c_1^a$ and $c_2 = c_2^{a*}$. We then calculate the difference between the expected profit with and without dishonest signalling for each firm. From this analysis, we make the following observations.

OBSERVATION 1. *When firms are complements in loss, if Firm 2 claims either a lower (higher) direct- or cross-risk elasticity, it increases (decreases) Firm 2's profits while decreasing (increasing) Firm 1's profits.*

OBSERVATION 2. *When firms are substitutes in loss, if Firm 2 claims a lower (higher) cross-risk elasticity, it increases (decreases) both firms' profits. Firm 2 claiming a higher direct-risk elasticity also decreases profits for both firms.*

OBSERVATION 3. *When firms are substitutes in loss, if Firm 2 claims a lower direct-risk elasticity, it increases Firm 2's profit and can increase Firm 1's profit if the difference between Firm 2's claim and the real direct-risk elasticity is not too large.*

When firms are complements in loss, observation 1 occurs because Firm 2 spends less (more) on security while Firm 1 spends more (less) when Firm 2 claims a lower (higher) direct- or cross-risk elasticity. Similarly, when firms are substitutes in loss, observations 2 and 3 occur because both firms will spend less (more) on security when Firm 2 claims a lower (higher) direct- or cross-risk elasticity. In summary, when firms are substitutes in loss, under-estimating the other firm's direct and cross-risk elasticities can lead to increased profits for both firms. However, when firms are

complements in loss, the firm with better information can increase its profits while harming the profitability of the other firm.

2.5. Conclusion

This research examines how adverse IT security events at all firms have an impact on customer demand, IT security spending, and profits for a firm. By understanding how customers will react to a competitor's adverse IT security events, managers will have a better understanding of the appropriate IT security spending to maximize their firm's profits. Our model examines the interaction effects of firms' actions when customers react to any adverse IT security event by changing their demand for both firms' products. We describe the impact between firms as the cross-risk elasticity of demand which can be described in three ways: firms may be substitutes in loss, unaffected, or complements in loss. In addition to obtaining analytical results when firms are symmetric, we extend our examination of the symmetric case by examining correlated arrivals of adverse events as well as examining the asymmetric case where customer reactions may be different for each firm. We examine the impact these demand changes have on security spending and profits.

Using a CTMC model, parameters include how customers react to a successful attack (customers go to another firm or leave the market), the duration of the reaction, and the riskiness of the environment. The continuous time model is a natural way to examine how the riskiness of the environment impacts security spending and expected profits given different types of customer reaction, and we derive several results regarding environment risk. Importantly, in riskier environments, there is a wider window of opportunity for firms to increase profits beyond market equilibrium through cooperation or regulation.

With our symmetric model we demonstrate in Proposition 2.2 that, when firms are complements in loss, it is possible for both firms to obtain increased profits over a range once there is a minimum mandatory spending level. With this proposition, we provide the range over which this case becomes the well-known "prisoner's dilemma" from game theory. When firms are asymmetric, it is still possible to increase profits, provided firms are sufficiently alike in direct- and cross-risk elasticities. As a result, it is possible for firms to cooperate, voluntarily increasing IT security spending in an attempt to increase customer confidence in the product. Unsurprisingly, we find

that whenever regulations cause firms to spend more on security, consumer surplus is increasing. When we examine the impact of correlation in adverse event arrivals, we find that as the degree of correlation between the firms increases, the system is more likely to be in a state where both firms have experienced adverse events and, as a result, equilibrium profits decrease for both firms.

Gal-Or and Ghose (2005) model an interaction between firms, in the form of information sharing, in order to obtain demand increases. Our analysis stands in contrast in that a firm can gain demand simply when the other firm spends more money on information security when the firms are complements in loss. Our model requires no interaction or exchange of information between the firms in order to obtain demand increases and increased profits for all firms. This substantial difference leads to very different implications regarding Information Sharing and Analysis Centers (ISACs) under Presidential Decision Directive (PDD) 63. Although information sharing may be a mechanism for coordinating security efforts and thus reducing security costs and improving efficiency of security efforts industry wide, the ISACs may also be a mechanism for enforcing minimum security standards in industries where security breaches in one firm may result in decreased demand at other firms (i.e., preventing negative spill over) thereby increasing profits for all participants.

Recognizing the characteristics of a firm or industry can aid decision makers. Sectors that fall into the complements in loss case would likely include those producing non-essential undifferentiated goods and services, perhaps sharing a common technology or service provider. An example of such a case lies anywhere a breach at a single firm in the industry can erode confidence in the industry itself, such as credit card use which led to the development of the PCI DSS industry standard (Sysnet 2012b). The Maple Leaf listeriosis contamination (Crawford 2008) provided another example of this type of customer reaction. We might also expect consumers to reduce their enrollment in loyalty programs in light of the recent Epsilon breach affecting the customers of at least 47 email marketing clients such as Best Buy, Ritz-Carlton, and McKinsey & Company (Lennon 2011). Sectors that fall into the substitutes in loss case would likely be producing necessities, undifferentiated goods, or not obviously sharing a common technology or service provider. Managers do not need to know the IT security spending amounts of other firms to make use of our results.

Signalling theory is concerned with the use of honest and dishonest signals

(Spence 2002). Regulation and verifiable inter-firm cooperation (such as externally audited information sharing) can be used to enforce honest signalling regarding disclosure of IT security breaches. Indeed, transparent disclosure of IT security breaches is required by law in certain situations (SB-1386 2002, HIPAA 1996), but in many others it is not. Our model does not require knowledge of industry security spending, but assumes that the elasticities and riskiness of the environment are common knowledge, in order to solve for equilibrium spending. Our preliminary work in this area shows that if firms can hide information regarding the risk elasticity parameters, then the calculated equilibria change. An interesting avenue for future research, then, is to expand this investigation to better understand whether firms have incentives to deceive other firms or consumers by providing false signals regarding their demand elasticities.

In our stationary and continuous time model, we obtain a constant stationary per unit security spending. Our numbers can be interpreted as "per unit"; demand is per unit of time, and spending is per unit sold. Periodic settings or sequential games could be used to investigate optimal investment timing or the effects of changes in the security spending level. An interesting application of this idea is building brand identity by spending heavily at the beginning, possibly altering customer reaction of future adverse events.

Additionally, the demand recovery process is modeled in our paper as an instantaneous recovery rather than a gradual return to normal demand levels. We performed a preliminary numerical analysis in correlated arrivals case and found that making the expected duration of adverse events state-specific does not substantially alter our observations. Future research could certainly more closely examine the impact of event duration on spending decisions as well as examining how firms might change the expected event duration. Incident response strategies, for example, have been shown to have an impact on the expected duration of events (Yue and Çakanyildirim 2007). It might be of interest to examine the trade-offs involved in spending to reduce arrival rates versus spending to favorably adjust event duration.

Prices in our model are fixed. In addition to changing prices in response to changing customer demand in response to a security breach, some customers may be willing to accept a less secure product if it came at a discounted price, thereby allowing firms to choose their security posture (set their IT security spending level)

and then price the product accordingly. Such actions could segment the market - high security firms would sell to customers willing to pay for higher security and low security firms would sell to customers unwilling to pay for higher security - or could result in the loss of either low-end or high-end security products altogether. Future work should examine customers with different risk seeking profiles.

Our model is successful in obtaining insights for different possible indirect effects of adverse IT security events on customer demand and firm IT security spending. This paper continues the work to more fully understand the complex interaction of firms' actions and customer reactions to stochastic adverse events, IT security being our primary example. By understanding the nature of these customer reactions, managers can know when it is in their best interest to cooperate with other firms on security or not.

Bibliography

Acohido, B. 2009. Cybercrooks stalk small businesses that bank online. *USA Today* .

Allan, A., P. Warden. 2011. Got an iPhone or 3G iPad? Apple is recording your moves. O'Reilly radar. URL `http://radar.oreilly.com/2011/04/apple-location-tracking.html`.

Arthur, C. 2011a. iPhone keeps record of everywhere you go. guardian.co.uk Technology Blog. URL `http://www.guardian.co.uk/technology/2011/apr/20/iphone-tracking-prompts-privacy-fears`.

Arthur, C. 2011b. Sony suffers second data breach with theft of 25m more user details. guardian.co.uk Technology Blog. URL `http://www.guardian.co.uk/technology/blog/2011/may/03/sony-data-breach-online-entertainment`.

Baker, L. B., J. Finkle. 2011. Sony PlayStation suffers massive data breach. *Reuters* .

Burke, R. J. 2011. Corporate reputations: Development, maintenance, change and repair. R. J. Burke, G. Martin, C. L. Cooper, eds., *Corporate Reputation: Managing Opportunities & Threats*. Psychological and Behavioural Aspects of Risk, Gower.

Cavusoglu, H., H. Cavusoglu, J. Zhang. 2008. Security patch management: Share the burden or share the damage? *Management Science* **54** 657–670.

Cezar, A., H. Cavusoglu, S. Raghunathan. 2010. Competition, speculative risks, and IT security outsourcing. T. Moore, D. Pym, C. Ioannidis, eds., *Economics of Information Security and Privacy*. Springer US, Boston, MA, 301–320.

Conz, N. 2008. Selling security – innovative carriers are nurturing a new kind of customer loyalty by establishing themselves as IT security stalwarts. *Insurance and Technology* **33**(3) 31.

Coyle, P. 2011. Pending cyber security bills 03-19-11. Digital Bond's SCADA Security Portal. URL `http://www.digitalbond.com/2011/03/22/pending-cyber-security-bills-03-19-11/`.

Crawford, T. 2008. Maple Leaf Foods plant reopens after listeriosis outbreak. *National Post* .

DPA. 1998. Data Protection Act 1998. *(UK)* .

Duh, R-R., S. Sunder, K. Jamal. 2002. Control and assurance in e-Commerce: Privacy, integrity, and security at eBay. *Taiwan Accounting Review* **3** 1–27.

Freedman, S.M., M. Schettini Kearney, M. Lederman. 2009. Product recalls, imperfect information, and spillover effects: Lessons from the consumer response to the 2007 toy recalls. *NBER Working Paper No. 15183* .

Gal-Or, E., A. Ghose. 2005. The economic incentives for sharing security information. *Information Systems Research* **16** 186–208.

Gordon, L. A., M. P. Loeb. 2002. The economics of information security investment. *ACM Transactions on Information and System Security* **5** 438–457.

Gordon, L. A., M. P. Loeb, W. Lucyshyn. 2003. Sharing information on computer systems security: An economic analysis. *Journal of Accounting and Public Policy* **22** 461–485.

HIPAA. 1996. Health Insurance Portability and Accountability Act of 1996. *PUBLIC LAW 104-191 (US)* .

IT Governance. 2012. PCI DSS. URL http://www.itgovernance.co.uk/iso27001.aspx.

Javelin. 2007. Data breaches and buyer behavior: Unfolding TJX Press release. *Javelin Strategy and Research Press Release* .

Krebs, B. 2009a. European cyber-gangs target smal U.S. firms, group says. *Washinton Post* .

Krebs, B. 2009b. The growing threat to business banking online. *Washington Post Security Fix Blog* .

Kunreuther, H., G. Heal. 2003. Interdependent security. *Journal of Risk and Uncertainty* **26**(2) 231–249.

Lennon, M. 2011. Massive breach at Epsilon compromises customer lists of major brands. *Securityweek* .

Ponemon, Symantec. 2011. 2010 annual study: U.S. cost of data breach .

SB-1386. 2002. California Senate Bill: Personal information: privacy. http://info.sen.ca.gov/pub/01-02/bill/sen/sb_1351-1400/sb_1386_bill_20020926_chaptered.html.

Smed, S., J.D. Jensen. 2005. Food safety information and food demand. *British Food Journal* **107** 173–186.

Spence, M. 2002. Signaling in retrospect and the informational structure of markets. *American Economic Review* **92** 434–459.

Sysnet. 2012a. ISO 27001 compliance services. URL http://www.sysnetglobalsolutions.com/en/Compliance_and_Standards/ISO_27001.aspx.

Sysnet. 2012b. PCI DSS. URL http://www.sysnetglobalsolutions.com/en/Compliance_and_Standards/PCI_DSS.aspx.

Varian, H. 2004. System reliability and free riding. L. Camp, Stephen Lewis, eds., *Economics of Information Security*, *Advances in Information Security*, vol. 12. Springer US, 1–15. URL http://dx.doi.org/10.1007/1-4020-8090-5_1.

Yu, T., R. H. Lester. 2008. Moving beyond firm boundaries: A social network perspective on reputation spillover. *Corporate Reputation Review* **11**(1) 94–108.

Yue, W., M. Çakanyildirim. 2007. Intrusion prevention in information systems: Reactive and proactive responses. *Journal of Management Information Systems* **24**(1) 329–353.

2.6. Appendices

2.6.1 Proof of Lemma 2.1

Proof. Proof of Lemma 2.1 As stated in Lemma 2.1, in this proof we are only interested in cases where $Z_1 > 0$. Part *(i)* follows from the respective partial derivatives, shown below:

$$\begin{aligned}
\frac{\partial c^e}{\partial \pi} &= \frac{Z_1}{\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}} > 0 \\
\frac{\partial c^e}{\partial Z_1} &= \frac{\pi+\rho}{\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}} > 0 \qquad \text{as both } \pi > 0 \text{ and } \rho > 0 \\
\frac{\partial c^e}{\partial Z_2} &= \frac{1}{2}\left(1 + \frac{Z_2}{\sqrt{Z_2^2 + 4\pi Z_1\frac{\pi+\rho}{\rho}}}\right)\rho > 0
\end{aligned} \tag{2.30}$$

Clearly, (2.30) is true when $Z_2 \geq 0$. Below, we also show it to be true for $Z_2 < 0$.

$$\begin{aligned}
\frac{1}{2}\left(1 + \frac{Z_2}{\sqrt{Z_2^2 + 4\pi Z_1 \frac{\pi+\rho}{\rho}}}\right)\rho > 0 & \\
\sqrt{Z_2^2 + 4\pi Z_1 \frac{\pi+\rho}{\rho}} > -Z_2 > 0 \qquad & \text{since } Z_2 < 0 \\
Z_2^2 + 4\pi Z_1 \frac{\pi+\rho}{\rho} > Z_2^2 & \\
4\pi Z_1 > 0 \qquad & \text{which always holds since both } \pi, Z_1 > 0.
\end{aligned} \tag{2.31}$$

To show part *(ii)*, we once again begin with the partial derivative as follows:

$$\frac{\partial c^e}{\partial \rho} = \frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}\right) - \frac{\pi Z_1}{\rho\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}} - 1 \tag{2.32}$$

$\frac{\partial c^e}{\partial \rho}$ is a continuous function of ρ when $\rho > 0$ and the only positive root of $\frac{\partial c^e}{\partial \rho}$ is ρ_t, as given by (2.19). Therefore, $\frac{\partial c^e}{\partial \rho}$ in (2.32) can change sign at most once. Given this fact, we complete the proof by first showing that spending is increasing in ρ when $\rho < \rho_t$ ($\frac{\partial c^e}{\partial \rho} > 0$ when $\rho < \rho_t$) and then showing that spending is decreasing ρ when $\rho > \rho_t$ ($\frac{\partial c^e}{\partial \rho} < 0$ when $\rho > \rho_t$).

To begin, we show that spending is increasing in ρ when $\rho < \rho_t$ ($\frac{\partial c^e}{\partial \rho} > 0$ when $\rho < \rho_t$) by considering a value, ρ^*, which is smaller than ρ_t. Given the Assumptions 1 and 2 on Z_1 and Z_2, we can find the smallest ρ_t possible, denoted by ρ_t^*, as follows:

$$\begin{aligned}
\rho_t &= \frac{\pi Z_1}{2[1-(Z_1+Z_2)] + \sqrt{(Z_2-2)^2[1-(Z_1+Z_2)]}} \\
&\geq \frac{\pi Z_1}{2[1-(1-1)] + \sqrt{(-1-2)^2[1-(1-1)]}} \\
&= \frac{\pi Z_1}{5} = \rho_t^*
\end{aligned}$$

We now pick a ρ value which is smaller than the smallest ρ_t possible, $\rho_t^* = \frac{\pi Z_1}{5}$. We

shall let $\rho^* = \frac{\pi Z_1}{16} < \rho_t^*$ and evaluate (2.32) for $\rho = \rho^*$.

$$\begin{aligned}
\left.\frac{\partial c^e}{\partial \rho}\right|_{\rho=\rho^*} &= \frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho^*}{\rho^*}}\right) - \frac{\pi Z_1}{\rho^*\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho^*}{\rho^*}}} - 1 \\
&\geq \frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi}{\rho^*}}\right) - \frac{\pi Z_1}{\rho^*\sqrt{Z_2^2 + 4Z_1\frac{\pi}{\rho^*}}} - 1 \\
&= \frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 64}\right) - \frac{16}{\sqrt{Z_2^2 + 64}} - 1 \\
&\geq \frac{1}{2}\left(-1 + \sqrt{Z_2^2 + 64}\right) - \frac{16}{\sqrt{Z_2^2 + 64}} - 1 \qquad \text{since } -1 \leq Z_2 \\
&\geq \frac{1}{2}\left(-1 + \sqrt{0 + 64}\right) - \frac{16}{\sqrt{0 + 64}} - 1 \;=\; 0.5 \;>\; 0 \qquad \text{since } -1 \leq Z_2 \leq 1
\end{aligned} \tag{2.33}$$

From (2.33), we conclude that $\frac{\partial c^e}{\partial \rho} > 0$ for $0 < \rho < \rho_t$. This is because, $\frac{\partial c^e}{\partial \rho}$ is a continuous function of $\rho > 0$ and ρ_t is its only root. Therefore $\frac{\partial c^e}{\partial \rho}$ has the same sign for all $0 < \rho < \rho_t$.

Now, we show that spending is decreasing in ρ when $\rho > \rho_t$ ($\frac{\partial c^e}{\partial \rho} < 0$ when $\rho > \rho_t$). First note that, as (2.19) shows, $\rho_t \to \infty$ when $Z_1 + Z_2 \to 1$. Therefore, we only need to consider cases with $Z_1 + Z_2 < 1$ because of two reasons: (i) when $Z_1 + Z_2 = 1$ the root ρ_t does not exist and hence there is no ρ value that satisfies $\rho > \rho_t$ and (ii) $Z_1 + Z_2 \leq 1$ by Assumption 2.1.

We show that spending is decreasing in ρ when $\rho > \rho_t$ and $Z_1 + Z_2 < 1$, by evaluating $\frac{\partial c^e}{\partial \rho}$ at a ρ value that is greater than ρ_t. We do this by considering the sign of (2.32) in the limit as ρ goes to infinity. That is, we want to consider

$$\begin{aligned}
\lim_{\rho\to\infty} \frac{\partial c^e}{\partial \rho} &= \lim_{\rho\to\infty} \left[\frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}\right) - \frac{\pi Z_1}{\rho\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}} - 1\right] \\
&= \frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1}\right) - 1 < 0 \qquad \text{when } Z_1 + Z_2 < 1
\end{aligned}$$

As we are only interested in cases with $Z_1 + Z_2 < 1$, the equilibrium spending, c^e is decreasing in ρ when $\rho > \rho_t$. □

2.6.2 Proof of Lemma 2.2

Proof. Proof of Lemma 2.2 As stated in Lemma 2.2, in this proof we are only interested in cases where $Z_1 > 0$. As in Lemma 2.1, the proof of parts *(i)* and *(ii)* follow

from the respective partial derivatives of $E[\Pi^e]$, shown below:

$$\frac{\partial E[\Pi^e]}{\partial \pi} = Q\left(1 - \frac{\rho\,(Z_1+Z_2)}{\rho + c^e}\right)\left(1 - \frac{\partial c^e}{\partial \pi}\right) + \frac{Q\rho(\pi - c^e)\,(Z_1+Z_2)\,\frac{\partial c^e}{\partial \pi}}{(\rho + c^e)^2} > 0 \tag{2.34}$$

where $\frac{\partial c^e}{\partial \pi}$ is shown in (2.20)

$$\frac{\partial E[\Pi^e]}{\partial Z_1} = -Q\left(1 - \frac{\rho\,(Z_1+Z_2)}{\rho + c^e}\right)\frac{\partial c^e}{\partial Z_1} - \frac{Q\rho\,(\pi - c^e)\left(\rho + c^e - (Z_1+Z_2)\,\frac{\partial c^e}{\partial Z_1}\right)}{(\rho + c^e)^2} < 0 \tag{2.35}$$

where $\frac{\partial c^e}{\partial Z_1}$ is shown in (2.21)

$$\frac{\partial E[\Pi^e]}{\partial Z_2} = -\frac{2\rho(\pi+\rho)Q\left(-2Z_1 - Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}\right)}{4Z_1(\pi+\rho) + \rho Z_2\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}\right)} < 0 \tag{2.36}$$

Finally, the proof of part *(iii)* follows from the partial derivative of $E[\Pi^e]$ w.r.t. ρ.

$$\frac{\partial E[\Pi^e]}{\partial \rho} = \frac{-Q}{(\rho + c^e)^2}\left[c^e(\pi - c^e)(Z_1+Z_2) + [2\rho c^e + c^{e\,2} + \rho(\rho - (\pi+\rho)(Z_1+Z_2))]\frac{\partial c^e}{\partial \rho}\right] \tag{2.37}$$

As $\frac{\partial E[\Pi^e]}{\partial \rho}$ is a continuous function of ρ when $\rho > 0$ and the only positive root of $\frac{\partial E[\Pi^e]}{\partial \rho}$ is ρ_T, as given by (2.24), therefore, $\frac{\partial E[\Pi^e]}{\partial \rho}$ in (2.37) can change sign at most once. Given this fact, we show that when $Z_2 \geq 0$, $\frac{\partial E[\Pi^e]}{\partial \rho}$ does not change sign and $\frac{\partial E[\Pi^e]}{\partial \rho} < 0$ always holds. On the other hand, when $Z_2 < 0$, $\frac{\partial E[\Pi^e]}{\partial \rho}$ changes sign at ρ_T: total expected profit is decreasing in ρ when $\rho < \rho_T$ ($\frac{\partial E[\Pi^e]}{\partial \rho} < 0$ when $\rho < \rho_T$) and total expected profit is increasing ρ when $\rho > \rho_T$ ($\frac{\partial E[\Pi^e]}{\partial \rho} > 0$ when $\rho > \rho_T$).

Simplifying $\frac{\partial E[\Pi^e]}{\partial \rho}$ given by (2.37), we find:

$$\frac{\partial E[\Pi^e]}{\partial \rho} = \frac{-Q}{(\rho + c^e)^2}A \qquad \text{where } A = \frac{Z_1(\pi - c^e)\left(c^e\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}} + Z_2\pi\right)}{\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}} \tag{2.38}$$

First note that $\frac{-Q}{(\rho+c^e)^2} < 0$ since $Q, c^e, \rho > 0$. Thus we need only consider the sign of the term A in (2.38). The denominator of A, $\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}} > 0$, thus we need

only look at the numerator. When $Z_2 \geq 0$, then $A > 0$ since $c^e\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}} > 0$ and $Z_1, \pi - c^e > 0$. Hence, $\frac{\partial E[\Pi^e]}{\partial \rho} < 0$ when $Z_2 \geq 0$.

Now, we focus on the case when $Z_2 < 0$ and find the range of ρ for which $A > 0$ (and hence $\frac{\partial E[\Pi^e]}{\partial \rho} < 0$). Since $Z_2 < 0$ in this case and $Z_1, \pi - c^e > 0$ in (2.38), $A > 0$ holds when $c^e\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}} > -\pi Z_2 > 0$, as analyzed below:

$$
\begin{aligned}
c^e\sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}} &> -\pi Z_2 \quad \text{substituting } c^e \text{ given by (2.18), we find} \\
\frac{4Z_1 + Z_2^2(2 - (Z_1 + Z_2))}{(2Z_1 + Z_2)^2} &< \frac{\pi+\rho}{\rho} \\
\rho &< \frac{\pi(2Z_1 + Z_2)^2}{(4Z_1 + Z_2^2)(1 - (Z_1 + Z_2))} = \rho_T \qquad (2.39)
\end{aligned}
$$

(2.39) shows that, for the $Z_2 < 0$ case, when $\rho < \rho_T$, $A > 0$ holds and equilibrium profit is decreasing in ρ. We know that $\partial E[\Pi^e]/\partial\rho = 0$ when $\rho = \rho_T$, and lastly when $\rho > \rho_T$, $A < 0$ holds and equilibrium profit is increasing in ρ. □

2.6.3 Proofs for Propositions

Proof. Proof of Proposition 2.1

Let R be the required security spending for firms. By the definition of binding regulation, we have

$$R - c_i^e \geq 0 \qquad \text{for } i = 1, 2. \qquad (2.40)$$

Let $E[\Pi_i^e]$ denote the profit calculated at equilibrium spending, c_i^e, and let $E[\Pi_i^R]$ denote the profit calculated at regulation spending, R, where $i = 1, 2$. As we are focusing on the symmetric case, we can drop the index i. We want to show:

$$
\begin{aligned}
E[\Pi^e] - E[\Pi^R] &\geq 0 \qquad \text{where} \\
E[\Pi^e] - E[\Pi^R] &= \left(1 - \frac{\rho(Z_1 + Z_2)}{\rho + c^e}\right)(\pi - c^e)Q - \left(1 - \frac{\rho(Z_1 + Z_2)}{\rho + R}\right)(\pi - R)Q \\
&= \frac{Q(R - c^e)[(\rho + c^e)(\rho + R) - \rho(Z_1 + Z_2)(\pi + \rho)]}{(\rho + c^e)(\rho + R)} \qquad (2.41)
\end{aligned}
$$

Therefore, to show that $E[\Pi^e] - E[\Pi^R] \geq 0$, we need to prove that (2.41) is nonnegative. By (2.40) we know that $R - c^e \geq 0$ and we know Q, $\rho + c^e$, and $\rho + R > 0$. So, to complete the proof, we need the following to hold in (2.41),

$$
\begin{aligned}
(\rho + c^e)(\rho + R) - \rho(Z_1 + Z_2)(\pi + \rho) &\geq 0 \\
(\rho + R)(\rho + c^e) &\geq \rho(Z_1 + Z_2)(\pi + \rho) \qquad (2.42)
\end{aligned}
$$

We want to show that (2.42) holds. From (2.40), we know that $R \geq c^e$. Therefore, the following holds

$$(\rho + R)(\rho + c^e) \geq (\rho + c^e)^2$$

Hence, we can prove (2.42) by showing that $(\rho + c^e)^2 \geq \rho(Z_1 + Z_2)(\pi + \rho)$.

$$\begin{aligned}
(\rho + c^e)^2 &\geq \rho(Z_1 + Z_2)(\pi + \rho) \\
\left\{\rho + \rho\left(\frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi + \rho}{\rho}}\right) - 1\right)\right\}^2 &\geq \rho(Z_1 + Z_2)(\pi + \rho) \\
Z_2\rho\left(\frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi + \rho}{\rho}}\right) - 1\right) &\geq Z_2\pi \\
Z_2 c^e &\geq Z_2\pi \qquad (2.43)
\end{aligned}$$

In this Proposition, we are considering the case where $Z_2 \leq 0$. When $Z_2 = 0$, (2.43) holds and when $Z_2 < 0$ we require $\pi \geq c^e$ for (2.43) to hold. Note that, $\pi \geq c^e$ holds due to the long-run participation constraint introduced in Section 2.3, as otherwise firms would not be willing to produce the product. Thus, we have proven that $E[\Pi^e] - E[\Pi^R] \geq 0$ for firms that are substitutes in loss or unaffected ($Z_2 \leq 0$). □

Proof. Proof of Proposition 2.2: First, we find a spending level, $\tilde{R}$, such that the profit at that spending level, $E[\Pi^{\tilde{R}}]$, is the same as the profit at equilibrium spending, $E[\Pi^e]$. That is, we want to find $\tilde{R}$ such that

$$\begin{aligned}
E[\Pi^e] &= E[\Pi^{\tilde{R}}] \\
\left(1 - \frac{\rho(Z_1 + Z_2)}{\rho + c^e}\right)(\pi - c^e)Q &= \left(1 - \frac{\rho(Z_1 + Z_2)}{\rho + \tilde{R}}\right)(\pi - \tilde{R})Q \qquad (2.44)
\end{aligned}$$

Solving the equation (2.44) for $\tilde{R}$, we obtain the following two roots

$$\tilde{R}(1) = c^e \quad \text{and} \quad \tilde{R}(2) = \rho\left(\frac{(Z_1 + Z_2)(\pi + \rho)}{c^e + \rho} - 1\right)$$

As we are interested in spending levels besides the equilibrium spending, c^e, we keep $\tilde{R}(2)$ as $\tilde{R}$.

Now we find the regulation spending level, R_{opt}, that maximizes the firm profit. To complete the proof, we then show that $c^e \leq R_{opt} \leq \tilde{R}$ and the firm profit between levels c^e and $\tilde{R}$ is greater than or equal to the equilibrium profit, $E[\Pi^e]$. Consider

the first and second partial derivatives of equilibrium profit provided in (2.23) with respect to c.

$$\begin{aligned}\frac{\partial E[\Pi]}{\partial c} &= Q\frac{\rho(Z_1+Z_2)(\pi+\rho)-(c+\rho)^2}{(c+\rho)^2} \\ \frac{\partial^2 E[\Pi]}{\partial c^2} &= \frac{-2Q\rho(Z_1+Z_2)(\pi+\rho)}{(c+\rho)^3} < 0 \end{aligned} \tag{2.45}$$

The second derivative (2.45) is negative since: (i) Q, c, π, and $\rho > 0$, and (ii) $Z_1 + Z_2 > 0$ in this Proposition. To see (ii), note that $Z_1 + Z_2 \geq 0$ by Assumption 2.1, $Z_1 \in [0, 1]$ by definition, however $Z_2 > 0$ in this Proposition and that makes $Z_1 + Z_2 > 0$. Therefore, $E[\Pi]$ is strictly concave in c and the first-order condition provided below will give us a spending level, R_{opt}, that maximizes the profit.

$$\left.\frac{\partial E[\Pi]}{\partial c}\right|_{c=R_{opt}} = Q\frac{\rho(Z_1+Z_2)(\pi+\rho)-(R_{opt}+\rho)^2}{(R_{opt}+\rho)^2} = 0 \tag{2.46}$$

Solving equation (2.46) for R_{opt} yields the following roots:

$$R_{opt}(1) = -\sqrt{(Z_1+Z_2)(\pi+\rho)\rho} - \rho \quad \text{and} \quad R_{opt}(2) = \sqrt{(Z_1+Z_2)(\pi+\rho)\rho} - \rho \tag{2.47}$$

It is evident that $R_{opt}(1) \leq 0$, so we eliminate $R_{opt}(1)$ and keep $R_{opt}(2)$ as R_{opt}, but we still need to show that $R_{opt}(2) \geq 0$. As $c^e > 0$, showing that $c^e \leq R_{opt}$ (which we do below as a step to show $c^e \leq R_{opt} \leq \tilde{R}$) will also prove that $R_{opt} = R_{opt}(2) \geq 0$ in (2.47). Below we show $c^e \leq R_{opt}$ holds, recalling that we are only interested in the case where $Z_2 > 0$.

$$\begin{aligned} c^e &\leq R_{opt} \\ \rho\left(\frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}\right) - 1\right) &\leq \sqrt{\rho(Z_1+Z_2)(\pi+\rho)} - \rho \\ Z_2\frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}\right) &\leq Z_2\frac{\pi+\rho}{\rho} \\ \rho\frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}\right) &\leq \pi+\rho \quad \text{as } Z_2 > 0 \\ \rho\left(\frac{1}{2}\left(Z_2 + \sqrt{Z_2^2 + 4Z_1\frac{\pi+\rho}{\rho}}\right) - 1\right) &\leq \pi \\ c^e &\leq \pi \end{aligned}$$

Note that, $\pi \geq c^e$ holds due to the long-run participation constraint introduced in Section 2.3, as otherwise firms would not be willing to produce the product. Thus,

$c^e \leq R_{opt}$ holds. Now we show that $c^e \leq R_{opt}$ implies $R_{opt} \leq \tilde{R}$. This concludes the proof of $c^e \leq R_{opt} \leq \tilde{R}$.

$$\begin{aligned} c^e &\leq R_{opt} \\ c^e &\leq \sqrt{\rho(Z_1+Z_2)(\pi+\rho)} - \rho \\ \sqrt{(Z_1+Z_2)(\pi+\rho)\rho} - \rho &\leq \rho\left(\frac{(Z_1+Z_2)(\pi+\rho)}{c^e+\rho} - 1\right) \\ R_{opt} &\leq \tilde{R} \end{aligned}$$

Thus, $c^e \leq R_{opt} \leq \tilde{R}$ holds. To summarize; when $Z_2 > 0$, the expected profit attains its maximum at R_{opt} and $c^e \leq R_{opt} \leq \tilde{R}$ holds, where $E[\Pi^e] = E[\Pi^{\tilde{R}}]$. We also showed that the expected profit is strictly concave in spending when $Z_2 > 0$, therefore for any spending level between c^e and $\tilde{R}$, non-inclusive, the expected profit is larger than the equilibrium profit, $E[\Pi^e]$. □

2.6.4 Correlated Arrivals

Here we present the calculations for steady state probabilities in the correlated arrivals model. We go on to show that the steady state probabilities are decreasing in γ for states gg, bg, and gb and increasing in γ for state bb. We then prove that $E[\Pi^e]$ is also decreasing in γ.

Following our methodology from Section 2.2, we focus on the symmetric case and substitute $\rho = \frac{\Lambda}{\mu}$. Thus, the steady-state probabilities, P_s, for the system in Figure 2.9 can be obtained from the following equations:

$$P_{gg}(\frac{\rho}{c_1} + \frac{\rho}{c_2}) = P_{bg} + P_{gb} \tag{2.48}$$

$$P_{bg}(\frac{\rho}{c_2} + 1) = P_{bb} + P_{gg}(1-\gamma)\frac{\rho}{c_1} \tag{2.49}$$

$$2P_{bb} = P_{bg}\frac{\rho}{c_2} + P_{gb}\frac{\rho}{c_1} + P_{gg}\gamma(\frac{\rho}{c_1} + \frac{\rho}{c_2}) \tag{2.50}$$

$$P_{gb}(1 + \frac{\rho}{c_1}) = P_{gg}(1-\gamma)\frac{\rho}{c_2} + P_{bb} \tag{2.51}$$

$$P_{gg} + P_{bg} + P_{gb} + P_{bb} = 1 \tag{2.52}$$

Where (2.48)-(2.51) are state balance equations and (2.52) normalizes the probabilities to 1. From (2.48)-(2.52), we obtain the steady-state probabilities as follows:

$$P_{gg} = \frac{c_1c_2(\rho c_2 + c_1(\rho + 2c_2))}{c_1^2(\rho + c_2)(\rho + \gamma\rho + 2c_2) + \rho^2c_2(\rho + (1+\gamma)c_2) + \rho c_1(\rho^2 + 4\rho c_2 + (3+\gamma)c_2^2)},$$
$$P_{bb} = \frac{\rho(\gamma c_1^2(\rho + c_2) + \rho c_2(\rho + \gamma c_2) + c_1(\rho^2 + 2\rho c_2 + \gamma c_2^2))}{c_1^2(\rho + c_2)(\rho + \gamma\rho + 2c_2) + \rho^2c_2(\rho + (1+\gamma)c_2) + \rho c_1(\rho^2 + 4\rho c_2 + (3+\gamma)c_2^2)},$$
$$P_{bg} = \frac{\rho c_2(\gamma c_1^2 + \rho c_2 + c_1(\rho - (-2+\gamma)c_2))}{c_1^2(\rho + c_2)(\rho + \gamma\rho + 2c_2) + \rho^2c_2(\rho + (1+\gamma)c_2) + \rho c_1(\rho^2 + 4\rho c_2 + (3+\gamma)c_2^2)},$$
$$P_{gb} = \frac{\rho c_1(c_1(\rho - (-2+\gamma)c_2) + c_2(\rho + \gamma c_2))}{c_1^2(\rho + c_2)(\rho + \gamma\rho + 2c_2) + \rho^2c_2(\rho + (1+\gamma)c_2) + \rho c_1(\rho^2 + 4\rho c_2 + (3+\gamma)c_2^2)} \tag{2.53}$$

If we let $\gamma = 0$, (2.53) reduces to the probabilities given in (2.8).

To examine how correlated arrivals will affect the expected profits for firms, we examine the partial derivatives of the steady-state probabilities with respect to γ, below:

$$\frac{\partial P_{gg}}{\partial \gamma} = \frac{-(\rho c_1c_2(\rho c_2^2 + c_1c_2^2 + c_1^2(\rho + c_2))(\rho c_2 + c_1(\rho + 2c_2)))}{(c_1^2(\rho + c_2)(\rho + \gamma\rho + 2c_2) + \rho^2c_2(\rho + (1+\gamma)c_2) + \rho c_1(\rho^2 + 4\rho c_2 + (3+\gamma)c_2^2))^2} \tag{2.54}$$

$$\frac{\partial P_{bg}}{\partial \gamma} = \frac{-(\rho c_2(-\rho c_1^3 + c_1(\rho^2 + (\rho - c_1)c_1)c_2 + (\rho + c_1)^2c_2^2)(\rho c_2 + c_1(\rho + 2c_2)))}{(c_1^2(\rho + c_2)(\rho + \gamma\rho + 2c_2) + \rho^2c_2(\rho + (1+\gamma)c_2) + \rho c_1(\rho^2 + 4\rho c_2 + (3+\gamma)c_2^2))^2} \tag{2.55}$$

$$\frac{\partial P_{gb}}{\partial \gamma} = \frac{-(\rho c_1(\rho c_2 + c_1(\rho + 2c_2))(-\rho c_2^3 + c_1^2(\rho + c_2)^2 + c_1c_2(\rho^2 + (\rho - c_2)c_2)))}{(c_1^2(\rho + c_2)(\rho + \gamma\rho + 2c_2) + \rho^2c_2(\rho + (1+\gamma)c_2) + \rho c_1(\rho^2 + 4\rho c_2 + (3+\gamma)c_2^2))^2} \tag{2.56}$$

$$\frac{\partial P_{bb}}{\partial \gamma} = \frac{\rho(\rho c_2 + c_1(\rho + c_2))(\rho c_2^2 + c_1c_2^2 + c_1^2(\rho + c_2))(\rho c_2 + c_1(\rho + 2c_2))}{(c_1^2(\rho + c_2)(\rho + \gamma\rho + 2c_2) + \rho^2c_2(\rho + (1+\gamma)c_2) + \rho c_1(\rho^2 + 4\rho c_2 + (3+\gamma)c_2^2))^2} \tag{2.57}$$

We show below that the steady state probabilities for the states gg, bg, and gb are all decreasing in γ while the steady-state probability for state bb is increasing in γ at c^e. The denominator is the same for all four partial derivatives listed in (2.54)-(2.57) and is non-negative. We then need only show the appropriate sign for the numerator of each partial derivative, evaluated at c^e.

State *gg* We want to show:

$$\begin{aligned} -\rho c^{e2}(\rho c^{e2} + c^{e3} + c^{e2}(\rho + c^e))(\rho c^e + c^e(\rho + 2c^e)) &\leq 0 \\ -4\rho c^{e5}(\rho + c^e)^2 &\leq 0 \\ (\rho + c^e)^2 &\geq 0 \end{aligned} \tag{2.58}$$

It is clear that (2.58) holds, thus $\frac{\partial P_{gg}}{\partial \gamma} \leq 0$.

States *bg* and *gb* We want to show:

$$-\left(\rho c^e(-\rho c^{e3} + c^e(\rho^2 + (\rho - c^e)c^e)c^e + (\rho + c^e)^2 c^{e2})(\rho c^e + c^e(\rho + 2c^e))\right) \leq 0$$
$$(-\rho c^{e3} + c^e(\rho^2 + (\rho - c^e)c^e)c^e + (\rho + c^e)^2 c^{e2})(\rho c^e + c^e(\rho + 2c^e)) \geq 0$$

As $(\rho c^e + c^e(\rho + 2c^e)) \geq 0$, we need only show:

$$(-\rho c^{e3} + c^e(\rho^2 + (\rho - c^e)c^e)c^e + (\rho + c^e)^2 c^{e2}) \geq 0$$
$$2\rho(\rho + c^e) \geq 0 \quad (2.59)$$

It is clear that (2.59) holds as all values are non-negative. Thus $\frac{\partial P_{bg}}{\partial \gamma} \leq 0$. And, due to symmetry, it also holds that $\frac{\partial P_{gb}}{\partial \gamma} \leq 0$.

State *bb* We want to show:

$$\rho(\rho c^e + c^e(\rho + c^e))(\rho c^{e2} + c^{e3} + c^{e2}(\rho + c^e))(\rho c^e + c^e(\rho + 2c^e)) \geq 0 \quad (2.60)$$

It is clear that (2.60) holds as all values are non-negative. Thus $\frac{\partial P_{bb}}{\partial \gamma} \geq 0$.

2.6.5 Expected Profit

Under correlated arrivals, the expected profit at equilibrium is given by the equation:

$$E[\Pi^e] = \frac{Q(\pi - c^e)\left(c^{e2} + c^e(2 + \gamma)\rho + \rho^2 - \rho(c^e + c^e\gamma + \rho)\left(Z_1 + Z_2\right)\right)}{c^{e2} + c^e(2 + \gamma)\rho + \rho^2} \quad (2.61)$$

When $\gamma = 0$, (2.61) reduces to the expected profit function given by (2.23). By examining the partial derivative of this expected profit function with respect to γ, we find that expected profits at equilibrium are decreasing in γ.

$$\frac{\partial E[\Pi^e]}{\partial \gamma} = \frac{-Q(\pi - c^e)c^{e2}\rho(c^e + \rho)\left(Z_1 + Z_2\right)}{\left(c^{e2} + c(2 + \gamma)\rho + \rho^2\right)^2} \leq 0$$

Since $\frac{\partial E[\Pi^e]}{\partial \gamma} \leq 0$ as $\pi - c^e \geq 0$, $Z_1 + Z_2 \geq 0$, and all other parameters are non-negative. Therefore, we can conclude that expected profits at equilibrium are decreasing in the correlation of attack arrivals.

2.6.6 Description of Numerical Analysis in Asymmetric Case

In the asymmetric case, one firm will have a lower equilibrium spending than the other. Without loss of generality, we examined parameter combinations where firm 1 has the higher equilibrium spending (i.e. $c_1^e > c_2^e$) and R thus the mandatory minimum spending will apply to firm 2 first. Here, there will always be a window for firm 1 to increase profits once firm 2 is required to spend R and we are interested in cases where it may be possible for firm 2 to also increase profits.

In the first stage, direct-risk elasticity (Z_{i1}) was the same for both firms and only the cross-risk elasticities differed between firms. Direct-risk elasticity for both firms was set to a low, medium, and high value (that is, $Z_{i1} \in \{.25, .5, .75\}$, as shown in Table 2.1) and cross-risk elasticity for firm 1 was set to a value without violating assumptions 1 or 2. It should be noted that a medium value for direct-risk elasticity of .5 allows for the greatest range of values for cross-risk elasticity while still meeting assumptions 1 and 2. Cross-risk elasticity for firm 2 was then set to a value between .025 and Z_{12}, according to the values shown in Table 2.1, Stage 1. In total, 326 combinations of values were examined in this stage. Assumptions 1 and 2 limit the possible combinations once direct-risk elasticity has been set.

In the second stage, cross-risk elasticity (Z_{i2}) was the same for both firms and the direct-risk elasticities differed between firms. Cross-risk elasticity for both firms was set to a low, medium, and high value (as shown in Table 2.1, Stage 2) and direct-risk elasticity for firm 1 was set to a value without violating assumptions 1 or 2. Direct-risk elasticity for firm 2 was then set. In total, 152 combinations of values were examined in this stage which met assumptions 1 and 2.

In the third stage, both direct- and cross-risk elasticity were differed between the firms, under the specific condition that firm 1's direct- and cross-risk elasticity values were greater than those of firm 2 (i.e. $Z_{11} > Z_{21}$, $Z_{12} > Z_{22}$). Here, Z_{11} was set to a low, medium, and high value as in the first stage and Z_{21} was then set to a value lower than firm 1's direct risk elasticity, as illustrated in Table 2.1, Stage 3. Next, Z_{12} was set to a low, medium, or high level and finally, Z_{22} was set at a percentage of firm 1's cross-risk elasticity. Thus, we examined cases where the demand elasticities for the firms differed by as little as 5% and by as much as 95%. The difference in direct-risk elasticities varied independently from the differences in

cross-risk elasticities. There were 218 combinations of values examined in this stage which met assumptions 1 and 2 along with the conditions $c_i^e > 0$ and $c_1^e > c_2^e$.

Finally, we examined the case where both direct- and cross-risk elasticity differed between firms with firm 1's direct-risk elasticity lower than firm 2's and firm 1's cross-risk elasticity greater than firm 2's (i.e. $Z_{11} < Z_{21}$, $Z_{12} > Z_{22}$ shown in Table 2.1, Stage 4). Here, direct-risk elasticity for firm 2 was set to a low, medium, and high value as above while the direct-risk elasticity for firm 1 was set to a value lower than firm 2's direct risk elasticity. Cross-risk elasticity of demand for firm 1, again, was set to a low, medium, or high level and cross-risk elasticity of demand for firm 2 was set at a percentage of firm 1's cross-risk elasticity. The difference in direct-risk elasticities was varied independently from the difference in cross-risk elasticities. There were 186 combinations of values examined in this stage which met assumptions 1 and 2 along with the conditions $c_i^e > 0$ and $c_1^e > c_2^e$.

Table 2.1: Parameter settings for numerical analysis in Asymmetrical Case

Direct Risk		Cross Risk		Total Cases
Firm 1 (Z_{11})	Firm 2 (Z_{21})	Firm 1 (Z_{12})	Firm 2 (Z_{22})	
STAGE 1: Direct risk for both firms is the same, cross risk for firm 1 is higher than cross risk for firm 2.				
.25, .5, .75	$Z_{21} = Z_{11}$	.025, .05, .075, .1, .15, .2, .25, .3, .4, .5 where $Z_{12} \leq \min(Z_{11}, 1 - Z_{11})$	[.01,.50], step .01 plus .025 & .075 where $Z_{22} \leq Z_12$	326
STAGE 2: Direct risk for firm 1 is higher than direct risk for form 2, cross risk for both firms is the same.				
.25, .3, .5, .65, .75, .9 where $Z_{11} \leq 1 - Z_{12}$	[.1, .9] step .01 where $Z_{21} \leq Z_{11}$	.1, .25, .45	$Z_{22} = Z_{12}$	152
STAGE 3: Vary both direct and cross risk for both firms where $Z_{11} > Z_{21}$ and $Z_{12} > Z_{22}$				
.25, .5, .75	$Z_{11}\times$ { .05, .2, .35, .5, .65, .8, .95}	.1, .25, .45 where $Z_{12} \leq \min(Z_{11}, 1 - Z_{11})$	$Z_{12}\times${.05, .2, .35, .5, .65, .8, .95} where $Z_{22} \leq \min(Z_{21}, 1 - Z_{21})$	201
STAGE 4: Vary both direct and cross risk for both firms where $Z_{11} < Z_{21}$ and $Z_{12} > Z_{22}$				
$Z_{21}\times$ { .05, .2, .35, .5, .65, .8, .95}	.25, .5, .75	.1, .25, .45 where $Z_{12} \leq \min(Z_{11}, 1 - Z_{11})$	$Z_{12}\times${.05, .2, .35, .5, .65, .8, .95} where $Z_{22} \leq \min(Z_{21}, 1 - Z_{21})$	186

Chapter 3

Minimum Mandatory Security Spending and Social Welfare

3.1. Introduction

Firm value is often used in understanding of the impact of adverse events on a firm, especially with respect to security incidents (Cavusoglu et al. 2004, Acquisti et al. 2006). Examining customer reaction can provide a more direct measure of the impact. Firm value changes are often a result of the uncertainty regarding future costs of a breach (due to litigation and fines, for example). Customer reaction is more immediate than litigation costs or regulatory fines, even if demand changes are temporary. One annual study shows the cost in lost business of IT security breaches has held steady at over 63% of the average cost of an IT security breach for the last three years (Ponemon and Symantec 2011); in 2010 this translated to an average of over $4 US million per breach in lost business. If we assume that firms invest in information security to the extent required by existing laws already (in an attempt to minimize the risk of litigation and fines), what additional investments might be necessary to reduce the negative impact of customer reactions to security breaches?

We want to examine the impact of minimum security spending on not only firm profits, but also consumer surplus and total social welfare in order to better understand where it will be necessary to regulate this minimum spending level rather than rely on voluntary industry standardization. By understanding the mechanisms required to internalize the externalities of security breaches, we will be better able to balance consumers' privacy interests with the business needs. By modeling customer utility, we are able to answer the question: "How are consumer surplus and total social welfare affected when firms increase security spending beyond market equilibrium?"

In this research, we continue to model firm behaviour using the continuous-time Markov chain (CTMC) model developed in Chapter 2. Assuming rational consumers, we use a Hotelling setting to model customer utility for each firm's product. Prior work (Cezar et al. 2010, Kolfal et al. 2010), has assumed a specific form for customer reaction to security breaches in a two firm setting. However, by explicitly modeling customer response using a utility model, we are able to examine more general forms for demand changes. Thus, we are able to demonstrate the value in understanding how consumers react to security breaches based on industry characteristics. Further, understanding consumer reaction is necessary for firms to be able to calculate both

the direct- and cross-risk elasticities of demand introduced in Chapter 2.

We are able to link the customer utility model to firm decision making regarding information security investments. By using the customer utility model, we are able to derive risk elasticity parameters for each firm, which may then be used in the CTMC model of information security spending in a competitive environment as proposed in Chapter 2. By modeling consumer demand in this way, we are able to show not only that consumer surplus is always increasing when a minimum mandatory spending level above market equilibrium is introduced, but also that there are conditions under which total social welfare is also increasing. This work shows that a more fine-grained understanding how customers react when both firms have suffered a breach is important for understanding where total social welfare may be increasing in spending. We are able to find an upper-bound on mandatory spending after which regulation has a negative impact on total social welfare; that is, it is no longer in the best interest of society to continue increasing security requirements.

Companies will find an equilibrium level of behaviour (Jamal et al. 2003) that is acceptable to the customer, effectively balancing the risk of demand losses due to adverse events with the cost of preventing them. Further, with respect to privacy in eCommerce, there is no significant difference in the actual practices between firms in regulated (UK) and non-regulated (US) environments, suggesting that a middle ground can be found without regulation (Jamal et al. 2005).

In Chapter 2, we assumed an additive functional form for demand reaction when both firms are in the bad state, as shown in (2.11). Other researchers have made different assumptions regarding how customers react in this state; Cezar et al. (2010), for example, assume that there are no cross effects on demand in the state bb, just direct demand change effects. If we model consumer reactions explicitly, we see that there are a several possibilities for customer reaction in state bb, each of which has a different impact on how minimum security spending will affect firm profits and, thus, total social welfare. In this work, we are able to examine the customer reaction possibilities more closely, identifying how this affects the propositions from Chapter 2, and identify the viable functional forms for the change in demand at each state. We may then use this information to derive the direct and cross-risk elastities, obtaining a deeper and better understanding of how customer reaction affects firm spending.

The paper proceeds as follows; Section 3.2 presents our model including the Hotelling model of customer utility, Section 3.3 presents our findings regarding regulation and conclusions are presented in Section 3.4.

3.2. Model

In this section, we present the details of our model. As in Chapter 2, we use a continuous-time Markov chain (CTMC) to model the evolution of firms' operating states. In particular, successful security related attacks follow a Poisson process and are i.i.d. with arrival rate of λ_i for firm i, $i = 1, 2$. The effects of a successful security related attack last for a stochastic duration of time following an exponential distribution with expected length $1/\mu_i$. Thus, at any point in time, each firm may be in either a 'good' or 'bad' state. Combining all possibilities for two firms, we have four possible states, as illustrated in Figure 2.1.

We model demand changes in a manner similar to that of Chapter 2, but with one significant difference. We introduce an additional parameter, Z_3, to capture the demand change in the state bb. With this change, the normalized demand functions for firm i in the symmetric setting are provided below:

$$D_{i,gg} = 1, \quad \text{for } i = 1, 2 \tag{3.1}$$

$$D_{1,bg} = D_{2,gb} = 1 - Z_1, \tag{3.2}$$

$$D_{1,gb} = D_{2,bg} = 1 - Z_2. \tag{3.3}$$

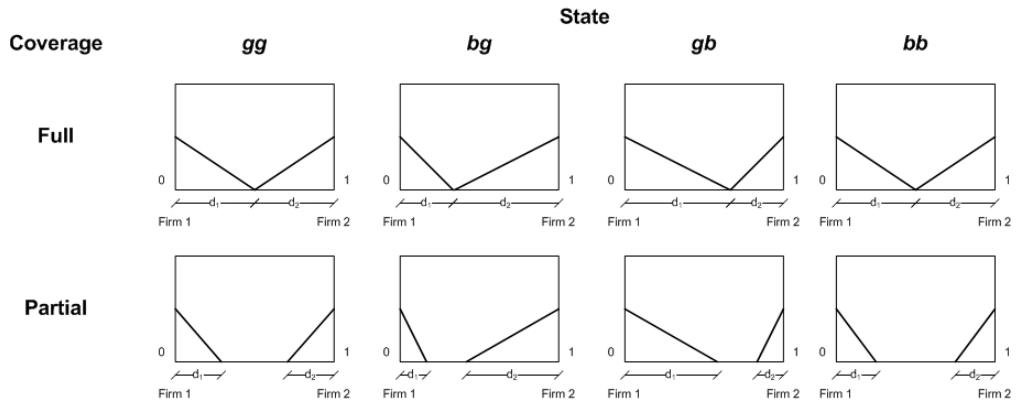
$$D_{i,bb} = 1 - Z_3 \quad \text{for } i = 1, 2 \tag{3.4}$$

where Z_1 is the percentage change in demand due to an adverse event in own firm and Z_2 is the percentage change in demand due to an adverse event in the other firm and Z_3 is the percentage change in demand due to an adverse event in both firms. In this setting, Z_1 is the *direct-risk elasticity of demand*, Z_2 is the *cross-risk elasticity of demand* and Z_3 is the *combined risk elasticity of demand*. Note that we can obtain the normalized demand equations given by (2.11) simply by setting $Z_3 = Z_1 + Z_2$; the demand in state bb from our model in Chapter 2 is just a special case of this more general model.

As we are concerned with understanding the demand effects of security breaches, we use customer utility models to derive the functional forms for Z_1, Z_2, and Z_3.

In particular, we model consumer behaviour as a Hotelling model where two firms are located at 0 and 1 on a horizontal axis. Consumers are located along the line between the two firms with unit density (total population is normalized to one) and they will select the product from the closest firm in the absence of adverse events. The market share for each firm is a function of the product's maximum utility to the consumer, product price, transportation cost, and each firm's state. When both firms are in a good state (state gg), the utility of the product offered by firm 1 to a customer at location $x \in [0,1]$ is given by the function $U_1 = \nu - t_d x$, where $\nu = u - p$ represents the products' inherent utility to the consumer, u, less the price of the product, p, and t_d is the "unit transportation cost" a customer incurs by having to travel to the firm's location. The utility of a product offered by firm 2 is then given by $U_2 = \nu - t_d(1-x)$. We can then solve this system of equations to find the market share, d_i, enjoyed by firm $i, i \in \{1,2\}$ in this state. If the difference between the utility and the price of the product is sufficiently high, the market will be fully covered, as shown in Figure 3.1 in the row *Full Coverage* and column gg, otherwise the market is only partially covered (Figure 3.1 row *Partial Coverage*, column gg). The necessary condition for full market coverage in this state is $\nu \geq \frac{t_d}{2}$. In the full coverage situation, we can solve for the location of the indifferent customer to find that the market share for each firm is 50% (that is, $d_1 = d_2 = .5$). In the partial market coverage situation, the market share for each is given by $d_1 = d_2 = \frac{\nu}{t_d}$.

Figure 3.1: Utility State Diagram.



The customer utility of buying from a firm depends on the state of both firms.

When there is an adverse security event at firm 1 only (state bg), there is an additional cost to the customer of that event (Figure 3.1, column bg). The utility

for firm 1's product to a customer becomes $U_1 = \nu - t_d x - T_1$ where $T_1 \geq 0$ is the cost to the consumer of an own-firm breach (the direct-breach cost). A breach in firm 1 may have indirect effects on the utility of firm 2's product, thus the utility for firm 2's product for this state is defined as $U_2 = \nu - t_d(1-x) - T_2$, where T_2 is the cost to the consumer of an other-firm breach (the cross-breach cost). There is no restriction on the sign of T_2. Once again, if the product's utility minus price remains high enough, the market will remain fully covered. The condition for full market coverage in this state is $\nu \geq \frac{t_d+T_1+T_2}{2}$. Solving for the indifferent customer in the full coverage situation (Figure 3.1 row *Full Coverage*, column bg), we find the market share for each firm is given by $d_1 = \frac{t_d-T_1+T_2}{2t_d}$ and $d_2 = \frac{t_d+T_1-T_2}{2t_d}$. In the partial market coverage situation, market share for each firm is $d_1 = \frac{\nu-T_1}{t_d}$ and $d_2 = \frac{\nu-T_2}{t_d}$. The state gb (Figure 3.1, column gb) is the mirror image of state gb, with the same condition for market coverage.

Finally, in the state bb (Figure 3.1, column bb), customers accrue both the cost of a direct breach (T_1) and the cost of a cross-breach (T_2), yielding utility functions for each firm of $U_1 = \nu - t_d x - (T_1 + T_2)$ and $U_2 = \nu - t_d(1-x) - (T_1 + T_2)$. The full market coverage condition in this state is $\nu \geq \frac{1}{2}t_d + T_1 + T_2$. Market share in the full coverage situation for each firm is $d_1 = d_2 = .5$ and in the partial coverage situation is $d_1 = d_2 = \frac{\nu-(T_1+T_2)}{t_d}$. The market shares and coverage conditions for all four states are given in Table 3.1.

We eliminate cases where total consumer utility increases with successful adverse events by using the following assumption.

Assumption 3.1. *Total consumer utility is not increasing when firms move from state gg to states bg or gb or from states bg or gb to state bb.*

We can calculate the percentage change in demand that each firm experiences as it transitions from one state to another. This percentage change in firm demand corresponds to the direct- and cross-risk elasticities (Z_1 and Z_2) in the model presented in Section 2.2. In order to analyze all possible coverage combinations presented in Figure 3.1, we allow the percentage change in demand as a result of being in state bb (denoted as Z_3) to be different from $Z_1 + Z_2$. Similar to our Assumption 2.1 that $Z_1 + Z_2 \geq 0$, we now require $Z_3 \geq 0$ to ensure demand bb is not greater than demand at gg. Thus, the assumptions of this model are:

Table 3.1: Market Coverage Conditions and Market Demand for each State.

State	Full Coverage Condition	Demand	
		Full	Partial
gg	$\nu \geq \frac{t_d}{2}$	$d_1 = .5$ $d_2 = .5$	$d_1 = \frac{\nu}{t_d}$ $d_2 = \frac{\nu}{t_d}$
bg	$\nu \geq \frac{t_d+T_1+T_2}{2}$	$d_1 = \frac{t_d-T_1+T_2}{2t_d}$ $d_2 = \frac{t_d+T_1-T_2}{2t_d}$	$d_1 = \frac{\nu-T_1}{t_d}$ $d_2 = \frac{\nu-T_2}{t_d}$
gb	$\nu \geq \frac{t_d+T_1+T_2}{2}$	$d_1 = \frac{t_d+T_1-T_2}{2t_d}$ $d_2 = \frac{t_d-T_1+T_2}{2t_d}$	$d_1 = \frac{\nu-T_2}{t_d}$ $d_2 = \frac{\nu-T_1}{t_d}$
bb	$\nu \geq \frac{1}{2}t_d + T_1 + T_2$	$d_1 = .5$ $d_2 = .5$	$d_1 = \frac{\nu-(T_1+T_2)}{t_d}$ $d_2 = \frac{\nu-(T_1+T_2)}{t_d}$

Assumption 3.2. $0 \leq Z_3 \leq 1$

Assumption 3.3. $Z_2 \leq Z_1$

Given the coverage possibilities presented in Figure 3.1, there are eight possible market coverage combinations as firms experience the full range of security breach states presented in Figure 2.1. For example, we might have the case FPP (where 'F' denotes full and 'P' denotes partial coverage) denoting full coverage at state gg, partial coverage at states bg and gb, and lastly partial coverage in state bb. With this notation, the eight cases are FFF, FFP, FPF, FPP, PFF, PFP, PPF, and PPP. Of these eight, in section 3.5.5 we show four are not possible (FPF, PFF, PFP, and PPF). The remaining four cases are discussed in sections 3.2.1-3.2.4 below.

3.2.1 Case 1-FFF

Here the market coverage is Full in all of the states. As presented in Table 3.1, the following case conditions must hold in order for us to have Full market coverage in all of the states:

$$\text{(i)}\ \ \nu \geq \frac{t_d}{2}, \quad \text{(ii)}\ \ \nu \geq \frac{t_d + T_1 + T_2}{2}, \quad \text{(iii)}\ \ \nu \geq \frac{t_d}{2} + T_1 + T_2 \tag{3.5}$$

Note that as demand is normalized to 1, at each state market coverage for a firm i, d_i, is the demand for that firm. For example, at state bg, the demand for firm 1, D_{bg}, is $d_1 = \frac{1}{2} - \frac{T_1-T_2}{2t_d}$, as presented in Table 3.1. Below we list the demand for firm

1 in each state.

$$\begin{aligned} D_{gg} &= \frac{1}{2} & D_{bg} &= \frac{1}{2} - \frac{T_1 - T_2}{2t_d} \\ D_{gb} &= \frac{1}{2} + \frac{T_1 - T_2}{2t_d} & D_{bb} &= \frac{1}{2} \end{aligned} \tag{3.6}$$

From the demand equations above, we can calculate the direct- and cross-risk elasticities of demand. For firm 1, Z_1 corresponds to the percentage change in demand between the states gg and bg, Z_2 corresponds to the percentage change in demand between the states gg and gb, and Z_3 corresponds to the percentage change in demand between the states gg and bb.

$$Z_1 = \frac{T_1 - T_2}{t_d}, \quad Z_2 = -\frac{T_1 - T_2}{t_d} = -Z_1, \quad Z_3 = 0 \tag{3.7}$$

From our Assumptions 3.2 and 3.3 placed on the direct- and cross- risk elasticities, we obtain the following conditions on the direct- and cross- breach costs, T_1 and T_2:

$$\begin{aligned} 0 \leq T_1 - T_2 \leq t_d \qquad & \text{since } Z_1 \in [0,1] \\ T_2 \leq T_1 \qquad & \text{from Assumption 3.3} \end{aligned}$$

The case conditions (3.5) can be used to further limit the direct and cross-risk elasticities of demand. The additional limitations are:

$$\begin{aligned} Z_1 &\geq \frac{T_1 - T_2}{2[\nu + (T_1 + T_2)]} \\ Z_2 &\leq -\frac{T_1 - T_2}{2[\nu + (T_1 + T_2)]} \\ Z_3 &= 0 \end{aligned}$$

In this case, Assumption 3.1, which states that consumer surplus be non-increasing in adverse security events, results in the following additional condition:

$$(T_1 + T_2) \geq Z_1^2 \frac{t_d}{2} \tag{3.8}$$

3.2.2 Case 2-FFP

Here the market coverage is Full in states gg, bg, and gb and Partial in state bb. As presented in Table 3.1, the following conditions must hold in order for us to have coverage as described in each of the states:

$$\text{(i)}\ \ \nu \geq \frac{t_d}{2}, \quad \text{(ii)}\ \ \nu \geq \frac{t_d + T_1 + T_2}{2}, \quad \text{(iii)}\ \ \nu < \frac{t_d}{2} + T_1 + T_2 \tag{3.9}$$

As presented in Table 3.1, we list the demand for firm 1 in each state:

$$\begin{aligned} D_{gg} &= \frac{1}{2} & D_{bg} &= \frac{1}{2} - \frac{T_1 - T_2}{2t_d} \\ D_{gb} &= \frac{1}{2} + \frac{T_1 - T_2}{2t_d} & D_{bb} &= \frac{\nu - (T_1 + T_2)}{t_d} \end{aligned} \tag{3.10}$$

From the demand equations above, we can calculate the direct- and cross-risk elasticities of demand:

$$Z_1 = \frac{T_1 - T_2}{t_d}, \quad Z_2 = -\frac{T_1 - T_2}{t_d} = -Z_1, \quad Z_3 = 1 - \frac{2(\nu - (T_1 + T_2))}{t_d} \tag{3.11}$$

Given case condition (3.9) *(iii)*, it is easily shown that $Z_3 > Z_1 + Z_2 = 0$.

Again, from Assumptions 3.2 and 3.3, we obtain the following conditions on the direct- and cross- breach costs:

$$\begin{aligned} 0 \leq T_1 - T_2 \leq t_d & \qquad \text{since } Z_1 \in [0,1] \\ T_2 \leq T_1 & \qquad \text{from Assumption 3.3} \\ \nu - t_d \leq T_1 + T_2 \leq \nu & \qquad \text{Since } Z_3 \geq 0 \end{aligned}$$

The case conditions (3.9) can be used to further limit the direct, cross, and combined-risk elasticities of demand as shown below:

$$\begin{aligned} \frac{T_1 - T_2}{2\nu - (T_1 + T_2)} &\leq Z_1 < \frac{T_1 - T_2}{2[\nu - (T_1 + T_2)]} \\ -\frac{T_1 - T_2}{2[\nu - (T_1 + T_2)]} &< Z_2 \leq -\frac{T_1 - T_2}{2\nu - (T_1 + T_2)} \\ 0 &< Z3 \leq \frac{T_1 - T_2}{2\nu - (T_1 + T_2)} \end{aligned}$$

In this case, Assumption 3.1, results in the following additional condition:

$$\frac{T_1 + T_2}{t_d} \geq 2Z_1^2 \Rightarrow Z_3 + \frac{2\nu}{t_d} - 1 \geq 4Z_1^2 \tag{3.12}$$

3.2.3 Case 3-FPP

Here the market coverage is Full in state gg and Partial in states bg, gb, and bb. As presented in Table 3.1, the following conditions must hold in order for us to have coverage as described in each of the states:

$$\text{(i)}\ \ \nu \geq \frac{t_d}{2}, \quad \text{(ii)}\ \ \nu < \frac{t_d + T_1 + T_2}{2}, \quad \text{(iii)}\ \ \nu < \frac{t_d}{2} + T_1 + T_2 \tag{3.13}$$

As presented in Table 3.1, we list the demand for firm 1 in each state:

$$\begin{aligned} D_{gg} &= \frac{1}{2} & D_{bg} &= \frac{\nu - T_1}{t_d} \\ D_{gb} &= \frac{\nu - T_2}{t_d} & D_{bb} &= \frac{\nu - (T_1 + T_2)}{t_d} \end{aligned} \tag{3.14}$$

From the demand equations above, we can calculate the direct-, cross-, and combined-risk elasticities of demand:

$$Z_1 = 1 - \frac{2[\nu - T_1]}{t_d}, \quad Z_2 = 1 - \frac{2[\nu - T_2]}{t_d}, \quad Z_3 = 1 - \frac{2[\nu - (T_1 + T_2)]}{t_d} \tag{3.15}$$

Given case condition (3.13) *(i)*, it can be shown that $Z_3 \geq Z_1 + Z_2$. From our Assumptions 3.2 and 3.3, we obtain the following conditions on the direct- and cross- breach costs:

$$\begin{aligned} \nu - \frac{t_d}{2} \leq T_1 \leq \nu \qquad & \text{since } Z_1 \in [0, 1] \\ \nu - t_d \leq T_2 \leq \nu \qquad & \text{since } Z_2 \in [-1, 1] \\ T_2 \leq T_1 \qquad & \text{From Assumption 3.3} \\ \nu - \frac{t_d}{2} \leq T_1 + T_2 \leq \nu \qquad & \text{from Assumption 3.2} \end{aligned}$$

The case conditions (3.13) can be used to further limit the direct-, cross- and combined-risk elasticities of demand. The additional limitations are:

$$\begin{aligned} \frac{T_1 - T_2}{2\nu - (T_1 + T_2)} &< Z_1 \leq \frac{T_1}{\nu} \\ -\frac{T_1}{[\nu - (T_1 + T_2)]} &< Z_2 \leq \frac{T_2}{\nu} \\ \frac{T_1 + T_2}{2\nu - (T_1 + T_2)} &< Z_3 \leq \frac{T_1 + T_2}{\nu} \end{aligned}$$

In this case, Assumption 3.1, results in the following additional condition:

$$\begin{aligned} (1 - Z_3)^2 &\leq \frac{2[t_d(4\nu - t_d) - 2(\nu^2 - 2T_1T_2)]}{t_d^2} \\ (1 - Z_3)^2 &\leq \frac{2(\nu^2 - 2T_1T_2)}{t_d^2(\nu - \frac{1}{2})} \end{aligned} \tag{3.16}$$

3.2.4 Case 4-PPP

Here the market coverage is Partial in all states. As presented in Table 3.1, the following conditions must hold in order for us to have Partial market coverage in each of the states:

$$\text{(i)}\ \ \nu < \frac{t_d}{2}, \quad \text{(ii)}\ \ \nu < \frac{t_d + T_1 + T_2}{2}, \quad \text{(iii)}\ \ \nu < \frac{t_d}{2} + T_1 + T_2 \tag{3.17}$$

As presented in Table 3.1, we list the demand for firm 1 in each state:

$$\begin{aligned} D_{gg} &= \frac{\nu}{t_d} & D_{bg} &= \frac{\nu - T_1}{t_d} \\ D_{gb} &= \frac{\nu - T_2}{t_d} & D_{bb} &= \frac{\nu - (T_1 + T_2)}{t_d} \end{aligned} \tag{3.18}$$

From the demand equations above, we can calculate the direct-, cross- and combined-risk elasticities of demand:

$$Z_1 = \frac{T_1}{\nu}, \quad Z_2 = \frac{T_2}{\nu}, \quad Z_3 = \frac{T_1 + T_2}{\nu} \tag{3.19}$$

One again, from Assumptions 3.2 and 3.3, we obtain the following conditions on the direct- and cross- breach costs:

$$\begin{aligned} 0 \le T_1 \le \nu \qquad & \text{since } Z_1 \in [0, 1] \\ -\nu \le T_2 \le \nu \qquad & \text{since } Z_2 \in [-1, 1] \\ T_2 \le T_1 \qquad & \text{From Assumption 2.2} \\ 0 \le T_1 + T_2 \le \nu \qquad & \text{From Assumption that } D_{gg} \ge D_{bb} \end{aligned}$$

The case conditions (3.13) can be used to further limit the direct-, cross-, and combined-risk elasticities of demand as follows:

$$\begin{aligned} Z_1 &> \frac{2T_1}{t_d} \\ Z_2 &> \frac{2T_2}{t_d} \\ Z3 &> \frac{2(T_1 + T_2)}{t_d} \end{aligned}$$

In this case, Assumption 3.1, results in the following additional condition:

$$Z_2 \ge \frac{1 - 2Z_1}{2(1 - Z_1)} \tag{3.20}$$

3.3. Regulations and the Consumer

For each of the four possible cases (FFF, FFP, FPP, PPP) we are able to show that consumer surplus is increasing under regulation.

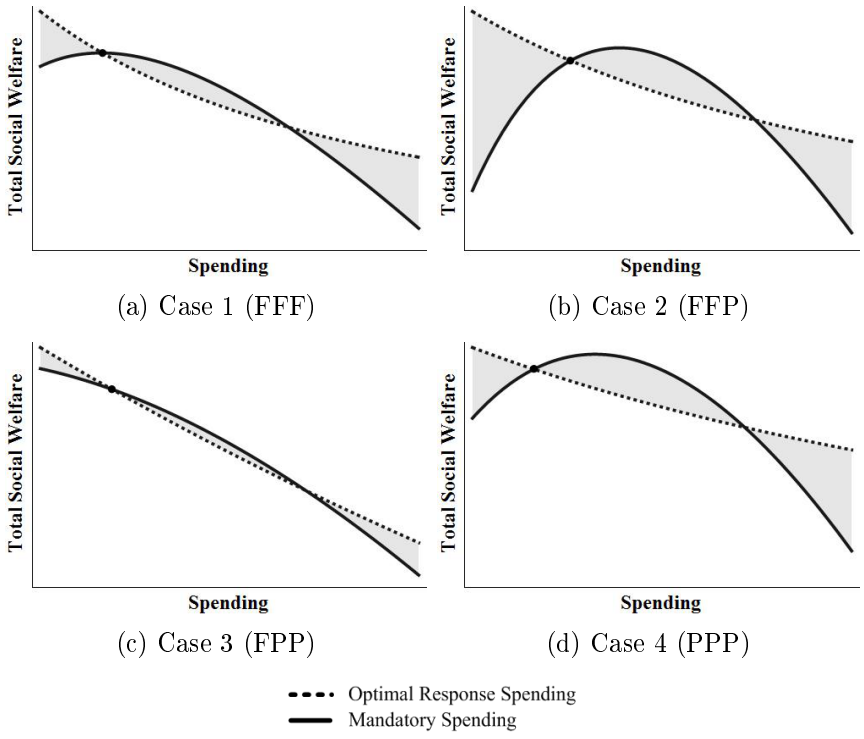
Proposition 3.1. *Consumer surplus is always increasing when a minimum mandatory spending level is introduced.*

Proof: Proof of the Proposition 3.1 is presented in Appendix 3.5. We are also able to derive the conditions under which total social welfare in increasing for each of the four cases.

Proposition 3.2. *When firms are substitutes in loss or unaffected, ($Z_2 \leq 0$) conditions exist under which total social welfare is increasing when a minimum mandatory spending level is introduced. These conditions are given by equations (3.29), (3.40), (3.48), and (3.59) in Appendix 3.5:*

Proof: Proof of the Proposition 3.2 is presented in Appendix 3.5.

Figure 3.2: Total Social Welfare curves when firms are substitutes in loss.



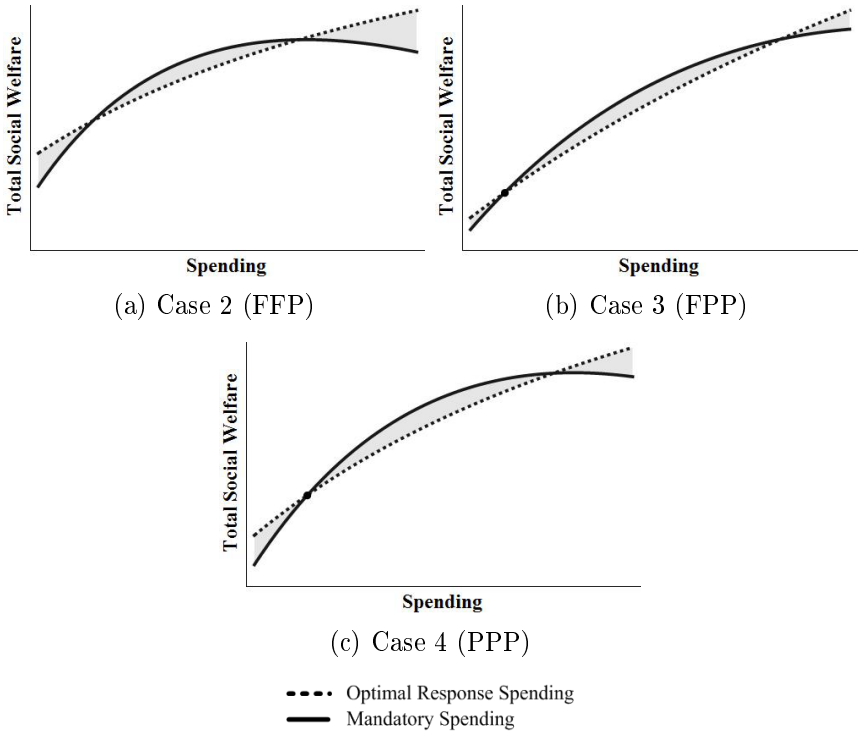
(a) Case 1 (FFF) (b) Case 2 (FFP)

(c) Case 3 (FPP) (d) Case 4 (PPP)

Proposition 3.3. *When firms are complements in loss ($Z_2 > 0$), total social welfare increases when a minimum mandatory spending level is introduced.*

Proof: Proof of the Proposition 3.3 follows directly from Propositions 2.1 and 3.1.

Figure 3.3: Total Social Welfare curves when firms are complements in loss.



(a) Case 2 (FFP)

(b) Case 3 (FPP)

(c) Case 4 (PPP)

In Chapter 2 we showed that, when firms are complements in loss, firm profits increase when minimum security spending levels are set appropriately. It is unsurprising, then, that when firms are complements in loss, total social welfare will also increase with minimum security spending since we have already shown that consumer welfare is always increasing in this case. What is interesting, is that we can show that despite a decrease in firm profits, total social welfare may also increase for a range of increased spending when firms are substitutes in loss. In general, such results arise when the increase in consumer welfare outpaces the decrease in firm profits as a result of mandatory spending above equilibrium.

3.4. Conclusion

By including consumer utility in our model, we are able to examine the effects of regulations or standardization on consumer surplus and total social welfare. A significant contribution of this paper is to allow unique insights regarding customer demand changes in response to adverse IT security events, and the impact these demand changes have on security spending and profits. Consumer surplus always

increases when minimum security spending is mandated either through regulations or voluntary industry standardization. We are able to obtain the conditions under which total social welfare is increasing, showing that regulation and industry standardization —both activities which introduce a minimum level of security spending —(*i*) will increase consumer surplus, (*ii*) can increase total social welfare for the cases analyzed in Proposition 2.1, even though firm profits are decreasing, and (*iii*) can increase both the firm profits and the total social welfare for the cases analyzed in Proposition 2.2. However, increasing total social welfare is dependent on limiting this minimum mandatory spending to a level that is not too expensive for firms. How restrictive the range on increased spending is depends greatly on industry characteristics. Thus, it is important to consider industry dynamics when making policy recommendations.

An interesting discovery from this work is that a more fine-grained understanding of how customers react in the bb state can really drive the results in terms of the conditions for which firms might be better off cooperating by jointly increasing security investments, even when firms are substitutes in loss. In all cases, however, we can find an upper-bound after which point regulation has a negative impact on total social welfare; where it is no longer in the best interest of society to continue increasing security investment requirements. In our work, we weight the consumer surplus and firm profits equally; a central planner could adjust these weights to account for social preferences, finding a different upper-bound on mandatory spending.

A firm may use our customer utility model to estimate the appropriate direct- and cross-risk elasticities. We assume that firms understand their customers' utility functions when no firm has experienced an adverse security event, thus the only new information to gather is how customers assess costs for security breaches in each firm. Firms need only observe how market shares change after security events in order to gather the necessary information. From this information, the direct- and cross-risk elasticity parameters can be estimated.

Measuring the impact of cross-risk elasticity directly is more difficult. In this case, a firm must correlate the customer churn and growth with events that occur outside the firm; information that may be hidden by their competitor. Future research should examine whether firms have an incentive to hide private information such as the frequency of successful security breaches, customer churn, and the duration of

demand change effects.

Firms seem unaware of the costs associated with security breaches at competitor firms. Governmental policy can encourage or require disclosure of security breaches which will help companies connect the dots, understanding how adverse events at other firms affect their demand. An example of such a regulation is California's breach disclosure law (SB-1386 2002). Enabling firms to establish causality of demand changes due to adverse events at other firms is an important policy initiative that must be encouraged. Given that many data breaches are transborder (e.g. the TJX data breach affected customers in not only the US but also Canada and the UK (Vijayan 2007)), international cooperation is vitally necessary.

Additional future research possibilities include an examination of asymmetric cases where the arrival rates of IT security events and the expected duration of these events are different for each firm. As an example of asymmetric cases, the interaction between firms with significant market power imbalances will affect IT security spending could be considered. If one firm is a dominant player, they might face a different risk profile (riskiness of the environment) than a smaller player. It is not hard to imagine there might be a difference - consider how Microsoft has traditionally attracted more attention by computer attackers than Apple (Chen et al. 2005). While Chapter 2 provided a numerical analysis of asymmetric firms, adapting this work to a periodic setting may allow for analytical results in this area.

Our use of the Hotelling model is designed to capture the behaviour of fully rational individuals. However, reaction to the loss of data may be more likely to be an emotional response in many cases. How, then, to best capture the idea that it's not rational to stop purchasing a product you need in light of a security incident - after all, the personally identifiable information has already been lost. Further, if firms are in no way equally susceptible to a particular security attack, is it rational for the customers of an unbreached firm to stop purchasing as well (the complements in loss case)? It would be interesting to see if customers are able to identify industries (firms) that are at risk of correlated events and which are not, then to examine if they rationally react in each of these situations.

Bibliography

Acquisti, A., A. Friedman, R. Telang. 2006. Is there a cost to privacy breaches? an event study. *ICIS 2006 Proceedings*.

Cavusoglu, H., B. Mishra, S. Raghunathan. 2004. The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce* **9** 70–104.

Cezar, A., H. Cavusoglu, S. Raghunathan. 2010. Competition, speculative risks, and IT security outsourcing. T. Moore, D. Pym, C. Ioannidis, eds., *Economics of Information Security and Privacy*. Springer US, Boston, MA, 301–320.

Chen, P., G. Kataria, R. Krishnan. 2005. Software diversity for information security.

Jamal, K., M. Maier, S. Sunder. 2003. Privacy in e-Commerce: Development of reporting standards, disclosure, and assurance services in an unregulated market. *Journal of Accounting Research* **41** 285–309.

Jamal, K., M. Maier, S. Sunder. 2005. Enforced standards versus evolution by general acceptance: A comparative study of e-Commerce privacy disclosure and practice in the United States and the United Kingdom. *Journal of Accounting Research* **43** 73–96.

Kolfal, B., R. Patterson, M.L. Yeo. 2010. Market impact on IT security spending.

Ponemon, Symantec. 2011. 2010 annual study: U.S. cost of data breach .

SB-1386. 2002. California Senate Bill: Personal information: privacy. `http://info.sen.ca.gov/pub/01-02/bill/sen/sb_1351-1400/sb_1386_bill_20020926_chaptered.html`.

Vijayan, J. 2007. TJX data breach: At 45.6m card numbers, it's the biggest ever. *Computer World* .

3.5. Appendix

The appendix includes proofs of the propositions as well as additional detail regarding the conditions under which total social welfare is increasing. Information is organized by case.

3.5.1 Case 1-FFF

Consumer Surplus under Regulation

In case FFF, we now show that consumer surplus is increasing under regulation.

Proof: Consumer surplus at equilibrium is given by:

$$CS^e = \frac{Q}{4(c^e+\rho)^2 t_d}\left[-(c^e+\rho)^2 t_d^2 + 2c^e\rho(T_1-T_2)^2 + 4(c^e+\rho)t_d(\nu(c^e+\rho)-\rho(T_1+T_2))\right] \tag{3.21}$$

We need to show that when $R \geq c^e$

$$CS^R - CS^e = \frac{Q(c^e-R)\rho}{2(c^e+\rho)^2(R+\rho)^2 t_d}\left[(c^eR-\rho^2)(T_1-T_2)^2 - 2(c^e+\rho)(R+\rho)t_d(T_1+T_2)\right] \geq 0 \tag{3.22}$$

We know $\frac{Q(c^e-R)\rho}{2(c^e+\rho)^2(R+\rho)^2 t_d} \leq 0$, so we must show

$$\begin{aligned}(c^eR-\rho^2)(T_1-T_2)^2 &\leq 2(c^e+\rho)(R+\rho)t_d(T_1+T_2)\\ (c^eR-\rho^2)Z_1^2 &\leq (c^eR+\rho^2+(c^e+R)\rho)2\left(Z_1+\frac{2T_2}{t_d}\right)\end{aligned} \tag{3.23}$$

(3.23) holds as:

$$\begin{aligned}(c^eR-\rho^2) &\leq (c^eR+\rho^2+(c^e+R)\rho),\\ Z_1^2 &\leq 2Z_1 \qquad \text{as } Z_1 \in [0,1], \text{ and}\\ (c^eR+\rho^2+(c^e+R)\rho)2\left(Z_1+\frac{2T_2}{t_d}\right) &\geq 0\end{aligned}$$

Total Social Welfare under Regulation

In case FFF, we now provide the conditions for which total social welfare increases for a range under regulation.

Proof: Total social welfare at equilibrium is given by

$$TSW^e = -\frac{Q((c^e+\rho)^2 t_d^2 - 2c^e\rho(T_1-T_2)^2 + 4(c^e+\rho)t_d((c^e-\pi-\nu)(c^e+\rho)+\rho(T_1+T_2)))}{4(c^e+\rho)^2 t_d} \tag{3.24}$$

Total social welfare under regulation, where $R \geq c^e$ is given by

$$TSW^R = -\frac{Q((R+\rho)^2 t_d^2 - 2R\rho(T_1-T_2)^2 + 4(R+\rho)t_d((R-\pi-\nu)(R+\rho)+\rho(T_1+T_2)))}{4(R+\rho)^2 t_d} \tag{3.25}$$

The difference is given by

$$\begin{aligned}&TSW^R - TSW^e =\\ &\frac{-Q(c^e-R)(\rho(c^eR-\rho^2)(T_1-T_2)^2 + 2(c^e+\rho)(R+\rho)t_d((c^e+\rho)(R+\rho)-\rho(T_1+T_2)))}{2(c^e+\rho)^2(R+\rho)^2 t_d}\end{aligned} \tag{3.26}$$

We can examine the sign of the derivative of (3.26) to determine whether total social welfare is increasing under regulation.

$$\frac{\partial(TSW^R - TSW^e)}{\partial R} = \frac{Q}{2(R+\rho)^3 t_d}\left(-(R-\rho)\rho(T_1-T_2)^2 - 2(R+\rho)t_d((R+\rho)^2 - \rho(T_1+T_2))\right) \tag{3.27}$$

We want to show that (3.27) is positive at $R = c^e$. By substituting $(T_1 - T_2) = Z_1 t_d$ and the the non-increasing consumer surplus condition given by (3.8) into (3.27), we find:

$$\frac{\partial(TSW^R - TSW^e)}{\partial R} \geq \frac{Q}{2(R+\rho)^3 t_d}\left(-(R-\rho)\rho(Z_1 t_d)^2 - 2(R+\rho)t_d((R+\rho)^2 - \frac{\rho Z_1^2 t_d}{2})\right) \tag{3.28}$$

If we can show that the right hand side of (3.28), evaluated at $R = c^e$ is positive, then we know that (3.27) is positive as well, and total social welfare is increasing under regulation.

$$\frac{Q}{2(c^e+\rho)^3 t_d}\left(-(c^e-\rho)\rho(Z_1 t_d)^2 - 2(c^e+\rho)t_d((c^e+\rho)^2 - \frac{\rho Z_1^2 t_d}{2})\right) \geq 0$$

We know $\frac{Q}{2(c^e+\rho)^3 t_d} > 0$ since $Q, \rho, t_d > 0$, and $c^e >= 0$, thus we need only show

$$-(c^e-\rho)\rho(Z_1 t_d)^2 - 2(c^e+\rho)t_d((c^e+\rho)^2 - \frac{\rho Z_1^2 t_d}{2}) \geq 0$$

Let $\rho C = c^e + \rho$. Then $C^2 = Z_1(\frac{\pi+\rho}{\rho} - C)$ in this case.

$$\begin{aligned}
-(c^e-\rho)\rho(Z_1 t_d)^2 - 2\rho C t_d((\rho C)^2 - \frac{\rho Z_1^2 t_d}{2}) &\geq 0 \\
-(C-2)\rho^2 Z_1^2 t_d^2 - 2\rho^3 C^3 t_d + \rho^2 C t_d^2 Z_1^2 &\geq 0 \\
Z_1^2 t_d - \rho C^3 &\geq 0 \\
Z_1^2 t_d &\geq \rho Z_1\left(Z_1(C - \frac{\pi+\rho}{\rho}) + C\frac{\pi+\rho}{\rho}\right) \\
Z_1 t_d &\geq \rho Z_1 C - \rho Z_1 \frac{\pi+\rho}{\rho} + \rho C \frac{\pi+\rho}{\rho} \\
Z_1 t_d &\geq \rho C(Z_1 + \frac{\pi+\rho}{\rho}) - \rho Z_1 \frac{\pi+\rho}{\rho} \\
Z_1 t_d &\geq (c^e+\rho)(Z_1 + \frac{\pi+\rho}{\rho}) - \rho Z_1 \frac{\pi+\rho}{\rho} \\
\rho\left(\frac{Z_1(t_d+\pi+\rho)}{\rho Z_1 + \pi + \rho} - 1\right) &\geq c^e \qquad (3.29)
\end{aligned}$$

Where (3.29) holds, total social welfare is increasing under regulation.

3.5.2 Case 2-FFP

Consumer Surplus under Regulation

In case FFP, we now show that consumer surplus is increasing under regulation.

Proof Consumer surplus at equilibrium is given by:

$$CS^e = \frac{Q\left(-c^e(c^e+2\rho)t_d^2+4c^e t_d(\nu(c^e+2\rho)-\rho(T_1+T_2))+2\rho(c^e(T_1-T_2)^2+2\rho(-\nu+T_1+T_2)^2)\right)}{4(c^e+\rho)^2 t_d} \tag{3.30}$$

We need to show that when $R \geq c^e$

$$\begin{aligned} &CS^R - CS^e = \\ &\frac{Q(R-c^e)\rho}{4(c^e+\rho)^2(R+\rho)^2}\left(4\nu(c^e+\rho)(R+\rho)+t_d(-2(c^e+\rho)(R+\rho)\right. \\ &\left.+(-2c^eR+2\rho^2)Z_1^2+Z_3(2(c^e+\rho)(R+\rho)-\rho(c^e+R+2\rho)Z_3))\right) \end{aligned} \tag{3.31}$$

Where Z_1, Z_2, and Z_3 are given in (3.11). We know $\frac{Q(R-c^e)\rho}{4(c^e+\rho)^2(R+\rho)^2} \geq 0$ since $R-c^e \geq 0$, we need to show:

$$\begin{aligned} &4\nu(c^e+\rho)(R+\rho) \\ &+t_d\left(-2(c^e+\rho)(R+\rho)+(-2c^eR+2\rho^2)Z_1^2+Z_3(2(c^e+\rho)(R+\rho)-\rho(c^e+R+2\rho)Z_3)\right) \geq 0 \end{aligned}$$

Let $A = c^eR + c^e\rho + \rho R + \rho^2 = (c^e+\rho)(R+\rho)$,

$$\begin{aligned} 4\nu A + t_d[-2A+(\rho^2-c^eR)2Z_1^2+Z_3(2A-(c^e\rho+\rho R+2\rho^2)Z_3)] &\geq 0 \\ 4\frac{\nu}{t_d}A-2A+(\rho^2-c^eR)2Z_1^2+2AZ_3-(c^e\rho+\rho R+2\rho^2)Z_3^2 &\geq 0 \end{aligned}$$

Let $B = c^e\rho + \rho R + 2c^eR$ (Thus, $A - B = \rho^2 - c^eR$)

$$4\frac{\nu}{t_d}A-2A+(A-B)2Z_1^2+2AZ_3-(2A-B)Z_3^2 \geq 0,$$

divide by $2A$

$$\begin{aligned} 2\frac{\nu}{t_d}-1+\frac{A-B}{2A}2Z_1^2+Z_3-\frac{2A-B}{2A}Z_3^2 &\geq 0 \\ 2\frac{\nu}{t_d}-1+Z_3+\frac{2A-B}{2A}2Z_1^2-Z_1^2-\frac{2A-B}{2A}Z_3^2 &\geq 0 \\ 2\frac{\nu}{t_d}-1+Z_3-Z_1^2+\frac{2A-B}{2A}(2Z_1^2-Z_3^2) &\geq 0 \end{aligned} \tag{3.32}$$

Since $-1 \leq \frac{A-B}{A} \leq 1$, this implies that $0 \leq \frac{2A-B}{2A} \leq 1$.

If $\frac{2A-B}{2A} = 0$, then $\frac{2\nu}{t_d} - 1 + Z_3 \geq Z_1^2$,which is true by (3.12) above.

If $\frac{2A-B}{2A} = 1$, then $\frac{2\nu}{t_d} - 1 + Z_3 - Z_1^2 + 2Z_1^2 - Z_3^2 \geq 0$ and, as $\frac{2\nu}{t_d} \geq 1$ (Case Condition (i)) and $Z_3 \geq Z_3^2$, this also holds. A complete proof follows:

If $\frac{2A-B}{2A} = x$, $x \in [0,1]$, then $\frac{2\nu}{t_d} - 1 + Z_3 - Z_1^2 + 2xZ_1^2 - xZ_3^2 \geq 0$.

If $x \geq \frac{1}{2}$, then:

$$\frac{2\nu}{t_d} - 1 + Z_3 - xZ_3^2 - (1-2x)Z_1^2 \geq 0,$$

As

$$\begin{aligned} \frac{2\nu}{t_d} &\geq 0 \qquad \text{By Case condition (i),} \\ 1 \geq Z_3 &\geq Z_3^2, \qquad \text{and} \\ 1 - 2x &\leq 0 \end{aligned}$$

If $x \leq \frac{1}{2}$, then:

$$\frac{2\nu}{t_d} - 1 + Z_3 - xZ_3^2 - (1-2x)Z_1^2 \geq 0 \tag{3.33}$$

By condition (3.12):

$$\begin{aligned} \frac{1}{4}(\frac{2\nu}{t_d} - 1 + Z_3) &\geq Z_1^2, \\ (1-2x)\frac{1}{4}(\frac{2\nu}{t_d} - 1 + Z_3) &\geq (1-2x)Z_1^2, \end{aligned}$$

thus, we can rewrite (3.33) as

$$(1-2x)\frac{1}{4}(\frac{2\nu}{t_d} - 1 + Z_3) - (1-2x)Z_1^2 + (1 - \frac{1-2x}{4})(\frac{2\nu}{t_d} - 1 + Z_3) - xZ_3^2 \geq 0 \tag{3.34}$$

Since $(1-2x)\frac{1}{4}(\frac{2\nu}{t_d} - 1 + Z_3) - (1-2x)Z_1^2 \geq 0$ and $\frac{2\nu}{t_d} - 1 \geq 0$, it remains for us to show:

$$\begin{aligned} (1 - \frac{1-2x}{4})Z_3 - xZ_3^2 &\geq 0 \\ (\frac{3-2x}{4})Z_3 &\geq xZ_3^2 \end{aligned}$$

Which always holds, as

$$\begin{aligned} (\frac{3-2x}{4}) \geq x \Rightarrow \frac{3}{2} \geq x, \text{ and} \\ Z_3 \geq Z_3^2 \end{aligned} \tag{3.35}$$

Total Social Welfare under Regulation

In case FFP, we now provide the conditions for which total social welfare increases for a range under regulation.

Proof: Total social welfare at equilibrium is given by

$$\begin{aligned} TSW^e =& \frac{Q}{4(c^e+\rho)^2}\big(-4(c^e+\rho)(c^{2e}-c^e(\pi+\nu-\rho)-\pi\rho) \\ &+t_d(-c^{2e}+2c^e\rho(Z_1^2-Z_3)+\rho^2(-1+Z_3)^2)+4(c^e-\pi)\rho^2Z_3\big) \end{aligned} \tag{3.36}$$

where Z_1, Z_2, and Z_3 are given in (3.11). Total social welfare under regulation, where $R \geq c^e$ is given by

$$\begin{aligned} TSW^R =& \frac{Q}{4(R+\rho)^2}\big(-4(R+\rho)(R^2-R(\pi+\nu-\rho)-\pi\rho) \\ &+t_d(-R^2+2R\rho(Z_1^2-Z_3)+\rho^2(-1+Z_3)^2)+4(R-\pi)\rho^2Z_3\big) \end{aligned} \tag{3.37}$$

The difference is given by

$$\begin{aligned} TSW^R - TSW^e =& \frac{Q(c^e-R)}{4(c^e+\rho)^2(R+\rho)^2}\big[4(c^e+\rho)(R+\rho)(c^e(R+\rho)+\rho(R-\nu+\rho)) \\ &+\rho\{-4\rho(R\pi+c^e(-R+y)+2\pi\rho+\rho^2)Z_3+t_d(2(c^e+\rho)(R+\rho) \\ &+2(c^eR-\rho^2)Z_1^2+Z_3(-2(c^e+\rho)(R+\rho)+\rho(c^e+R+2\rho)Z_3))\}\big] \end{aligned} \tag{3.38}$$

We can examine the sign of the derivative of (3.38) to determine whether total social welfare is increasing under regulation.

$$\begin{aligned} \frac{\partial(TSW^R-TSW^e)}{\partial R} =& \frac{-Q}{2(R+\rho)^3}\big(2(R+\rho)(R^2+2R\rho-\nu\rho+\rho^2)+2(R-2\pi-\rho)\rho^2Z_3 \\ &+\rho t_d(R+\rho+(R-\rho)Z_1^2-(R+\rho)Z_3+\rho Z_3^2)\big) \end{aligned} \tag{3.39}$$

If we let $R = c^e$ in (3.39) then solve for roots, we can show that $\frac{\partial(TSW^R-TSW^e)}{\partial R} \geq 0$ when

$$\begin{aligned} c^e \leq& \frac{1}{2}(-\rho(2+t_d(1+Z_1^2-Z_3))-2\rho^2Z_3 \\ &+\sqrt{\rho(\rho(2+t_d(1+Z_1^2-Z_3))+2\rho Z_3)^2+4(\nu-\rho+2\rho(2\pi+\rho)Z_3+\rho t_d(-1+Z_1^2+Z_3-Z_3^2))}) \end{aligned} \tag{3.40}$$

Thus, when (3.40) holds we know that total social welfare is increasing at $R = c^e$.

3.5.3 Case 3-FPP

Consumer Surplus under Regulation

In case FPP, we now show that consumer surplus is increasing under regulation.

Proof: Consumer surplus at equilibrium is given by:

$$CS^e = \frac{Q(c^{2e}(4\nu - t_d)t_d + 4\rho^2(-\nu + T_1 + T_2)^2 + 4c^e\rho(T_1^2 + T_2^2 - 2\nu(-\nu + T_1 + T_2)))}{4(c^e + \rho)^2 t_d} \tag{3.41}$$

We need to show that, when $R \geq c^e$, $CS^R - CS^e \geq 0$. That is, we need to show:

$$\begin{aligned}\frac{-Q(R - c^e)}{4(c^e + \rho)^2(R + \rho)^2}[&-4\nu(R\rho + c^e(R + \rho)) + (c^e(R + \rho) \\ &+ \rho(R - ((Z_1 - 2)Z_1 + (Z_2 - 1)^2 + 2Z_3 - Z_3^2)\rho))t_d] \geq 0\end{aligned} \tag{3.42}$$

We know $\frac{-Q(R-c^e)}{4(c^e+\rho)^2(R+\rho)^2} \leq 0$, thus we need only show

$$-4\nu(R\rho + c^e(R+\rho)) + (c^e(R+\rho) + \rho(R - ((Z_1 - 2)Z_1 + (Z_2 - 1)^2 + 2Z_3 - Z_3^2)\rho))t_d \leq 0$$

We let $A = (c^e + \rho)(R + \rho)$ and rearrange to get

$$\begin{aligned}(A - \rho^2)t_d - \rho^2 t_d[(Z_1 - 2)Z_1) + (Z_2 - 2)Z_2 + (2 - Z_3)Z_3 + 1] &\leq 4\nu(A - \rho^2) \\ -\rho^2 t_d[(Z_1 - 2)Z_1) + (Z_2 - 2)Z_2 + (2 - Z_3)Z_3 + 1] &\leq (A - \rho^2)(4\nu - t_d)\end{aligned}$$

We can show $(A - \rho^2)(4\nu - t_d) \geq 0$ as $4\nu - t_d = 2(2\nu - t_d/2) \geq 0$ by case condition (3.13) *(i)*, thus we want to show:

$$-\rho^2 t_d[(Z_1 - 2)Z_1) + (Z_2 - 2)Z_2 + (2 - Z_3)Z_3 + 1] \leq 0 \tag{3.43}$$

As we know $Z_3 \geq Z_1 + Z_2$ in this case, let us examine (3.43)

$$\begin{aligned}-(Z_1 - 2)Z_1) + (Z_2 - 2)Z_2 + (2 - Z_3)Z_3 + 1 &\geq 0 \\ Z_1^2 - 2(Z_1 + Z_2) + Z_2^2 - Z_3^2 + 2Z_3 + 1 &\geq 0 \\ Z_1^2 - 2(Z_1 + Z_2) + Z_2^2 - Z_3^2 + 2Z_3 + 1 &\geq Z_1^2 - 2Z_3 + Z_2^2 - Z_3^2 + 2Z_3 + 1 \\ &\geq 0 \quad \text{as } Z_3 \geq Z_1 + Z_2, \text{ thus} \\ 1 + Z_1^2 + ZZ_2^2 &\geq Z_3^2 \\ 1 + (Z_1 + Z_2)^2 - 2Z_1Z_2 &\geq 1 + Z_3^2 - 2Z_1Z_2 \\ &\geq Z_3^2 \text{ since } Z_3 \leq 1,\ Z_3^2 \leq Z_3, \text{ thus} \\ 1 - 2Z_1Z_2 &\geq 0\end{aligned} \tag{3.44}$$

Since (3.44) clearly holds for $Z_2 \leq 0$, let us show it is also true for $Z_2 > 0$

$$\begin{aligned}1 &\geq 2Z_1Z_2 \\ 1 \geq 2Z_1(1 - Z_1) &\geq 2Z_1Z_2 \quad \text{as } Z_1 + Z_2 \leq 1\end{aligned} \tag{3.45}$$

As $Z_1(1-Z_1)$ attains a maximum value of 0.25 when $Z_1 = .5$, we know that (3.45) holds and thus (3.43) holds. Thus, consumer surplus is increasing under regulation.

Total Social Welfare under Regulation

In case FPP, we now provide the conditions for which total social welfare increases for a range under regulation.

Proof: Total social welfare at equilibrium is given by

$$\begin{aligned} TSW^e = &\frac{Q}{4(c^e+\rho)^2 t_d}\big(-8c^e(c^e-\pi)\nu\rho + t_d(-4c^{e3} + c^e\rho(t_d(2+(-2+Z_1)Z_1 \\ &+(-2+Z_2)Z_2) - 4(\pi-\rho)(-1+Z_3)) + \rho^2(-4\pi + t_d(-1+Z_3))(-1+Z_3) \\ &+c^{e2}(-t_d + 4(\pi+\nu-\rho+\rho Z_3))))\big) \end{aligned} \tag{3.46}$$

Total social welfare under regulation, where $R \geq c^e$ is given by

$$\begin{aligned} TSW^R = &\frac{Q}{4(R+\rho)^2 t_d}\big(-8R(R-\pi)\nu\rho + t_d(-4R^3 + R\rho(t_d(2+(-2+Z_1)Z_1 \\ &+(-2+Z_2)Z_2) - 4(\pi-\rho)(-1+Z_3)) + \rho^2(-4\pi + t_d(-1+Z_3))(-1+Z_3) \\ &+R^2(-t_d + 4(\pi+\nu-\rho+\rho Z_3))))\big) \end{aligned} \tag{3.47}$$

We can examine the sign of the derivative of $TSW^R - TSW^e$ to determine whether total social welfare is increasing under regulation.

$$\begin{aligned} &\frac{\partial(TSW^R - TSW^e)}{\partial R} = \\ &\frac{Q}{4(R+\rho)^3 t_d}(8\nu\rho(\pi\rho - R(\pi+2\rho)) \\ &+ t_d(-4(R^3 + R(3R-\pi-2\nu)\rho + (R+\pi)\rho^2 + \rho^3) + 4\rho(R+\rho)(\pi+\rho)Z_3 \\ &+\rho t_d(-4R - (R-\rho)((-2+Z_1)Z_1 + (-2+Z_2)Z_2) + 4\rho Z_3 - 2\rho Z_3^2)))\big) \end{aligned} \tag{3.48}$$

If we let $R = c^e$ in (3.48) then solve for roots, we can show that there is a condition on c^e that will result in $\frac{\partial(TSW^R - TSW^e)}{\partial R} \geq 0$. With Mathematica, we have verified that there is only one positive root and that as long as c^e is less than this root, total social welfare is increasing.

3.5.4 Case 4-PPP

Consumer Surplus under Regulation

In case PPP, we now show that consumer surplus is increasing under regulation.

Proof: Consumer surplus at equilibrium is given by:

$$\begin{aligned} CS^e =& \frac{Q}{2(c^e+\rho)^2 t_d}\left(-c^{2e}(2\nu^2 - 4\nu t_d + t_d^2)\right. \\ &\left. +2\rho^2(-\nu+T_1+T_2)^2 + 2c^e\rho(2\nu^2 - 2\nu T_1 + T_1^2 - 2\nu T_2 + T_2^2)\right) \end{aligned} \tag{3.49}$$

We need to show that, when $R \geq c^e$, $CS^R - CS^e \geq 0$. That is, we need to show:

$$\begin{aligned} &\frac{Q(c^e-R)\rho}{(c^e+\rho)^2(R+\rho)^2 t_d}\left[(c^e+\rho)(R+\rho)T_1^2 - (c^e+\rho)(R+\rho)(2\nu - T_2)T_2\right. \\ &\qquad \left. +2T_1(-\nu(c^e+\rho)(R+\rho) + \rho(c^e+R+2\rho)T_2)\right] \geq 0 \end{aligned} \tag{3.50}$$

We know $\frac{Q(c^e-R)\rho}{(c^e+\rho)^2(R+\rho)^2 t_d} \leq 0$ as $(c^e - R) \leq 0$, thus we need only show

$$\begin{aligned} &\left[(c^e+\rho)(R+\rho)T_1^2 - (c^e+\rho)(R+\rho)(2\nu - T_2)T_2\right. \\ &\left. +2T_1(-\nu(c^e+\rho)(R+\rho) + \rho(c^e+R+2\rho)T_2)\right] \leq 0 \end{aligned}$$

Let $A = (c^e+\rho)(R+\rho)$, then show:

$$AT_1^2 + AT_2^2 + \rho(c^e+R+2\rho)2T_1T_2 \leq A2\nu(T_1+T_2)$$

$$A(T_1^2+T_2^2) + A2T_1T_2(\rho^2 - c^eR)2T_1T_2 \leq A2\nu(T_1+T_2)$$

$$A(T_1+T_2)^2 + (\rho^2 - c^eR)2T_1T_2 \leq A\nu(T_1+T_2) + A\nu(T_1+T_2)$$

We first show that $A(T_1+T_2)^2 \leq A\nu(T_1+T_2)$

$$\nu \geq (T_1+T_2) \text{ as } D_{bb} = \frac{\nu - (T_1+T_2)}{t_d} \geq 0 \text{ in Case 4}$$

We now need to show that:

$$\begin{aligned} (\rho^2 - c^eR)2T_1T_2 &\leq A\nu(T_1+T_2) \\ (\rho^2 - c^eR)2T_1T_2 &\leq (\rho^2 - c^{e2})2T_1T_2 \qquad \text{as } R \geq c \\ \text{Show:}(\rho^2 - c^{e2})2T_1T_2 &\leq A\nu(T_1+T_2) \end{aligned}$$

Since $(\rho^2 - c^{e2}) \leq A = c^eR + c^e\rho + R\rho + R^2$ and $A \geq 0$, this means we must show:

$$\begin{aligned} 2T_1T_2 &\leq \nu(T_1+T_2) \\ \frac{T_1}{\nu}\frac{T_2}{\nu} &= Z_1Z_2 \Rightarrow T_1T_2 = Z_1Z_2\nu^2, \text{ and} \\ \frac{T_1}{\nu} + \frac{T_2}{\nu} &= Z_1 + Z_2 \Rightarrow T_1 + T_2 = \nu(Z_1+Z_2) \end{aligned}$$

Show:

$$2\nu^2 Z_1 Z_2 \leq \nu^2(Z_1+Z_2)$$

Above holds if $Z_2 < 0$ (as $Z_1 + Z_2 \geq 0$ by our assumptions and $Z_1 Z_2 \leq 0$)

When $Z_2 \geq 0$ (so $Z_1, Z_2 \in [0,1]$ and $0 \leq Z_1 + Z_2 \leq 1$)

$$\begin{aligned} 2Z_1Z_2 &\leq (Z_1 + Z_2) \\ Z_1^2 + 2Z_1Z_2 + Z_2^2 - (Z_1^2 + Z_2^2) &\leq (Z_1 + Z_2) \\ (Z_1 + Z_2)^2 + (Z_1^2 + Z_2^2) &\leq (Z_1 + Z_2) \\ (Z_1 + Z_2)^2 + (Z_1^2 + Z_2^2) \leq (Z_1 + Z_2)^2 &\leq (Z_1 + Z_2), \text{as } Z_1, Z_2 \in [0,1] \text{ and } Z_1 + Z_2 \leq 1 \end{aligned}$$

Thus, consumer surplus is increasing under regulation.

Total Social Welfare under Regulation

In case PPP, we now provide the conditions for which total social welfare increases for a range under regulation.

Proof: Total social welfare at equilibrium is given by

$$\begin{aligned} &\frac{Q}{2(c^e+\rho)^2 t_d}\big(-4c^{3e}\nu + 2c^e\nu\rho(2(2\pi+\nu-\rho) - 2(\pi+\nu-\rho)Z_1 \\ &+ \nu Z_1^2 - 2(\pi+\nu-\rho)Z_2 + \nu Z_2^2) + 2\nu\rho^2(-1+Z_3)(-2\pi-\nu+\nu Z_3) \\ &+ c^{2e}(4\nu t_d - t_d^2 - 2\nu(-2\pi+\nu+4\rho-2\rho Z_3))\big) \end{aligned} \tag{3.51}$$

Total social welfare under regulation, where $R \geq c^e$ is given by

$$\begin{aligned} &\frac{Q}{2(R+\rho)^2 t_d}\big(-4R^3\nu + 2R\nu\rho(2(2\pi+\nu-\rho) - 2(\pi+\nu-\rho)Z_1 \\ &+ \nu Z_1^2 - 2(\pi+\nu-\rho)Z_2 + \nu Z_2^2) + 2\nu\rho^2(-1+Z_3)(-2\pi-\nu+\nu Z_3) \\ &+ R^2(4\nu t_d - t_d^2 - 2\nu(-2\pi+\nu+4\rho-2\rho Z_3))\big) \end{aligned} \tag{3.52}$$

We can examine the sign of the derivative of $TSW^R - TSW^e$ to determine whether total social welfare is increasing under regulation.

$$\begin{aligned} &\frac{Q}{(R+\rho)^3 t_d}\big(-2\nu((R+\rho)^3 + 2R\rho\nu) + \rho(4Rt_d - Rt_d^2 \\ &+ \nu(-\nu(R+\rho)(Z_1^2+Z_2^2) + 2(R+\rho)(\pi+\nu+\rho)(Z_1+Z_2) - 4\nu\rho Z_1 Z_2))\big) \end{aligned} \tag{3.53}$$

As we know $\frac{Q}{(R+\rho)^3 t_d} > 0$, we need only to show, at $R = c^e$, the following holds.

$$\begin{aligned} &-2\nu((c^e+\rho)^3 + 2c^e\rho\nu) + \rho(4c^e t_d - c^e t_d^2 \\ &+ \nu(-\nu(c^e+\rho)(Z_1^2+Z_2^2) + 2(c^e+\rho)(\pi+\nu+\rho)(Z_1+Z_2) - 4\nu\rho Z_1 Z_2)) \geq 0 \end{aligned} \tag{3.54}$$

Once again, we shall let $\rho C = (c^e + \rho)$. Thus we need to show

$$\begin{aligned} -2\nu((\rho C)^3 + 2\rho^2(C-1)\nu) + \rho(\rho(C-1)t_d(4-t_d) \\ +\nu(-\nu\rho C(Z_1^2+Z_2^2) + 2\rho C(\pi+\nu+\rho)(Z_1+Z_2) - 4\nu\rho Z_1 Z_2)) \geq 0 \end{aligned} \tag{3.55}$$

Concentrating on the left hand side of (3.55), and recognizing from case condition (i) that $4\nu^2 < t_d^2$, we find

$$\begin{aligned} &(C-1)(4\nu t_d - t_d^2) - 2\nu\rho C^3 - 4\nu(C-1) \\ &\qquad + \nu(2C(\pi+\rho+\nu)(Z_1+Z_2) - \nu C(Z_1^2+Z_2^2) - 4\nu Z_1 Z_2) \\ &> (C-1)(4\nu t_d - 4\nu^2) - 2\nu\rho C^3 - 4\nu(C-1) \\ &\qquad + \nu(2C(\pi+\rho+\nu)(Z_1+Z_2) - \nu C(Z_1^2+Z_2^2) - 4\nu Z_1 Z_2) \end{aligned}$$

We know $Z_1 + Z_2 \geq 0$, thus:

$$\begin{aligned} &(C-1)(4t_d - 4\nu) + 2C(\pi+\rho+\nu)(Z_1+Z_2) - \nu C(Z_1^2+Z_2^2) - 4\nu Z_1 Z_2 - 2\rho C^3 - 4(C-1) \\ \geq &(C-1)(4t_d - 4\nu) - \nu C(Z_1^2+Z_2^2) - 4\nu Z_1 Z_2 - 2\rho C^3 - 4(C-1) \end{aligned}$$

We know $Z_1^2 + Z_2^2 \leq 2$, thus:

$$\begin{aligned} &4(C-1)(t_d - \nu - 1) - \nu C(Z_1^2+Z_2^2) - 4\nu Z_1 Z_2 - \rho C^3 \\ &\qquad \geq 2(C-1)(t_d - \nu - 1) - \nu C - 2\nu Z_1 Z_2 - \rho C^3 \end{aligned}$$

Since $\frac{t_d}{2} > \nu$,

$$\begin{aligned} &2(C-1)(t_d - \nu - 1) - \nu C - 2\nu Z_1 Z_2 - \rho C^3 \\ &\quad \geq 2(C-1)(\nu - 1) - \nu C - 2\nu Z_1 Z_2 - \rho C^3 \end{aligned} \tag{3.56}$$

So, as the left hand side of (3.55) is greater than (3.56), we can show that

$$\begin{aligned} 2(C-1)(\nu-1) - \nu C - 2\nu Z_1 Z_2 - \rho C^3 &\geq 0 \\ C(\nu-2) + 2(1 - \nu(1+Z_1 Z_2)) - \rho Z_2^2 C - Z_1(1+Z_2)(\pi+\rho) &\geq 0 \\ C(\nu - 2 - \rho Z_2^2) + 2 - 2\nu(1+Z_1 Z_2) - Z_1 Z_2(\pi+\rho) - Z_1(\pi+\rho) &\geq 0 \end{aligned}$$

As $Z_2^2 \leq 1$,

$$\begin{aligned} &C(\nu - 2 - \rho Z_2^2) + 2 - 2\nu(1+Z_1 Z_2) - Z_1 Z_2(\pi+\rho) - Z_1(\pi+\rho) \\ \geq &C(\nu - 2 - \rho) + 2 - 2\nu(1+Z_1 Z_2) - Z_1 Z_2(\pi+\rho) - Z_1(\pi+\rho) \\ \geq &0 \end{aligned}$$

Solving for C, we find

$$C \geq \frac{2\nu(1+Z_1Z_2)+Z_1Z_2(\pi+\rho)+Z_1(\pi+\rho)-2}{\nu-2-\rho}$$

Finally, substituting $\rho C = c^e + \rho$ back in, we can obtain the condition in terms of c^e:

$$c^e \geq \rho\left(\frac{2\nu(1+Z_1Z_2)+Z_1Z_2(\pi+\rho)+Z_1(\pi+\rho)-\nu+\rho}{\nu-2-\rho}\right) \tag{3.57}$$

Since we know $c^e \geq 0$, we can examine where the right hand side of (3.57) is non-positive to show that there does indeed exist some region for which total social welfare is increasing under regulation, even when firms are substitutes in loss or unaffected (i.e. when $Z_2 \leq 0$). The denominator in (3.57) is less than zero as long as $\nu < 2+\rho$. First, we show that the numerator is always non-negative for $Z_2 > 0$ below:

$$2\nu(1+Z_1Z_2)+Z_1Z_2(\pi+\rho)+Z_1(\pi+\rho)-\nu+\rho \geq 0$$

Since $0 \leq Z_1Z_2 \leq 1$,

$$2\nu(1+Z_1Z_2)+Z_1Z_2(\pi+\rho)+Z_1(\pi+\rho)-\nu+\rho \geq 2\nu(1)+Z_1(\pi+\rho)-\nu+\rho \geq 0$$

$$\nu \geq -Z_1(\pi+\rho)-\rho \tag{3.58}$$

We know that (3.59) always holds as ν, Z_1, π, and ρ are all non-negative values. Thus, for $0 \leq \nu \leq \rho+2$ and $Z_2 > 0$, total social welfare is increasing under regulation for some range when firms are complements in loss ($Z_2 > 0$). Now we show the condition under which the numerator is also non-negative when $Z_2 \leq 0$.

$$2\nu(1+Z_1Z_2)+Z_1Z_2(\pi+\rho)+Z_1(\pi+\rho)-\nu+\rho \geq 0$$

Since now $-1 \leq Z_1Z_2 \leq 0$,

$$\begin{aligned} &2\nu(1+Z_1Z_2)+Z_1Z_2(\pi+\rho)+Z_1(\pi+\rho)-\nu+\rho \\ &\qquad \geq 2\nu(1-1)-Z_1(\pi+\rho)+Z_1(\pi+\rho)-\nu+\rho \geq 0 \\ &\rho \geq \nu \end{aligned} \tag{3.59}$$

When firms are substitutes in loss or unaffected, then, we require that $\nu \leq \rho$ in order for there to be a range for which total social welfare is increasing under regulation.

3.5.5 Eliminated Cases

In this section, we will show that the four cases FPF, PFF, PFP, and PPF are not possible in our Hotelling setting.

Case FPF

Here the market coverage is Full in both gg and bb states, but only Partial in the bg and gb states. The case conditions for this state are:

$$\text{(i)}\ \ \nu \geq \frac{t_d}{2}, \quad \text{(ii)}\ \ \nu < \frac{t_d + T_1 + T_2}{2}, \quad \text{(iii)}\ \ \nu \geq \frac{t_d}{2} + T_1 + T_2 \tag{3.60}$$

By considering pair-wise case conditions, we can eliminate this case from possibility:

First compare (3.60) *(i)* and *(ii)*

$$\frac{t_d}{2} \leq \nu < \frac{t_d + T_1 + T_2}{2} \qquad \text{yielding} \quad T_1 + T_2 > 0 \tag{3.61}$$

Now compare (3.60) *(ii)* and *(iii)*

$$\frac{t_d}{2} + T_1 + T_2 \leq \nu < \frac{t_d + T_1 + T_2}{2} \qquad \text{yielding} \quad T_1 + T_2 < 0 \tag{3.62}$$

Since we cannot simultaneously satisfy (3.61) and (3.62), this case is eliminated.

Case PFF

Here the market coverage is Partial in the gg state and Full in the bg, gb, and bb states. The case conditions for this state are:

$$\text{(i)}\ \ \nu < \frac{t_d}{2}, \quad \text{(ii)}\ \ \nu \geq \frac{t_d + T_1 + T_2}{2}, \quad \text{(iii)}\ \ \nu \geq \frac{t_d}{2} + T_1 + T_2 \tag{3.63}$$

It is not possible to meet both our requirement that $D_{gg} \geq D_{bb}$ (i.e. $Z_3 \geq 0$) and case condition (3.63) *(i)* simultaneously, thus this case is eliminated.

Case PFP

Here the market coverage is Partial in both gg and bb states, but Full in the bg and gb states. The case conditions for this state are:

$$\text{(i)}\ \ \nu < \frac{t_d}{2}, \quad \text{(ii)}\ \ \nu \geq \frac{t_d + T_1 + T_2}{2}, \quad \text{(iii)}\ \ \nu < \frac{t_d}{2} + T_1 + T_2 \tag{3.64}$$

By considering pair-wise case conditions, we can eliminate this case from possibility:

First compare (3.64) *(i)* and *(ii)*

$$\frac{t_d + T_1 + T_2}{2} \leq \nu < \frac{t_d}{2} \qquad \text{yielding} \quad T_1 + T_2 < 0 \tag{3.65}$$

Now compare (3.64) *(ii)* and *(iii)*

$$\frac{t_d + T_1 + T_2}{2} \leq \nu < \frac{t_d}{2} + T_1 + T_2 \qquad \text{yielding} \quad T_1 + T_2 > 0 \tag{3.66}$$

Since we cannot simultaneously satisfy (3.65) and (3.66), this case is eliminated.

Case PPF

Here the market coverage is Partial in *gg*, *bg*, and *gb* states and Full in the *bg* state. The case conditions for this state are:

$$\text{(i)} \;\; \nu < \frac{t_d}{2}, \quad \text{(ii)} \;\; \nu < \frac{t_d + T_1 + T_2}{2}, \quad \text{(iii)} \;\; \nu \geq \frac{t_d}{2} + T_1 + T_2 \tag{3.67}$$

It is not possible to meet both our assumption that $D_{gg} \geq D_{bb}$ (i.e. $Z_3 \geq 0$) and (3.67) *(i)* simultaneously, thus this case is eliminated.

Chapter 4

Risk Mitigation Decisions for IT Security

A version of this chapter has been submitted for review.

4.1. Introduction & Literature

Enterprises are under increasing pressure to better manage operational risks, including information risks. As an example, in 2004, an accounts payable clerk used her computer to access her firm's accounting system and issued 127 checks payable to herself and others. Checks written were cashed or deposited into her account or the accounts of her accomplices. The clerk was able to alter the electronic check registers to make it appear as if the checks had been made payable to the firm's legitimate vendors. The firm lost at least $875,035. The clerk was caught, pled guilty to two counts of computer fraud and faced a maximum sentence of five years in prison and a $250,000 fine (DoJ 2004).

Headlines in influential media outlets routinely recall the latest information security breach affecting yet another organization. Unfortunately, the costs of such breaches add up to real money. According to the Identity Theft Resource Center, there were 16,167,542 records reported as breached in 2010 (ITRC 2010). If the estimates provided by the 2010 Annual Study: U.S. Cost of a Data Breach of $214 per record are close to accurate, the total cost in 2010 of data breaches is approximately $3.5 Billion in the United States alone (Ponemon Institute and Symantec 2011). This estimate only accounts for breaches of confidential information, such as credit card numbers, social security numbers, drivers' license data, bank account numbers, etc. as this information is required to be reported to the state attorney general in most US jurisdictions. Firms are also growing increasingly aware of the value of informational assets and how attractive these assets could be to the wrong parties; assets such as patent applications, engineering designs, chemical formulations, corporate strategy documents, research and development documentation, among other potentially high-value information.

Numerous frameworks for managing risks to information and technology resources abound, ranging from the ISO series on risk management (ISO 31000, ISO 31010) and information security management (ISO 27000 series) to The Committee of Sponsoring Organizations of the Treadway Commission (COSO) to COBiT to the NIST standards for risk management and information security, and others. These standards and frameworks share many similarities in that information risks must be identified, assessed, and managed. Such risks are managed by making decisions

on which risks to accept, which to transfer via sourcing agreements, insurance or both, and which to mitigate or reduce to a more acceptable level. Risks are typically mitigated by placing one or more controls at a specific step in a business process. A control might be a specific technology, for example an access control mechanism, or it might be a procedure, such as having a supervisor signature on an override. Controls also have varying degrees of reliability in terms of preventing or detecting erroneous or fraudulent data moving through a system. While each framework has strengths and weaknesses, each one defaults to a generic prescriptive approach, which can be more or less implemented as a type of systems checklist. Despite being generic and in theory, customizable to each organization's unique set of systems and processes, the checklist approach becomes extremely difficult for managers to use with today's complex arrays of processes and technologies.

The checklist approach falls short in at least two areas. First, workflows change over time, as do the threats. Appropriate controls may not be used for many reasons, such as the system complexity might be greater than anticipated by the creators of the checklist or the introduction of new technologies might limit the effectiveness controls. An example would be the introduction of a wireless access point in a warehouse management system by an employee outside of the IT organization. Second, managers might overspend or misallocate funds for controls because they are unable to assess the impact of the interaction between the controls available, potential attacks, and business processes. For example, an expensive control might be placed on a check printer which limits who can pick up a printed check. The printer might be located in a highly secured area which requires remodeling with expensive materials and a trained guard checks identification of those few employees allowed to print and pick up checks. However, if no background check is performed (a relatively inexpensive control) on the few employees allowed to print and pick up checks, then additional risk is introduced into the system despite the checklist.

The concept of the organizations' workflows can be used to define the focus of security controls (Rodríguez et al. 2011). Indeed, this was the motivation behind Section 404 of the 2002 Sarbanes-Oxley Act (SOX) in the US, which requires explicit management of internal controls over financial reporting processes. By focusing on the ways in which people, data, documents, forms, processes, etc. interact to accomplish organizational goals, we can then make better decisions about which

controls need to be placed in which workflow locations, in order to better manage the organization's overall information risk profile.

Not unlike physical sensor systems (for example, waterflow contamination detection systems (Watson et al. 2009)), multiple problems arise in selecting, placing, and managing internal controls for information risk management within organizational workflows. The problems of selecting and placing internal controls have long been addressed by heuristics, meaning that internal audit practitioners have developed checklists and guidelines for the selection and placement of such controls. The same is generally true for information security management. Good security managers follow prescriptive practices in selecting technology and policy controls typically generated by outside agencies and augmented by internal institutional experiences. While the checklist approach generally meets legislative requirements, this approach is likely sub-optimal from an enterprise information risk management perspective. Do the controls selected in the locations in which they are placed within the organizational workflow provide an optimal level of risk management? There is significant need to create an integrated, contextually holistic view of information risk management given the workflow processes of the organization.

The orientation of this paper is to develop decision models for managers to place controls, and then simulate the expected effectiveness of these controls against risk exposures. The goal of this research is to enable decision makers to integrate the analysis of controls into the workflow context. We formalize a representation of the investment and control placement problem within the overlapping and interconnected workflows of the organization, as well as propose insights and solutions to the problem. This work falls under the category of design science modeling; we model the organizational workflows and place controls to mitigate information risks. We test three solution methods to place controls, two of which are heuristics based on checklist-style decision methods. The third method uses an integer linear programming (IP) technique. We solve this problem with a budget constraint and then test the solutions with a period of simulated incident attacks. Depending on the controls selected, damages may or may not be mitigated. The incremental risk exposure of the three decision methods, compared to the lowest cost control expenditures, are used to evaluate relative effectiveness. This work is important because currently there is no method to effectively integrate information management risks within the

context of the organizational workflow.

The rest of the paper is organized as follows. The literature review is in section 4.2. We present our problem statement and formulation model in section 4.3. Section 4.3 also describes two heuristic procedures for adopting controls. Section 4.4 presents the computational experiments and results, and section 4.5 ends the paper with discussion and conclusions.

4.2. Literature Review

Earlier, we identified two central themes in risk management investment decisions; the need for both controls and an integrated view of risk in the context of workflows. In this section we review studies related to these themes starting with works related to making investments to manage information risk. This allows us to then present our model of control investment and placement within workflows in order to manage risk.

Gordon and Loeb (2002) proposed one of the earliest models for making economically rational information security investments. Their model takes into account the vulnerability of the information to be protected and the resources available to protect that information. They found that in certain scenarios, firms should only spend a fraction of their expected losses to prevent security breaches, which is contrary to the popular belief at this time, which is that information security investments should be continually increasing. Bodin et al. (2005) incorporate the Analytic Hierarchy Process (AHP) into the earlier Gordon and Loeb (2002) model, in order to take advantage of qualitative information in making security investments.

Other researchers have since presented more complex and detailed models. Kumar et al. (2008a) use a portfolio model of information security countermeasures to simulate the value of various portfolios against various attacks. They were able to demonstrate through simulation experiments that the interaction effects of the various security countermeasures can offer more protection to the organization than just the sum of each countermeasure's benefit, which indicates that an overall strong information security infrastructure can mitigate a weaker component of the infrastructure. Kumar et al. (2007) present an analytical model of investment decisions and countermeasures for protecting against availability and confidentiality-type attacks. Their model results in guidance to managers regarding investments and al-

locations to divisions for both availability-protecting and confidentiality protecting mechanisms. Herath and Herath (2008) propose a real options analysis (ROA) model for evaluating information security investments and present a Bayesian learning and post-audit function, in order to incorporate continuous information into their model. Cavusoglu et al. (2008) compare game-theoretic models to decision-theoretic models of security investment and report that game-theoretic approaches can result in better outcomes to the firm under certain conditions, which emphasizes the need to consider information security management a dynamic and strategic problem. While these papers take different approaches to modeling investments in information security, they all consider the interaction effects among security technologies, which is an important development in the literature.

The papers mentioned above generally tend to focus on security technologies, such as intrusion detection systems and anti-virus protection. The more general concepts of internal controls, which include access control technologies and internal audit processes, as well as technologies used to protect the confidentiality and integrity of data, have also been studied in the information security context. Researchers have examined the specific nature of controls used in protecting information systems. For example, Weber (1989) examined electronic funds transfer systems, and found a need to balance speed and ease of use with security. Wood (1990) prescribed twenty-three principles for designing controls in software, ranging from cost effectiveness to maintaining a low profile for the control.

Basu and Blanning (1997) define a workflow as "the flow of information and work through one or more organizational entities involved in business processes," (Basu and Blanning 1997, pp. 359-360). Workflows are critical to organizations, as they depict the business processes and rules within the organization, and are necessary for systems analysis and design activities, as well as for efficiency and control purposes. Basu and Blanning (1997, 2000, 2003) propose using metagraphs as a formalization of organizational workflows, which allows for formal analysis of workflows and business processes.

Cernauskas and Tarantino (2009) suggested that combining business process management and process control can improve risk transparency and reduce operational losses. Kumar et al. (2008b) examine different policies for countermeasure placement given information asymmetry between the CIO and division managers. In

the context of auditing, Krishnan et al. (2005) provide a formal method of assessing data reliability that helps auditors choose the controls to review, balancing the cost and accuracy of assessment requirements. Their set covering model could be adapted to help answer the question about which controls to implement in order to reach a desired level of data reliability. Extensions to the work of Krishnan et al. (2005) provide a framework for managing data quality risks in accounting information systems by modeling error propagation through the system, where the system is represented at the business process level (Bai et al. 2007, 2011b). A Markov decision model then allows for the determination of the optimal control policy, for specific control procedures. This model assumes that at each error source, there is one control procedure that will be implemented, as opposed to selecting from a portfolio of controls, which could be a combination of technologies and procedures, each with varying costs and benefits.

To our knowledge there is scant literature which provides managers with support to establish the strategic placement of controls within workflows. Bai et al. (2011a) examine access control for information privacy and confidentiality within a workflow context. While the problem of access control is a critical and complex issue, our work examines the more general issue of control placement from an overall investment and risk management perspective.

Having identified the gap in the literature regarding the need for a model for strategic placement of controls, we formalize our problem statement in the next section. Following this, we present our model for optimal control placement within a workflow framework with the goal of mitigating information risk subject to budgetary constraints.

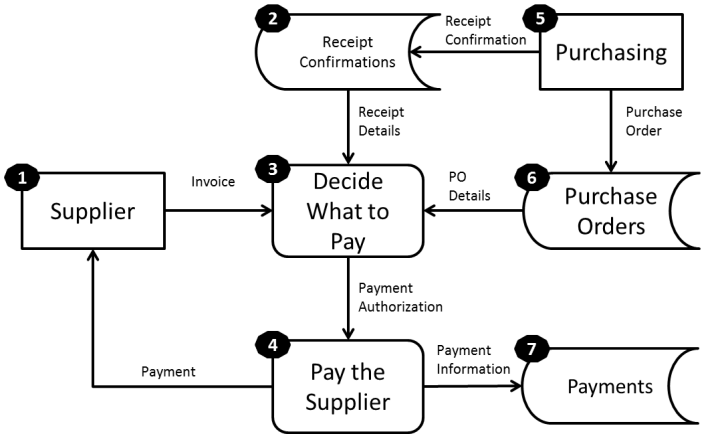
4.3. Problem Statement & Model Development

The placement of controls is a matter of deciding on how to best guard against potential information security breaches given constraints. For a single workflow, there are multiple security scenarios that must be considered, each with multiple protection choices and implementation locations in order to address confidentiality, integrity, and availability concerns. Workflow controls have costs associated with their acquisition, implementation, and management. The multitude of choices in multiple scenarios becomes a combinatorial problem; multiple workflows amplify

this combinatorial problem. Since most organizations will have clear constraints as to the budgets that can be spent on information security, we conclude that the resulting management problem is a combinatorial optimization problem with budget constraints.

Consider a standard purchasing workflow, as illustrated by Figure 4.1. We assume that a security breach may occur at any node in the workflow, although we recognize that some nodes might be more vulnerable than others. Our goal is to place high quality (efficient) controls that minimize the potential damage to data contained within the workflow.

Figure 4.1: Process Flow Example.

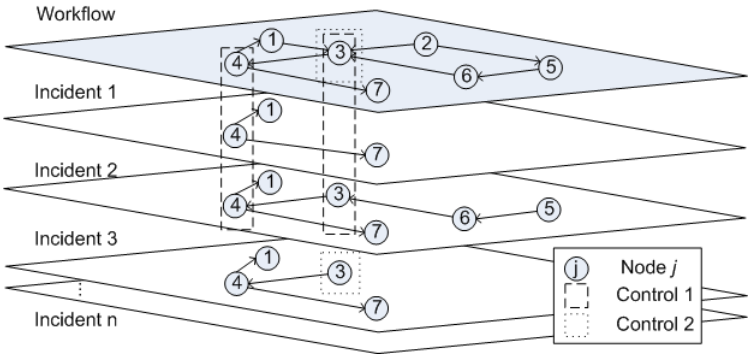


In Figure 4.2 we present the workflow illustrated in Figure 4.1 simply as the set of nodes and edges to allow us to easily illustrate some hypothetical incidents such as breach of confidentiality (incident 1), loss of data integrity (incident 2), and impaired availability of data (incident 3). Incident 1 could be an intercepted electronic funds transfer (EFT). Incident 2 could be the deliberate altering of PO information. Incident 3 could be the loss of access to a database server due to power outage. To detect these incidents, many controls can be placed in many different locations, and the same control can even be placed in multiple locations. In general, the damage of an incident is lower the earlier we detect it. This decrease in damage, though, must be balanced by the cost of placing controls in multiple locations.

Some controls at particular locations will detect some but perhaps not all incidents. Control 1, placed at node 3, can detect incident 2 whereas control 2, placed at node 3, can detect incident 3. Control 1, placed at node 4, can detect both incidents

1 and 2. However, if we rely on control 1 placed at location 4 to detect incident 2, we will incur increased damages related to that incident compared to placing control 1 at location 3. Thus, it might be worthwhile placing control 1 at nodes 3 and 4, even though there is redundant coverage for incident 2. Thus, in Figure 4.2 we see that Control 1 has been placed at nodes 3 and 4, and Control 2 has been placed at node 3 and all three incidents may be detected.

Figure 4.2: Incident Examples.



4.3.1 Model

Risk reduction will occur through the strategic placement of controls within the workflow, given the costs and benefits of the controls under consideration, as well as the impact of the controls given the specific activities at each location within the workflow structure. We assume budget constraints, which limit the availability and effectiveness of the controls. Our approach is similar to the approach taken in the placement of sensors to detect contamination in water networks (Leskovec et al. 2007, Murray et al. 2009, Watson et al. 2009). Leskovec et al. (2007) also apply the approach to model the spread of information in blogs, identifying key blogs that quickly cover the majority of "information cascades." We can also apply this approach to contamination of information in organizational workflows.

Graph Theoretic Definition:

Given a graph with a set of location nodes (J) and edges (E), we define a set of incident scenarios (I) each describing an incident such as a security breach or the spread of unwanted data, and a set of controls (K) to mitigate incidents. An incident, $i \in I$, is initiated from a single node in the workflow, and spreads through the workflow in a pattern described by the incident tree formed by the incident i's

connections to other affected nodes; that is, a set of arcs $\{\alpha, \beta\} \in E$. The collection of arcs $\{\alpha, \beta\}$ for an incident i depict a damage dissemination flow created by that incident. Incidents can be detected and controlled for by installing a control $k \in K$ at any affected location. For each location, $j \in J$, there are zero or more control options, where each control will apply to one or more of the incidents $i \in I$. The use of a control at a location would incur a cost that may be location dependent. For any location, we may elect to use zero or more controls to guard against each incident. Each control type used for an incident at a location implies a unique level of potential damage resulting from the incident. The flow associated with each incident is used as a proxy for the damage from an incident given its control location and type. Once an incident is detected at a location, we assume that all issues related to that incident are resolved. If, for some reason, this is not the case, then a separate incident must be constructed.

Data:

Incident and damage data are described by several variables. First, we must identify which incidents, $i \in I$ can effectively be controlled by a control $k \in K$ if it is placed at location $j \in J$. We store this information in the variable a_{ijk}, defined as follows:

$$a_{ijk} = \begin{cases} 1 & \text{if incident } i \text{ is covered by a control of type } k \text{ at location } j, \\ 0 & \text{otherwise} \end{cases}$$

Each control has a cost, c_{jk} associated with it and the cost may vary based on its location in the workflow. Each incident, if not detected, will cause the firm to incur damages, D_i. However, by placing a control k at location j, this damage may be reduced by an amount d_{ijk}. As we assume that workflows may be described by spanning trees and that the first control along a path that can detect an incident will detect it, we must also keep track of the set of paths, P_i, for each incident. A budget, B, limits the total monetary resources available to purchase (and presumably implement and manage) the controls. The goal of this paper is to select locations and types of controls in such a manner that the total expected damage is minimized, while complying with the budget constraint. Table 4.1 summarizes the notation used in our IP.

Decision Variables:

Table 4.1: Notation for IP formulation.

Term	Name	Description
I	Incidents	Set of incidents where $i \in I$
J	Nodes	Set of nodes within a workflow where $j \in J$
K	Controls	Set of controls where $k \in K$
B	Budget	Limit on amount to spend for controls
a_{ijk}	Applicability	Denotes which incidents i are controllable by control k at location j
c_{jk}	Control cost	Cost for deploying control k at location j
D_i	Uncontrolled damage	Damage of incident i with no controls
d_{ijk}	Damage reduction	Reduction in damage for incident i for deploying control k at location j
P_i	Paths	Set of paths defining incident i
s_{jk}	Selected controls	Decision variable
x_{ijk}	Incident controls	Decision variable

In our model for multiple coverage, there are two decision variables. It is possible to purchase multiple controls at each and every location in our workflows. Thus, we define:

$$s_{jk} = \begin{cases} 1 & \text{if control of type } k \text{ is implemented at location } j, \\ 0 & \text{otherwise} \end{cases}$$

We must also decide which of the controls purchased at each location will be used for detecting a given incident. Thus, we define:

$$x_{ijk} = \begin{cases} 1 & \text{if incident } i \text{ is covered by a control of type } k \text{ at location } j, \\ 0 & \text{otherwise} \end{cases}$$

Note that, in data generation, for every incident i we require $\sum_{j \in J, k \in K} d_{ijk} \leq D_i$ so that damage is always non-negative, even when controls are used. We accomplish this by using the distance between nodes to calculate both D_i and the discounts, d_{ijk}. A full description of how the data variables are generated is provided in appendix 4.6.1.

Problem: Flow Risk Reduction (FRR)

Our solution is a two stage process where we first select controls to minimize damage. It is possible that different solutions, at different costs, can result in the same minimal damage. Thus, we perform a second stage where the value of the objective function from stage 1 becomes a constraint in stage 2 where we find the minimal cost solution.

In this formulation for stage 1, we want to minimize the damage for not placing controls or placing ineffective controls:

$$\min[\sum_{i\in I} D_i - \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} d_{ijk}x_{ijk}] \tag{4.1}$$

Subject to:

Can only purchase a control of type k at location j if that control at that location is used to detect at least one incident, i.

$$s_{jk} \leq \sum_{i\in I} x_{ijk} \qquad \forall j \in J, k \in K \tag{4.2}$$

Breach observed only if a control exists:

$$x_{ijk} \leq a_{ijk}s_{jk} \qquad \forall i \in I, j \in J, k \in K \tag{4.3}$$

For each incident i, at most one control is active on each path p in the set of paths P_i from the root node to each terminal node in the incident:

$$\sum_{j\in p}\sum_{k\in K} x_{ijk} \leq 1 \qquad \forall i \in I, p \in P_i \tag{4.4}$$

Total spending on controls must not exceed the budget amount.

$$\sum_{j\in J}\sum_{k\in K} c_{jk}s_{jk} \leq B \tag{4.5}$$

$$s_{jk} \in \{0,1\}, \quad x_{ijk} \in \{0,1\} \tag{4.6}$$

In this formulation for stage 2, we wish to minimize the total cost of controls with the constraint that the damages from this solution must not exceed the value of the objective function found in stage 1. Thus our new objective function becomes:

$$\min \sum_{j\in J}\sum_{k\in K} c_{jk}s_{jk} \tag{4.7}$$

Subject to:

Constraints (4.2) to (4.6) from stage 1 plus:

Total damage must not exceed the value of the objective function found in stage 1, Γ.

$$\sum_{i\in I} D_i - \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} d_{ijk}x_{ijk} \leq \theta\Gamma \tag{4.8}$$

The parameter θ in (4.8) is a slight relaxation of the damage restriction, Γ, necessary to accommodate for rounding errors in the solver. In our testing, we used parameter values of 1.000004 and 1.000009 as necessary.

The knapsack problem is NP-Complete (Karp 1972, 2010). The FRR problem is a special case of the knapsack problem, where categories of items are available but you can pick at most one item from each category. These categories are the incidents in the FRR problem. The FRR problem is NP-Complete.

4.3.2 Heuristic Decision Making

Current management practice is to use checklists or heuristic rules of thumb to guide the placement of controls in workflows. We compare our formulation with two heuristic decision making models selected to mimic the most likely processes used by human decision makers. The first heuristic selects controls for locations that will result in the maximum reduction in damage across all incidents in an iterative manner. If the first choice of a control and location exceeds the budget, the heuristic will search through the remaining choices to see if there is a control at a location that can be afforded within the budget. It continues in this manner until the budget is reached or there are no more affordable controls.

1. Calculate the discount across all incidents for deploying a control of type k at location j. That is, calculate $\sum_i a_{ijk} d_{ijk} \forall j \in J, k \in K$

2. Select the control which results in the largest discount across all incidents. That is, set $s_{j^*k^*} = 1$ for the control k^* at location j^* which results in the largest value for $\sum_i a_{ij^*k^*} d_{ij^*k^*}$, if it fits within our budget.

   (a) If the first choice of control does not fit in the budget, look for the next best control that is affordable.

3. Continue steps 1 and 2 until no more controls can be purchased within the budget.

The second heuristic determines the incident that has the largest expected damage and selects the control at a location that will minimize the damage for that incident. It then recalculates the expected damage for all incidents given the control and location selected, determines the incident with the highest remaining damages,

selects the control at a location that minimizes the damage for that incident and repeats these steps until the entire budget is exhausted. Like the first heuristic, if the first choice of a control and location exceeds the budget, the heuristic will search through the remaining choices to see if there is a control at a location that can be afforded within the budget, stopping only once it cannot purchase any more controls within the budget.

1. Select the incident, i', which has the highest expected damage if no controls are selected.

2. Select the control k' at location j' which will result in the largest reduction in damage for incident i'. That is, find the largest value of $d_{i'jk}$ for this incident and set $s_{j'k'} = 1$, if it fits within our budget.

   (a) If the first choice of control does not fit in the budget, look for the next best control.

   (b) If no control for this incident fits within the budget, look at the incident with the next highest expected damage and repeat step 2.

3. Recalculate the expected damages for all incidents given the control k' at location j' has been deployed. Thus, for any incident for which the control k' at location j' is effective, the expected damage should be adjusted by the discount, $d_{i'jk}$.

4. Repeat steps 1 through 3 until no more controls can be purchased within the budget.

Next, we test the effectiveness of each of these three control placement methods against a barrage of simulated attacks. The relative effectiveness of each method is measured as the reduction of risk achieved. The barrage of attacks will simulate a year of attacks with different distributions of realization.

4.4. Computational Experiments and Results

We created a collection of 162 unique data sets with randomly generated node locations, incidents and controls along with all other input data for the model defined in section 3. This data set embodies the manager's expectations regarding attacks

Table 4.2: Listing of variables for IP formulation.

Variable	**Values**
Incidents (I)	50, 100, 150
Locations (J)	50, 275, 500
Controls (K)	10, 15, 20
Budget Scale (BP)	0.05, 0.1
Maximum cost of control ($maxC$)	900, 950, 1000

and control effectiveness as related to organizational workflows. The data generation is explained in detail in Appendix 4.6.1. A summary of the data generation parameters is presented in Table 4.2 below.

The IP control selection method and the two heuristics defined earlier are used to select a set of controls for each data set. We refer to the solutions obtained by each method IP, H1 and H2, respectively. Algorithms H1 and H2 will spend the entire budgeted amount, BP. To create an apples-to-apples comparison between the IP method and the heuristics, we restrict the budget for the heuristic algorithms to match the cost obtained by the IP method, calling the two restricted budget methods H1a and H2a, respectively. In every case, the cost of the IP method was below the budget. For one data set the IP stage 1 solution was found but the IP stage 2 solutions was too difficult for CPLEX to solve in a timely manner and was terminated after ten days without finding an IP solution. This data set was replaced.

All experiments were conducted using MatLab and AMPL to call CPLEX 11.0.1 running on an IBM X-series 3550 with eight Intel Xenon processors running at 3.1GHz and 32GB of RAM. The operating system is Windows Server 2003 R2 Enterprise x64 Edition with Service Pack 2 applied.

Finding controls is a fairly efficient process. Each stage of this process is timed using MatLab's TIC and TOC stopwatch functions, which records elapsed wall clock time between the start and finish of the code segments described in Table 4.3. The IP method takes the longest time to solve, taking on average 6,038.912 seconds (1.68 hours). The worst case took 372,876.8 seconds (4.32 days) to find the solution, with stage 2 consuming most of the time. Stage 1 finds a solution with the lowest expected damage in an average time of 21.63 seconds, and stage 2 finds the cheapest solution among all solutions with the lowest expected damage. In general, the number of node locations in the workflow has the most dramatic effect on the length of time

Table 4.3: CPU time to find controls, in seconds

	Average	**Minimum**	**Maximum**
IP stage 1	21.630	0.385	205.592
IP stage 2	6017.282	1.212	372837.500
IP total	6038.912	1.608	372876.800
H1 algorithm	0.013	0.001	0.093
H1 formulation	4.985	0.177	20.480
H1 total	4.998	0.178	20.521
H1a algorithm	0.012	0.001	0.042
H1a formulation	4.634	0.174	18.716
H1a total	4.646	0.175	18.758
H2 algorithm	0.108	0.007	0.493
H2 formulation	4.855	0.182	20.043
H2 total	4.963	0.189	20.536
H2a algorithm	0.021	0.001	0.122
H2a formulation	4.610	0.174	18.832
H2a total	4.631	0.177	18.897

it takes both stage 1 and stage 2 of the IP method to solve. In comparison, the heuristic methods take at most 20.536 seconds in the worst case and under 5 seconds on average. Note that the solution from each heuristic was input to a modified version of the formulation to efficiently calculate the damages; the s_{jk} values were input as data and constraint (4.2) is not applicable. The time to complete this step for each heuristic is recorded in the respective line labeled "formulation" in Table 4.3.

Each data set effectively serves as a training set to create controls using the five alternative methods of choosing controls which will protect against future attacks. To determine if the control placements are effective, the system is placed under simulated attack. That is, to test the solutions given by the IP method and the heuristics under different attack scenarios, we generate a weekly set of attacks for an entire year (52 sets of attacks). Descriptions of the attack scenarios are provided in Appendix 4.6.2.

The performance of each solution against attack simulation 1 which follows a uniform random distribution is presented in Table 4.4. To analyze the performance of each solution against the simulated attack scenarios, the reduction of risk for each solution is considered. The risk reduction measure is calculated as the difference between the damage of the suite of attacks when no controls have been placed versus

Table 4.4: RR and RROI for Attacks following Uniform Distribution

	RR					**RROI**				
	IP	H1	H1a	H2	H2a	IP	H1	H1a	H2	H2a
Average	100	97.7	81.8	88.6	62.1	100	4.3	81.8	3.8	62.1
Minimum	100	74.2	33.0	39.7	6.9	100	0.2	33.0	0.2	6.9
Maximum	100	100.5	100.0	100.0	97.5	100	54.7	100.0	54.2	97.5

Table 4.5: RR and RROI for Attacks following Expected Distribution

	RR					**RROI**				
	IP	H1	H1a	H2	H2a	IP	H1	H1a	H2	H2a
Average	100	97.6	80.5	88.2	61.1	100	4.3	80.5	3.8	61.1
Minimum	100	80.3	63.7	41.4	7.8	100	0.2	63.7	0.2	7.8
Maximum	100	100.3	100.0	100.1	89.8	100	53.6	100.0	53.0	89.8

the damage that is incurred under the solution of interest, where:

$$RR = damage\ without\ controls\ -\ damage\ given\ selected\ controls$$

In Table 4.4, the average, minimum, and maximum risk reduction (RR) of each method as a percentage of the risk reduction found with the IP method is presented. The risk reduction on investment (RROI) is calculated by dividing the risk reduction by the cost of the solution solution method, where:

$$RROI = RR/cost\ of\ solution$$

In Table 4.4, the average, minimum, and maximum RROI of each method as a percentage of the risk reduction on investment found with the IP method is presented. Thus, the values for the IP method are set to 100 in every instance since this is the baseline. A value below 100 indicates a worse solution, and a value over 100 indicates a better result than the IP method baseline.

The same results for Attack 2 which is also random, but follows the expected distribution, are presented in Table 4.5. Results indicate that the IP method for finding controls is superior for both attack scenario 1 and 2 in terms of both risk reduction (RR) and risk reduction on investment (RROI).

Some general observations are made. Heuristic H1 may occasionally find the best solution under attack, and often finds solutions as good as IP, but at significantly higher cost. Heuristic H2 does not find the best solution, even at additional cost, but occasionally finds a better solution than the IP method. When constrained to the

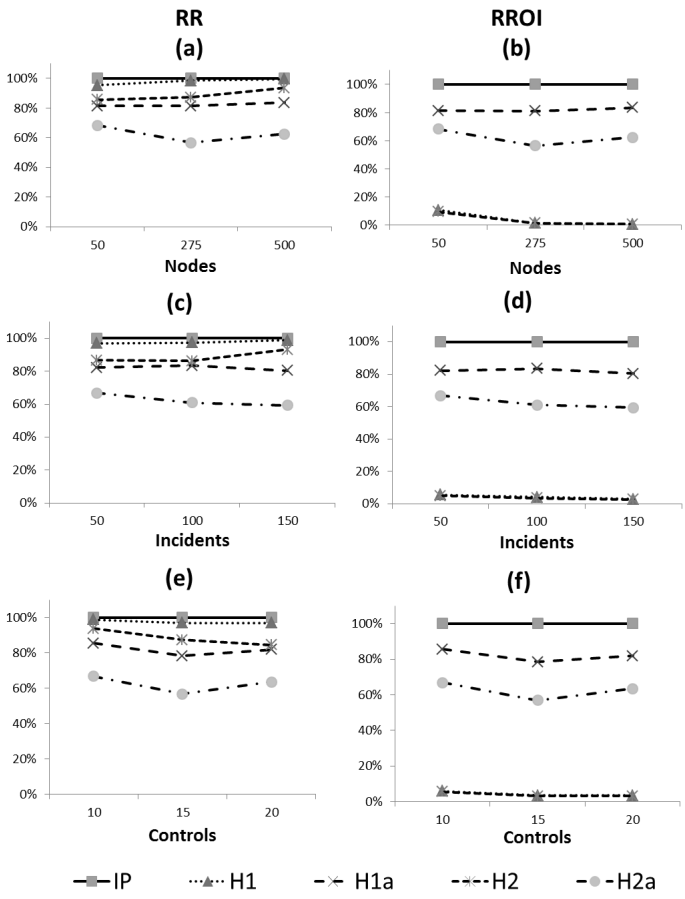
same cost as the IP stage 2 solution, the performance of both heuristic methods, H1a and H2a, deteriorate drastically in terms of both RR and RROI. Thus, we conclude that if the IP method is unsolvable due to excessive problem size, then the heuristics can be utilized to solve the problem. However, significantly poorer risk reductions are likely and the cost will be higher, driving the RROI down substantially compared to using the IP method.

The reason for the poor performance of the heuristic solutions H1 and H2 in terms of RROI is that they drastically overspend on controls while yielding slightly worse control configurations on average. The heuristics cannot make valid judgments with respect to when the marginal benefit of adding the next control outweighs its cost; the stopping rule for the heuristics is to continue until all available money is spent. As a result, while it is sometimes possible to increase the risk reduction (e.g., H1 occasionally finds a better solution than the IP in terms of RR), the cost of doing so makes the overall return on the investment much smaller than the IP's solution. Having said that, in some situations the overriding factor is risk reduction rather than RROI. An example is when risk represents loss of life, and measuring the RROI of reducing the loss of life is considered to be unethical.

To compare how performance changes with the various methods as the parameter settings change, we illustrate the average performance under the simulated attacks for the number of node locations in Figure 4.3 (a) and (b), the number of incidents in Figure 4.3 (b) and (c), and the number of controls in Figure 4.3 (c) and (d). We see that the IP method is superior on average for all three parameters over all values tested. Figure 4.4 presents the equivalent results under the attacks simulated using the second method (Attack 2) presented in Appendix 4.6.2.

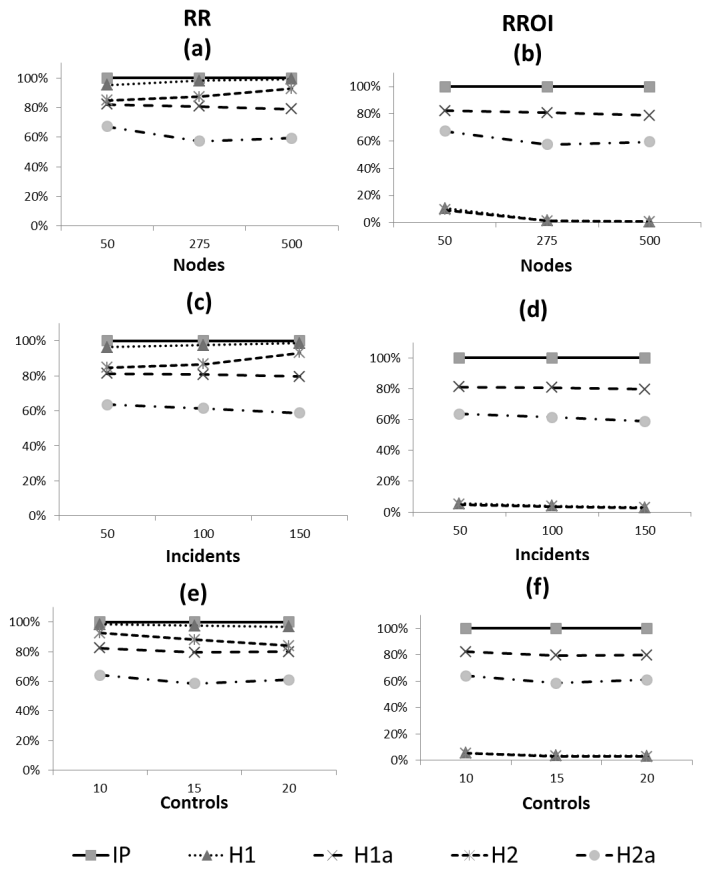
Why doesn't the IP method produce the best set of controls to protect against attack in every situation, without exception? The IP method does produce the optimal solution for the expected incidents and attack frequency, which is why it performs so well across the board (note that all methods use the same expected values to produce control configurations). However, when the realized attacks are made, the actual attack occurrences in the experiment deviate randomly from the expected arrival rates. When reality differs from expectations, which is almost always the case, then the sub-optimal heuristics can sometimes produce better realized solutions than the "optimal" IP method. This is borne out in our experimental results.

Figure 4.3: Attack 1 results relative to the IP solution by Nodes (a, b), Incidents (c, d) and Controls (e, f).



In summary, we can say that the IP method is the overwhelmingly best overall approach regardless of budget if the problem is small enough to be solved with this procedure. Having said that, we found no instances where the IP method failed to solve the problem, and we tested very large problem sizes of up to 500 control locations, 150 incidents, and 20 controls. What if solutions using the IP method cannot be found because the number of node locations becomes too large? Then a solution heuristic would have to be used, and heuristic H1 is superior to H2. Reducing the budget of the heuristics to compensate for overspending, as is done in H1a and H2a, results in dramatically poorer solutions. Thus, if the heuristic procedures are utilized, then artificially lowering the budget to prevent overspending will result in substantially more risk being carried by the firm. Thus, based on the experimental results, we can conclude that the use of the heuristics is never appropriate in a

Figure 4.4: Attack 2 results relative to the IP solution by Nodes (a, b), Incidents (c, d) and Controls (e, f).



low budget environment. Additionally, even in a high budget environment, the IP method results in less risk an average being carried by the firm, and the heuristics are much worse in the worst case scenarios.

4.5. Discussion and Conclusions

Insights from this paper are twofold: first, we specify the FRR control placement decision problem using formal methods. This leads to a better understanding of the problem, and shows important connections between security investment decisions and information risk management outcomes. Second, we demonstrate how this problem can be solved using integer programming methods as well as heuristics. We demonstrate how trade-offs can be made with respect to security investments within the context of organizational workflows.

While the model currently assumes perfect detection of an incident by a control,

it does not assume that the cost of all controls is the same. Future work could allow controls to detect incidents with probabilistic reliability and also could allow for differentiated damage prevention, meaning some controls are more effective than others against a given incident. However, our current model accommodates both the "worst case" and "expected value" views of the world.

The decision model allows for finding cost-efficient ways of protection against information security scenarios in the form of prevention or detection controls, or both - we do not make this distinction in the paper. The controls may have future impact on likelihoods of incidents, but this would have to be considered when preparing to solve the model again at some future time. Also, the model assumes that all control decisions are implemented in the current period. However, many control systems actually evolve over time and the decision to implement controls are made in the context of the control infrastructure that already exists and future controls that will take some time to implement, given the high time and cost to reallocate control resources from one task or location to another. For example, employees may need to be relocated to different cities, or new employees may need to be hired gradually over time. Future work can look at the multi-period dimensions of the problem, helping to identify not only optimal controls, but also optimal control implementation ordering.

There are multiple areas worth further exploration. More complex instantiations of controls, incidents, and workflows can all be considered, as mentioned previously. The model could be extended to consider the additional risk reduction afforded by the purchase of so-called "cyberinsurance." The model could be further tested using actual incident data. Finally, given the extent of business process outsourcing, the model could be extended to examine cross-organizational workflows, building on work by Patterson et al. (2006).

The ultimate goal of this line of research is to build improved decision support tools for managers faced with managing the information risk in their enterprises. While the current practice of using intuition, experience, and best practices is an important starting point, managers who incorporate more formal methods, such as the proposed FRR model, can further improve resource allocation decisions and information risk management outcomes.

Bibliography

Bai, X., R. Gopal, M. Nunez, D. Zhdanov. 2011a. On the prevention of fraud and privacy exposure in process information flow. *INFORMS Journal on Computing* forthcoming.

Bai, X., M. Nunez, J. R. Kalagnanam. 2011b. Managing data quality risk in accounting information systems. *Information Systems Research* forthcoming.

Bai, X., R. Padman, R. Krishnan. 2007. A risk management approach to business process design. *Proceedings from the International Conference on Information Systems.*

Basu, A., R. Blanning. 1997. Metagraph transformations and workflow analysis. *Proceedings of the Thirtieth Hawaii International Conference on System Sciences*, vol. 4. 359–366.

Basu, A., R.W. Blanning. 2000. A formal approach to workflow analysis. *Information Systems Research* **11**(1) 17–36.

Basu, A., R.W. Blanning. 2003. Synthesis and decomposition of processes in organizations. *Information Systems Research* **14**(4) 337–355.

Bodin, L. D., L. A. Gordon, M. A. Loeb. 2005. Evaluating information security investments using the analytic hierarchy process. *Communications of the ACM* **48**(2) 79–83.

Cavusoglu, H., S. Raghunathan, W. T. Yue. 2008. Decision-theoretic and game-theoretic approaches to it security investment. *Journal of Management Information Systems* **25**(2) 281–304.

Cernauskas, Deborah, Anthony Tarantino. 2009. Operational risk management with process control and business process modeling. *Journal of Operational Risk* **4**(2) 3–17.

DoJ. 2004. Vallejo woman admits to embezzling more than $875,035.

Gordon, Lawrence A., Martin P. Loeb. 2002. The economics of information security investment. *ACM Trans. Inf. Syst. Secur.* **5**(4) 438–457.

Herath, H. S. B., T. C. Herath. 2008. Investments in information security: A real options perspective with bayesian postaudit. *Journal of Management Information Systems* **25**(3) 337–375.

ITRC. 2010. 2010 breach list.

Karp, Richard M. 1972. Reducibility among combinatorial problems. *Complexity of Computer Computations* 85–103.

Karp, R.M. 2010. Reducibility among combinatorial problems. M. Junger, T.M. Liebling, D. Naddef, G.L. Nemhauser, W.R. Pulleyblank, G. Reinelt, G. Rinaldi, L.A. Wolsey, eds., *50 Years of Integer Programming 1958-2008*. Springer Berlin Heidelberg, 219–241.

Krishnan, Ramayya, James Peters, Rema Padman, David Kaplan. 2005. On data reliability assessment in accounting information systems. *INFORMATION SYSTEMS RESEARCH* **16**(3) 307–326.

Kumar, R. L., S. Park, C. Subramaniam. 2008a. Understanding the value of countermeasure portfolios in information systems security. *Journal of Management Information Systems* **25**(2) 241–279.

Kumar, V., R. Telang, T. Mukhopadhyay. 2007. Optimally securing interconnected information systems and assets. *Workshop on the Economics of Information Security 2007*. Pittsburgh, PA.

Kumar, V., R. Telang, T. Mukhopadhyay. 2008b. Optimal information security architecture for the enterprise. *Working Paper-SSRN* .

Leskovec, Jure, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, Natalie Glance. 2007. Cost-effective outbreak detection in networks. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, San Jose, California, USA, 420–429.

Murray, R., W. E. Hart, C. A. Phillips, J. Berry, E. G. Boman, R. D. Carr, L. A. Riesen, J.-P. Watson, T. Haxton, J. G. Herrmann, R. Janke, G. Gray, T. Taxon, J. G. Uber, K. M. Morley. 2009. US environmental protection agency uses operations research to reduce contamination risks in drinking water. *Interfaces* **39**(1) 57–68.

Patterson, R., E. Rolland, M.L. Yeo. 2006. Security and privacy in outsourcing with customer-specified risk tolerance. *eJETA* **2**(1).

Ponemon Institute, Symantec. 2011. 2010 annual study: U.S. cost of a data breach.

Rodríguez, Alfonso, Eduardo Fernández-Medina, Juan Trujillo, Mario Piattini. 2011. Secure business process model specification through a UML 2.0 activity diagram profile. *Decision Support Systems* **51**(3) 446–465.

Watson, Jean-Paul, Regan Murray, William E. Hart. 2009. Formulation and optimization of robust sensor placement problems for drinking water contamination warning systems. *Journal of Infrastructure Systems* **15**(4) 330–339.

Weber, Ron. 1989. Controls in electronic funds transfer systems: A survey and synthesis. *Computers and Security* **8** 123–137.

Wood, Charles Cresson. 1990. Principles of secure information systems design. *Computers and Security* **9** 13–24.

4.6. Appendix

4.6.1 Data Generation

J node locations are randomly generated in a 100 by 100 space. We then randomly select an incident set, I, comprised of randomly selected subsets of nodes of a random size between 1 and J. The set of K controls is available to mitigate risk at each node in a the workflow. In addition, the cost of deploying a control, $k \in K$, at node $j \in J$ varies but is bounded by a value, $maxC$. Finally, the total budget for all controls is a percentage, $BudgetScale$ (BP), of total damages associated with the incidents when no controls are chosen. The data used for each iteration of the problem was generated in the following steps.

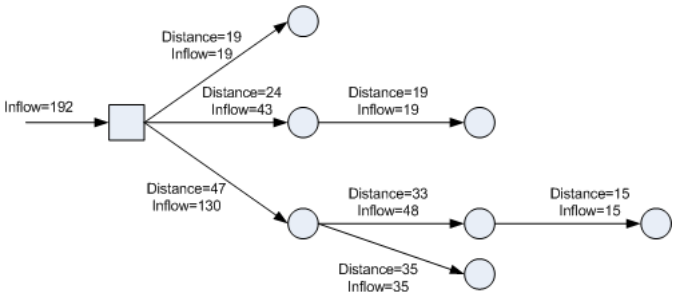
Nodes (Locations)

- Select coordinates in a 100×100 grid, randomly from a uniform distribution
- Calculate distances between each pair of nodes

Incidents

- For each incident, randomly select nodes to include. Each node has a 70% chance of being included in each incident.
- If an incident has no nodes chosen, randomly select one node for inclusion.
- For each incident, calculate the minimum spanning tree using Prim's algorithm and the distance matrix calculated in the node generation routine.
- For each incident, calculate every path in the spanning tree.
- For each incident, calculate the damage discount that would apply at each node if a control of type k was placed there, for all controls.
  - Calculate the total distance along all paths in the spanning tree.
  - For each node in the incident, calculate the total inflow to the node as the distance between the node and its upstream neighbour plus the total distance from the node to each endpoint in the node's path(s). See Figure 4.5 for an illustration.
  - The damage discount for each node is defined as the inflow to that node. That is, by placing a control k at node j, we are able to block damage from flowing any further and thus, discount the damage of incident i by the outflow distance at node j.

Figure 4.5: Calculating inflow.



Incident-Node-Control Applicability array (a_{ijk})

For each incident, for each node, for each control, if the node is part of this incident, then randomly decide if this control will be applicable here (i.e. set $a_{ijk} = 1$) by randomly drawing from a uniform distribution. There is a 50% chance that the control will be applicable. Otherwise, set the $a_{ijk} = 0$.

Table 4.6: Intervals used for generating uncontrolled damages.

Random Draw Interval	**Value Interval**	**P(incident i)**
$[0, 0.0001)$	[100 million, 10 billion]	0.0001
$[0.0001, 0.005)$	[10 million, 100 million]	0.0049
$[0.005, 0.05)$	[10,000, 10 million]	0.045
$[0.05, 0.95)$	[1000, 10,000]	0.90
$[0.95, 0.995)$	[10, 1000]	0.045
$[0.995, 0.9999)$	[0, 10]	0.0049
$[0.9999, 1)$	0	0.0001

Cost matrix (c_{jk}) For each activity, for each location, randomly select an integer between 0 and the maximum cost of controls, $maxC$.

Discount matrix (d_{ijk}) The damage discount is the reduction in damage if the incident is detected at this node with this control. For each incident, node, and control combination, define the damage discount as the inflow from node j in incident i as calculated when generating the incidents as illustrated in Figure 4.5.

Total expected damages for incident (D_i)
In essence, we provide a structure to the damages such that rare incidents have extremely high (or extremely low) damages and the most common incidents have medium damages.

- For each incident, calculate the total, uncontrolled damage realized if this incident occurs:
  - Draw a random probability from a $\mathcal{U}(0, 1)$ distribution and record.
  - According to the break down in Table 4.6, add a value drawn from the appropriate interval to the total damage discount recorded when generating the discount matrix and set the probability that incident i will occur, P(incident i), accordingly.
- Sum uncontrolled damages across all incidents to use in calculating the budget for this data set
- Calculated the expected damages (D_i) for each incident by multiplying the uncontrolled damages of incident i by the probability that incident i will occur.

Budget Calculate the budget by dividing the total expected damages for all incidents divided by the number of controls in this data set then multiply by a budget

scale (BP) defined for this data set. That is, $Budget = BP * \sum_{i \in I} E[D_i]/|K|$ where $|K|$ is the magnitude of K (i.e. number of controls).

4.6.2 Attack Description

Attack Simulation 1

In the first simulation, we select attacks from a uniform distribution with no consideration for the actual probability of seeing any particular incident.

1. Randomly select an integer, n from $\mathcal{U}(0, I)$, of attacks for this day.
2. Randomly select a set of n attacks from I, without replacement. That is, the same attack may not be seen more than once in any given day.
3. Calculate the actual damage incurred for each solution, given this set of attacks has occurred.

We can then total the realized damages over the entire period of 365 days and compare solutions.

Attack Simulation 2

The attacks in this simulation are drawn such that they follow the same probability distribution as the incidents.

1. For each incident in I, if the probability of this incident is greater than or equal to a randomly drawn value, $\mathcal{U}(0, 1)$, add it to the set of attacks for this day. Once again, the same attack may not be seen more than once in any given day.
2. Calculate the actual damage incurred for each solution, given this set of attacks has occurred.

We can then total the realized damages over the entire period of 365 days and compare solutions.

Chapter 5

General Discussion and Conclusions

The purpose of this thesis was to explore formal methods for determining the level and location for information security investments so that managers can make more informed choices when it comes to information risk management. In Chapter 2, market forces (in the form of customer reactions to security breaches) are used to explore the nature of security investments in a duopoly. This approach allows an examination of how the introduction of a minimum mandatory security spending level would impact firm profits, finding that under a particular customer reaction type (i.e. when firms are complements in loss) that a prisoner's dilemma arises where firms would be better off if they cooperated, voluntarily increasing security spending above equilibrium. Chapter 3 expands on this work by incorporating customer utility into the model, allowing for an examination into the effects of mandatory spending above equilibrium on total social welfare and pointing to areas where it would be an appropriate policy decision to require such a minimum level of spending. Finally, in Chapter 4, the allocation of a given information security investment is explored by formulating the flow risk reduction (FRR) problem then solving it to optimally select control placements in a workflow context.

The ultimate goal of this line of research is to build improved decision support tools for managers faced with managing the information risk in their enterprises. While the current practice of using intuition, experience, and best practices is an important starting point, managers who incorporate more formal methods, such as the proposed CTMC and FRR models, can further improve investment and allocation decisions as well as information security outcomes.

In terms of theoretical implications, Chapters 2 and 3 offer a glimpse into how Adam Smith's 'invisible hand'(Smith 1790) functions to regulate a market place. In particular, this work begins to articulate the links between customer reactions to information security breaches and firm incentives to invest appropriately to guard against risk. The goal in exploring this linkage is to challenge the notion of that information security investment as costs only and replace it with the idea that information security is a 'value add' that could lead to competitive advantage.

Practically speaking, understanding the nature of customer reactions is important for both managers and policy makers. As firms within an industry understand when it is in their best interest to cooperate, they should only need a coordination mechanism, such as the Information Sharing and Analysis Centers (ISACs) consid-

ered in Gal-Or and Ghose (2005), to monitor compliance. However, if it is not the case that cooperation will increase profits for firms, then a central planner will need to step in and mandate increased security investments where there would be an increase in total social welfare.

Along with deciding the appropriate level of investment, managers must decide what security measures to implement. By taking a process (workflow) view of the firm to consider the optimal placement of controls within those workflows, firms are better able to identify the high-value controls and to focus efforts there. By comparing the optimal solution of the flow risk reduction problem to heuristic approaches, this work helps managers identify new measures of investment effectiveness - the risk reduction and the return on risk reduction - as calculated by simulating attacks against each solution found.

Bibliography

Gal-Or, E., A. Ghose. 2005. The economic incentives for sharing security information. *Information Systems Research* **16** 186–208.

Smith, A. 1790. The theory of moral sentiments. URL `http://www.econlib.org/library/Smith/smMSCover.html`.

Chapter 6

Appendices

Permissions from my co-authors will be placed here before submitting the final Thesis document.