Essays on Individual Incentives and Private Information

by

Shouqiang Wang

Department of Business Administration
Duke University

Date: ______________________

Approved:

____
Alexandre Belloni, Co-Supervisor

____
Giuseppe Lopomo, Co-Supervisor

____
Peng Sun

____
Paul H. Zipkin

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Business Administration
in the Graduate School of Duke University
2011

Abstract

(Business Administration)

Essays on Individual Incentives and Private Information

by

Shouqiang Wang

Department of Business Administration
Duke University

Date: ____________________

Approved:

____
Alexandre Belloni, Co-Supervisor

____
Giuseppe Lopomo, Co-Supervisor

____
Peng Sun

____
Paul H. Zipkin

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University
2011

Abstract

This dissertation explores the incentive issues and strategic interactions among decentralized parties in three operations management environments: inventory systems, revenue management and healthcare policies. The first model studies the impact of multilateral asymmetric information about inventories in a two-echelon inventory systems. The second model applies optimization techniques to solve a monopolist's revenue problem where the seller's cost function is not separable across buyers with multidimensional private information. The third model uses a game-theoretical approach to study the decentralized resource allocation between self-interested countries to control an epidemic disease.

To my parents and my wife.

Contents

List of Tables

List of Figures

Acknowledgements

Faculty, friends, and family members have continuously helped me these years during my pursuit of my doctoral degree. I would like to take this opportunity to express my sincere gratitude to these individuals who made this dissertation possible through their kind guidance, assistance and support.

First and foremost, I am deeply grateful for my advisors, Alex Belloni and Pino Lopomo, who have been everything a student could ask for in a mentor. I cannot count how many Saturdays and Sundays they sacrificed to discuss research with me and provide invaluable insights. I learned what it means to be a true scholar through their high standard and rigorous attitude towards research. They also generously share with me their wisdom and experience as an academic professional, from which my own career will benefit for sure. I will always carry your example with me in my long journey ahead.

I must also thank Peng Sun, Francis de Véricourt and Paul Zipkin. Peng and Francis introduced me to my first research project. Part of this dissertation is attributed to the collaboration with them, which is a very enjoyable, hands-on experience and helped me change my mind-set from a knowledge-receiver to a knowledge-contributor. In particular, Peng offered enormous amount of invaluable advices during my time here at Fuqua. I feel indebted to him for his counsel on all matters great and small. Paul taught me a lot about inventory theory and operations management in general. I also benefit a lot, not only intellectually but also mentally, from all the casual but witty chats with him, which I will definitely miss a lot in the future.

My further thanks go to Fernando Bernstein, Gürhan Kök, Jeannette Song, Saša Pekeč, Bob Nau, Kevin Shang and all the other faculty members in decision science area and operation management area here at Fuqua. I feel so fortunate to learn from all these intelligent scholars and I gratefully appreciate all their generous help along the way.

It's been my privilege to meet many smart, funny, and kindhearted friends during my time here at Duke. To Kai, Zhengliang, Bobby, Zhuo, Dong, Victor, Liu, Fang, Saed, Hang, Fan, Bin, Zhenghua, and others I'm sure I have left out: thank you!

Last but certainly not the least, I am deeply indebted to my parents and my lovely wife:

To my parents: thank you very much for giving up so many things for me. Without your unconditional love and support, I wouldn't have gone this far: from Guiyang to Beijing and now Durham, from Beida to Uppsala and now Duke. You were and will always be my source of warmth, strength and courage.

To my incredible wife, Yingbo: thank you so much for standing by my side and cheering for me all the way through this. Your constant care and support are the key ingredients of this dissertation. Without you, my life wouldn't be a complete one. I love you.

1

Introduction

Traditional operations management research focuses on centralized systems, where a single firm is able to access to all the information along every stage of the supply chain and control all the decision making processes. However, in reality, the products and services flow among multiple parties who are motivated by their own interests and possess information unknown to the others. Conflicts of incentives are prevalent in many decentralized operations management environments. This dissertation explores the strategic interactions involved among different parties in three OM settings: inventory, revenue management and health care. The rest of the dissertation is organized as follows.

Chapter 2[1] studies the impact of multilateral private information about inventory levels on the overall performance of two-echelon supply chains. We focus on two polar cases: i) multiple suppliers serving a single retailer; and ii) a single supplier facing multiple retailers. In both cases the single party has all the bargaining power, i.e., it can commit to take-it-or-leave-it offers. In the former case, we find that the inventory policy that solves the centralized cost minimization problem remains optimal for the

[1] This chapter is based on the recently submitted paper (Belloni et al., 2011).

retailer, even in the presence of multilateral private information. Moreover, the retailer can appropriate the cost reduction of the entire supply chain. In the latter case, the contract that is optimal for the supplier induces a simple ranking allocation rule, based on critical fractiles of adjusted demand distributions that account for the incentives generated by private information. This rule reverses the order of all served retailers' inventory positions. For intermediate values of the supplier's initial inventory level, the rule "overshoots" the positions of some retailers, i.e., pushes them beyond their optimal levels without private information. An important managerial insight implied by our results is that supply chains in which parties that are more exposed to demand uncertainty have stronger bargaining positions perform better in terms of operating cost.

Chapter 3[2] is motivated by a classic problem in revenue management, where a monopolist offers products that can be sold at different quality levels. However, we consider the case in which all buyers have privately known valuations for each quality level (i.e. multidimensional types), and the seller has only a single object to sell (i.e., the seller's cost function is not separable). For example, the Federal Communications Commission (FCC) sells spectrum licenses with different levels of regulatory restrictions. The analysis of multidimensional mechanism design has proven significantly more complex than the one-dimensional case. Building upon Border and taking advantage of symmetry, we reformulate the seller's problem into a more parsimonious representation that eliminates the dimensionality dependence on the number of buyers. We next show that the associated infinite-dimensional optimization problem posed by the theoretical model can be approximated arbitrarily well by a sequence of finite-dimensional linear programming problems. To deal with the exponentially large set of constraints, we use a cutting-plane method that dynamically incorporates constraints as needed. In particular, we provide an efficient algorithm to verify the

[2] This chapter is based on the published paper (Belloni et al., 2010).

feasibility of the Border constraints and incentive compatibility constraints. This implies that our finite-dimensional approximation is solvable in polynomial time. The numerical solutions of the finite-dimensional approximations shed light on the qualitative nature of optimal solutions to the original infinite-dimensional problem. In particular, for the case with two quality levels, we propose an auction named "exclusive buyer mechanism" for practical implementation: the buyers compete in a second-price auction for the right to be the only buyer who gets to choose between buying the object of low quality at no additional cost, or the object of high quality for a prescribed add-on price.

Chapter 4[3] examines how two countries would allocate resources at the onset of an epidemic when they seek to protect their own populations. We model this situation as a game between selfish countries, where players strategically allocate their resources in order to minimize the total number of infected individuals in their respective populations during the epidemic. We study this problem when the initial number of infectives is very small, which greatly simplifies the analysis. We show in this framework that selfish countries always allocate their resources so as to bring the effective reproduction ratio below one and avoid a major outbreak. When a major outbreak is avoidable, we further identify the necessary and sufficient conditions under which the individual allocation decisions of selfish countries match the decision that a central planner would make in order to minimize the total number of infectives in the whole population (without distinguishing between countries). More specifically, the decentralized equilibrium allocation coincides with the centralized optimal allocation only in the cases of total selfishness (i.e., both countries keep their resources for themselves) and total altruism (i.e., one of the countries will give up all its resources to the other).

[3] This chapter is based on the published paper (Wang et al., 2009).

2

Private Information about Inventories

2.1 Motivation

Consider a two-echelon supply chain in which retailers face stochastic demand. The centralized problem of minimizing the total operating cost in this setting has been extensively studied, and the main features of the optimal allocation are well understood (Zipkin, 2000). Recently, several decentralized supply chains models have highlighted various aspects of the fundamental tension between the unilateral incentives of suppliers and retailers versus the benefits of system coordination, see e.g. Cachon and Lariviere (1999), Lee and Whang (1999), Cachon and Fisher (2000), Lee and Whang (2000), Cachon (2001), Cachon and Lariviere (2001), Chen (2007), Chen et al. (2001), Cachon and Lariviere (2005), Cachon and Zhang (2006), Cachon (2004) and Zhang (2010). In this chapter, we study the impact of multilateral private information regarding inventory levels on both the performance of the entire supply chain and the main features of the optimal allocation. We consider a single period model in which all parameters except initial inventory levels are common knowledge. Thus we isolate the effect of private information about inventory levels on the performance of the supply chain. More specifically, we address the following questions:

- Can a decentralized supply chain be as efficient as its centralized counterpart?
- What are the main properties of the optimal contract?

We find that the answers to these questions depend upon the structure of the supply chain. Our analysis focuses on two scenarios: i) multiple suppliers serving a single retailer (MSSR), and ii) a single supplier facing multiple retailers (SSMR). In both scenarios retailers face stochastic demands, all parties have privately known inventory levels, and the single party (the retailer in the MSSR case and the supplier in the SSMR case) has all the bargaining power, i.e. can commit to a "take-it-or-leave-it" contract offer. The optimal mechanisms minimize the overall cost (operating plus monetary) for the single party in the presence of strategic agents and multilateral private information about initial inventory levels.

The MSSR case applies to situations in which a large retail company procures its products from many suppliers on a regular basis. Typically, the retailer's inventory can hardly be monitored by any individual supplier. On the other hand, it may be costly for the retailer to gather accurate and timely information about all its suppliers' current inventory levels.

To motivate the SSMR case, consider a new hot-selling product during a holiday shopping season. For example, *Zhu Zhu Pet* became the hottest toy during the Christmas season of 2009 in the US (Mabrey and Janik, 2009; Wernau, 2009). The overnight frenzy created a drastic shortage, and the manufacturer Cepia LLC had to deal with the problem of restocking its retailers, having limited knowledge about their current inventory positions and local demand conditions.

Our first main result is that private information about inventories matters in the SSMR case, and is irrelevant in the MSSR case. The irrelevance result hinges on two reasons. First, the retailer's private information does not matter because its cost function is not directly affected by the suppliers' inventory levels. In the literature on mechanism design by an informed principal, this property is labeled "private values" (Maskin and Tirole, 1990). Second, each supplier's private information is inconsequential because the benefit of selling any additional unit is constant up to its inventory level. In our static environment, where production has already taken place, this is equivalent to the standard assumption of constant marginal handling cost. In turn, this is also due to the fact that suppliers are not directly exposed to any demand uncertainty. Thus in the MSSR case the presence of multilateral private inventory information does not hinder the implementability of the inventory policy that minimizes the total operating cost of the system (without private information). Moreover, the retailer can appropriate the full cost reduction of the entire supply chain using a straightforward fixed price contract.

In contrast, in the SSMR case the centralized optimal inventory policy is no longer optimal for the supplier. This is because the exposure to demand uncertainty makes each retailer's cost function non-linear – the cost reduction from an additional unit is decreasing.

Contracts that are optimal for the supplier are characterized by a simple ranking allocation rule, based on critical fractiles of an appropriately adjusted demand distribution, which accounts for the incentives generated by the presence of private information. Even in symmetric environments, where retailers differ only with respect to their privately known inventory levels, it is optimal for the supplier to reverse the order of the inventory positions of all served retailers. As it is well-known, the solution to the centralized problem (without private information) aims at balancing all retailers' inventory positions. Thus the "reverse ranking" of final inventory positions is entirely due to the presence of private information. We also show that the optimal allocation is more selective, i.e., it serves a weakly smaller number of retailers relative to the centralized allocation.

Confirming a well known feature of environments with private information, the total number of units sold by the supplier cannot exceed the one prescribed by the centralized solution. However, in our model, even with symmetric retailers, the quantities allocated to some retailers may "overshoot," i.e. go beyond the levels prescribed by the centralized solution. We provide necessary and sufficient conditions for the occurrence of overshooting, which, to the best of our knowledge, is unique in the literature.

The rest of the chapter is organized as follows. Sections 2.2 and 3.2 are devoted to the analyses of the MSSR model and the SSMR model respectively. In Section 2.4, we restrict our attention to the symmetric case of the SSMR model and establish that the optimal allocation entails reverse ranking of inventory positions and overshooting distortions. Section 2.5 contains numerical examples to illustrate our findings. All proofs and technical comments are relegated to the appendices.

2.1.1 Literature Review

Centralized supply chains have been extensively studied. The main properties of the allocations that minimize inventory operating costs in many different situations are well understood (see Allen (1958) and Zipkin (2000)). More recently, there has been a wealth of research on the tension between competition and coordination in decentralized supply chains. We defer to Cachon (2003) for an extensive literature review, and a general discussion on how various types of contracts can be used to manage incentive conflicts. In particular, Section 10 of Cachon (2003) focuses on issues generated by private information in supply chain environments. In this work we consider models that contribute to the decentralized supply chain literature by incorporating multilateral private information about inventory levels and examining different bargaining power configurations under capacity constraints.

In their pioneering work, Cachon and Lariviere (1999) study a model with a capacity constrained supplier serving multiple retailers that are privately informed about their optimal stocking level. They show that several empirically relevant contracts with a fixed wholesale price are vulnerable to manipulation.

Deshpande and Schwarz (2005) characterize optimal contracts in an abstract model similar to Cachon and Lariviere (1999). Both papers also consider the capacity choice problem under the optimal contract.

Corbett (2001), Corbett and de Groote (2000) and Cachon and Zhang (2006) study principal-agent problems in supply chain environments with one-sided private information on cost parameters. Lau et al. (2008) also consider private information on cost but allow for price-dependent demand. In a similar environment Zhang (2010) characterizes optimal contracts that specify a service-level attribute in addition to quantity.

Chen (2007) considers the optimal procurement auction for a buyer facing multiple suppliers with private information about their cost structures. Iyengar and Kumar (2008) study a similar model, where suppliers have private information about cost parameters as well as capacity levels.

Lee and Whang (2000) point out that informational decentralization and incentives are key factors for the overall supply chain performance. Lee and Whang (1999) investigate performance measurement schemes to improve the decentralized decisions. Porteus and Whang (1991) develop mechanisms which include an internal futures market to account for different incentives in the supply chain.

Zhang et al. (2009) investigate a model in which a supplier interacts over time with a retailer who has private information about its current inventory level. Their single period model is essentially identical to the SSMR case we consider, with a single retailer, no capacity constraint and no private information on the supplier's side. In the dynamic model, they show that under certain conditions on the parameter values and demand distributions it is optimal for the supplier to offer a batch contract in every period.

2.2 Multiple Suppliers and a Single Retailer

In this section we study a two-echelon supply chain system with M suppliers serving a single retailer.

The retailer is modeled as a newsvendor, with a privately known inventory level x_r, facing the standard trade-off between holding excess inventory, which entails a unit cost h_r, and not being able to meet its stochastic demand D_r which generates a per-unit penalty cost b_r. After receiving a total quantity Q from all suppliers, its expected operating cost can be written as

$$C_r(x_r, Q) = h_r \, \mathbb{E}_{D_r}\left[(x_r + Q - D_r)^+\right] + b_r \, \mathbb{E}_{D_r}\left[(x_r + Q - D_r)^-\right], \tag{2.1}$$

where the expectation is taken with respect to the demand D_r.

Similarly, supplier $i \in \{1, \ldots, M\}$ is privately informed about its inventory x_i and incurs unit holding and shipping/transaction costs h_i and c_i, respectively. Thus its total cost of supplying q_i units is given by

$$C_i(x_i, q_i) = h_i(x_i - q_i) + c_i q_i, \quad 0 \leq q_i \leq x_i. \tag{2.2}$$

The last two inequalities capture the facts that inventory cannot flow back from any retailer to the supplier and production has already taken place, hence the quantity sold q_i cannot exceed the inventory.

Relabeling suppliers if necessary, we can assume that $c_1 - h_1 \leq \cdots \leq c_M - h_M$. To avoid uninteresting cases we also assume that $h_i < h_r$ and $c_i < b_r$ for all $i = 1, \ldots, M$.

We assume that the retailer has all the bargaining power, i.e. can to commit to a take-it-or-leave-it contract offer. Its problem is to design a procurement contract that minimizes its total expected cost given by the sum of its monetary payments to the suppliers plus its operating cost. By the revelation principle (Myerson, 1979; Dasgupta et al., 1979), without loss of generality, we can restrict attention to the set of incentive compatible and individually rational revelation mechanisms. Formally, a revelation mechanism specifies the quantity $q_i\left(x_r, x\right)$ sold by supplier i and the payment $m_i(x_r, x)$ received from the retailer, for each inventory profile x_r, $x := (x_1, \ldots, x_M)$, and each $i = 1, ..., M$.

The incentive compatibility and individual rationality constraints can be stated as: for all x_r and x_r'

$$\mathbb{E}_x\left[C_r\left(x_r, \sum_{i=1}^{M} q_i(x_r, x)\right) + \sum_{i=1}^{M} m_i(x_r, x)\right] \leq \mathbb{E}_x\left[C_r\left(x_r, \sum_{i=1}^{M} q_i(x_r', x)\right) + \sum_{i=1}^{M} m_i(x_r', x)\right], \tag{IC$_r$}$$

and

$$\mathbb{E}_x\left[C_r\left(x_r, \sum_{i=1}^{M} q_i(x_r, x)\right) + \sum_{i=1}^{M} m_i(x_r, x)\right] \quad \leq \quad C_r(x_r, 0), \tag{IR$_r$}$$

where the expectation is taken with respect to the retailer's belief about the suppliers' inventory profile x. Similarly, and for each $i = 1, \ldots, M$, and for all x_i, x_i',

$$\mathbb{E}_{x_r, x_{-i}}\left[C_i(x_i, q_i(x_r, x)) - m_i(x_r, x)\right] \leq \mathbb{E}_{x_r, x_{-i}}\left[C_i(x_i, q_i(x_r, (x_{-i}, x_i'))) - m_i(x_r, (x_{-i}, x_i'))\right], \tag{IC$_i$}$$

where $x_{-i} := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_M)$, and

$$\mathbb{E}_{x_r, x_{-i}}\left[C_i(x_i, q_i(x_r, x)) - m_i(x_r, x)\right] \quad \leq \quad C_i(x_i, 0) \tag{IR$_i$}$$

where the expectation is taken with respect to supplier's i belief about all other parties' inventory levels.

The retailer's problem, given its privately known inventory level x_r, can be stated as[1]

$$\begin{aligned} &\min_{q,m} \quad \mathbb{E}_x \left[C_r \left(x_r, \sum_{i=1}^{M} q_i(x_r, x) \right) - \sum_{i=1}^{M} m_i(x_r, x) \right] \\ &\text{s.t.} \quad (\text{IC}_r),\ (\text{IR}_r),\ (\text{IC}_i),\ (\text{IR}_i) \text{ and } 0 \le q_i(x_r, x) \le x_i, \quad i = 1, \dots, M. \end{aligned} \tag{2.3}$$

Our first theorem establishes that the retailer's problem (2.3) is solved by the allocation function q_i° which also solves the centralized problem: for each inventory profile x_r, x

$$\begin{aligned} q^\circ(x_r, x) \in \arg\min_{q} \quad & C_r \left(x_r, \sum\nolimits_{i=1}^{M} q_i \right) + \sum\nolimits_{i=1}^{M} C_i(x_i, q_i) \\ & 0 \le q_i \le x_i \qquad \text{for} \;\; i = 1, .., M. \end{aligned} \tag{2.4}$$

Theorem 2.1. *In the MSSR model, the following contract is optimal for the retailer's problem. For any inventory profile x_r, x, there exists an integer i° such that, for each $i = 1, 2, \dots, M$*

$$q_i^\circ(x_r, x) = \begin{cases} x_i, & \text{if} \quad i < i^\circ, \\ G_r^{-1} \left(\frac{h_i + b_r - c_i}{h_r + b_r} \right) - x_r - \sum\limits_{j=1}^{i-1} x_j, & \text{if} \quad i = i^\circ, \\ 0, & \text{if} \quad i > i^\circ, \end{cases}$$

where G_r denotes the cumulative distribution function of the retailer's demand D_r, and

$$m_i^\circ(x_r, x) = (c_i - h_i)\; q_i^\circ(x_r, x). \tag{2.5}$$

Since the allocation rule q° defined in (2.4) already minimizes the entire system cost for any given inventory profile x_r, x without any incentive constraints, the proof of Theorem 2.1 consists in showing that the payments m° defined in (2.5) together with q° satisfy all incentive constraints and at the same time generate no

[1] Technically, the retailer faces an "informed principal problem." Appendix A.1 provides the details of the formal approach leading to (2.3).

cost reduction for any supplier. This immediately implies that the retailer is able to appropriate the full cost reduction of the entire supply chain. We record this observation for future reference in the following corollary.

Corollary 2.1. *In the MSSR model, the retailer appropriates the full cost reduction of the entire supply chain.*

The contract characterized in Theorem 2.1 would remain optimal even if all inventory levels were publicly known. Thus the presence of private information has no impact on the efficiency of entire supply chain in the MSSR model. Intuitively, there are two reasons for this results. First, the irrelevance of the retailer's private information is due to the fact that its cost function is not directly affected by the suppliers' private information.[2] Second, the irrelevance of each supplier's private information hinges on the property that its benefit of selling any additional unit is constant, which in turn follows from the fact that they are not directly exposed to any demand uncertainty.

Remark 2.1. *The contract characterized in Theorem 2.1 satisfies all incentive constraints "ex post": the retailer's constraints* (IC_r) *and* (IR_r) *for each realization of the suppliers' inventory profile* x*, and supplier* i*'s satisfies* (IC_i) *and* (IR_i), $i = 1, \dots, M$*, for all realizations of* x_{-i} *and* x_r*. In particular the implementation of the optimal contract does not require any knowledge by the retailer of the suppliers' beliefs about the others' inventory levels.*[3]

[2] This result was first established by Maskin and Tirole (1990). Mylovanov and Tröger (2008) provide sufficient conditions in more general environment for the irrelevance result. See Appendix A.1 for a more detailed discussion on this issue, which also arises in the SSMR model studied in the next sections.

[3] Formally, this contract satisfies *ex post* incentive compatibility and individual rationality for all parties.

2.3 A Single Supplier and Multiple Retailers

We now turn to the case with a single supplier, indexed by s, serving N retailers. As in the previous model, each party is privately informed about its own inventory level.

The supplier's cost of delivering q_i units to retailer $i = 1, \ldots, N$ is given by

$$C_s(x_s, q) = \ h_s \left(x_s - \sum_{i=1}^N q_i \right) + \sum_{i=1}^N c_i q_i, \qquad \text{for} \ \ \sum_{i=1}^N q_i \leq x_s, \ \ q \geq 0, \tag{2.6}$$

where x_s denotes its inventory level, h_s denotes the unit holding cost for unsold inventory, and c_i the shipping/transaction cost for any unit sold to retailer i. The supplier cannot allocate more than its inventory/capacity x_s, cannot discard its inventory, nor can receive inventory from retailers.

Retailer i is modeled as a newsvendor with unit holding cost $h_i \geq h_s$, unit penalty cost $b_i \geq c_i$, facing stochastic demand D_i with cdf and pdf denoted by G_i and g_i. Its expected cost function given by

$$C_i\left(x_i, q_i\right) = h_i \ \mathbb{E}_{D_i}\left[(x_i + q_i - D_i)^+ \right] + b_i \ \mathbb{E}_{D_i}\left[(x_i + q_i - D_i)^- \right], \tag{2.7}$$

where x_i denotes the retailer's (privately known) inventory level.

The supplier's problem[4] is to design a contract that minimizes its expected total cost given by its operating cost minus the payments made by the retailers. The supplier's belief about the profile $x := (x_1, \ldots, x_N)$ is represented by the joint probability distribution F^N on the rectangular support $\mathcal{X} = \prod_{i=1}^N \mathcal{X}_i = \prod_{i=1}^N [\underline{x}_i, \bar{x}_i]$ with marginals $F_1, \ldots, F_N$. Similarly, retailer i believes that retailer j's inventory is distributed accordingly to the cdf F_j and the supplier's inventory level is distributed accordingly to F_s with support $\mathcal{X}_s$.

[4] As in the previous section there is a potential signaling issue due to the fact that the contract designer also has private information. Potentially, the other participants may make inferences about the designer's private information through the offered contract and exploit this information. Appendix A.1 provides a more detailed discussion of this issue.

The incentive constraints that determine the feasible set for the supplier's problem are: for each retailer $i = 1, \ldots, N$, and all $x_i \in \mathcal{X}_i$, $x_i' \in \mathcal{X}_i$,

$$\mathbb{E}_{x_s,x_{-i}}\left[C_i\left(x_i, q_i\left(x_s, x\right)\right) + m_i\left(x_s, x\right)\right] \leq \mathbb{E}_{x_s,x_{-i}}\left[C_i\left(x_i, q_i\left(x_s, \left(x_i', x_{-i}\right)\right)\right) + m_i\left(x_s, \left(x_i', x_{-i}\right)\right)\right], \tag{IC$_i$}$$

for all $x_i \in \mathcal{X}_i$

$$\mathbb{E}_{x_s,x_{-i}}\left[C_i\left(x_i, q_i\left(x_s, x\right)\right) + m_i\left(x_s, x\right)\right] \leq C_i\left(x_i, 0\right), \tag{IR$_i$}$$

for any $x_s, x_s' \in \mathcal{X}_s$,

$$\mathbb{E}_x\left[C_s\left(x_s, q(x_s, x)\right) - \sum_{i=1}^{N} m_i(x_s, x)\right] \leq \mathbb{E}_x\left[C_s\left(x_s, q(x_s', x)\right) - \sum_{i=1}^{N} m_i(x_s', x)\right], \tag{IC$_s$}$$

and for all $x_s \in \mathcal{X}_s$

$$\mathbb{E}_x\left[C_s\left(x_s, q(x_s, x)\right) - \sum_{i=1}^{N} m_i(x_s, x)\right] \leq C_s\left(x_s, 0\right), \tag{IR$_s$}$$

where $x_{-i} := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_M)$, and all expectations are taken to their corresponding distributions.

Formally, the supplier's problem can be stated as

$$\begin{aligned} \min_{q,m} \quad & \mathbb{E}_x\left[C_s\left(x_s, q(x_s, x)\right) - \sum\nolimits_{i=1}^{N} m_i(x_s, x)\right] \\ \text{s.t.} \quad & (\mathrm{IC}_s),\ (\mathrm{IR}_s),\ (\mathrm{IC}_i),\ (\mathrm{IR}_i),\ q_i(x_s, x) \geq 0 \quad i = 1, \ldots, N \\ & \sum\nolimits_{i=1}^{N} q_i(x_s, x) \leq x_s. \end{aligned} \tag{2.8}$$

The results for the SSMR model will be established under the following regularity conditions.

Assumption 2.1. *The survival function of the demand $\bar{G}_i(\cdot) = 1 - G_i(\cdot)$ is log-concave for $i = 1, 2, \ldots, N$.*

Assumption 2.2. *The retailers' inventory levels $x_1, ..., x_N$ are distributed independently, with marginal densities $f_1, ..., f_N$, and cumulative distributions $F_1, ..., F_N$. For each $i = 1, , ..., N$ the reversed hazard rate $\frac{f_i}{F_i}$ is non-increasing.*

Assumption 2.1 allows for a variety of demand distributions. It is weaker than the log-concavity of the probability density functions g_i which is imposed in Zhang et al. (2009). Assumption 2.2 is standard in the literature (e.g. Myerson (1979), Corbett (2001) and Zhang et al. (2009)). We are now ready to state the main result of this section.

Theorem 2.2. *In the SSMR model, under Assumptions 2.1 and 2.2, the following contract is optimal for the supplier's problem (2.8): for each $x \in \mathcal{X}$ and $x_s \in \mathcal{X}_s$*

$$q^\star(x_s, x) \in \arg\min_{q \geq 0} \quad \pi^\star(x, q) \qquad \sum\nolimits_{i=1}^{N} q_i \leq x_s \tag{2.9}$$

where $\pi^\star(x, q) := \sum_{i=1}^{N} \left[(c_i - h_s)q_i + C_i(x_i, q_i) + (h_i + b_i)\frac{F_i(x_i)}{f_i(x_i)} G_i(x_i + q_i) \right]$, *and each retailer payment is given by*

$$m_i^\star(x_s, x) = C_i(x_i, 0) - C_i(x_i, q_i^\star(x_s, x)) - (h_i + b_i) \int_{x_i}^{\bar{x}_i} \left[G_i(z + q_i^\star(x_s, (z, x_{-i}))) - G_i(z) \right] dz.$$

The function $\pi^\star(x, q)$ is given by the sum of all retailers' "virtual costs".[5] It adjusts the supply chain costs by taking into account the incentives generated by the presence of private information.

The allocation $q^\star$ characterized in Theorem 2.2 may differ significantly from the centralized solution, due to the presence of the retailers' private information in the SSMR model. Indeed, the supplier's private information is irrelevant – the contract

[5] The expression "virtual utility" was introduced by Myerson (1979).

of Theorem 2.2 would remain optimal even if x_s were publicly known. This is because the supplier's cost function is not directly affected by the retailers' private information.[6] The retailers' private information limits the supplier's gains because each retailer's cost reduction of receiving additional units is decreasing. This feature in turn is a consequence of the retailers bearing the entire supply chain risk.

Remark 2.2. *The contract characterized in Theorem 2.2 satisfies* (IC_i) *and* (IR_i) *for each* $i = 1, \ldots, N$*, for all realizations of* x_{-i} *and* x_s*. Therefore, the implementation of the optimal contract does not require any knowledge by the supplier of the retailers' beliefs about the others' inventory levels.*

In the remainder of this section and in Section 2.4 we derive additional structural properties of the optimal contract.

2.3.1 Fractile-Based Interpretation of the Optimal Allocation

This section provides a characterization of the optimal contract derived in Theorem 2.2 based on critical fractiles of appropriately adjusted demand distributions. This allows for a direct comparison with the centralized solution. We begin by defining the following family of adjusted cumulative distribution functions, parameterized by the inventory level x_i

$$\widetilde{G}_i(y|x_i) := G_i(y) + \frac{F_i(x_i)}{f_i(x_i)} g_i(y), \quad i = 1, \ldots, N. \tag{2.10}$$

In general, $\widetilde{G}_i(\cdot|x_i)$ is not a cumulative probability distribution as it can be both non-monotone and larger than one. It turns out however that, under Assumption 2.1, $\min\{1, \widetilde{G}_i(\cdot|x_i)\}$ is a cumulative distribution function,[7] and its fractiles agree

[6] This result was first established by Maskin and Tirole (1990). Mylovanov and Tröger (2008) provide sufficient conditions in more general environment for the irrelevance result.

[7] Under Assumption 2.1, the non-monotonicity can only occur if $\widetilde{G}_i(y|x_i) \geq 1$.

with the inverse of $\widetilde{G}_i(\cdot|x_i)$ restricted to $(0,1)$. This allows us to speak meaningfully, albeit informally, of the "fractiles" of $\widetilde{G}_i(\cdot|x_i)$.

To formally state our results let

$$\alpha\text{-fractile of } \widetilde{G}_i(\cdot|x_i) := \inf\{y \in [\underline{x}_i, \bar{x}_i] \ : \ \widetilde{G}_i(y|x_i) \geq \alpha\}, \quad i = 1, \ldots, N, \quad (2.11)$$

and

$$\mu_i(x_i) := (h_i + b_i)\widetilde{G}_i(x_i|x_i) - (h_s + b_i - c_i). \quad (2.12)$$

Note that the index μ_i coincides with the derivative of $\pi^\star(x, q)$ with respect to q_i evaluated at x_i. The next theorem characterizes the optimal allocation defined in (2.9) by first ranking the retailers based on indices $\mu_i(x_i)$, and then expressing the allocated quantity to each retailer as a fractile of the adjusted distribution $\widetilde{G}_i$.

Theorem 2.3 (Allocation Rule via Ranking and Fractiles)**.** *In the SSMR model, fix an inventory profile $x_s \in \mathcal{X}_s$, $x \in \mathcal{X}$, and relabel retailers, if necessary, so that*

$$\mu_1(x_1) \leq \mu_2(x_2) \leq \cdots \leq \mu_N(x_N). \quad (2.13)$$

Under Assumption 2.1, there exist a Lagrange multiplier $U^\star \in [\min_{i=1,\ldots,N}\{c_i - h_s - b_i\}, 0]$ and an integer $n^\star$ such that the solution $q^\star$ to (2.9) is determined by:

(i) for all $i \leq n^\star$, we have $q_i^\star > 0$ and $x_i + q_i^\star = \left(\frac{U^\star + h_s + b_i - c_i}{h_i + b_i}\right)$-fractile of $\widetilde{G}_i(\cdot|x_i)$;

(ii) for all $j > n^\star$, we have $q_j^\star = 0$ and $x_j \geq \left(\frac{U^\star + h_s + b_j - c_j}{h_j + b_j}\right)$-fractile of $\widetilde{G}_j(\cdot|x_j)$; and

(iii) $U^\star \cdot (\sum_{i=1}^N q_i^\star - x_s) = 0$.

The idea of ranking retailers to characterize optimal allocations under a capacity constraint is already present in Zipkin (1980).[8] Theorem 2.3 shows that this characterization extends to the problem in (2.9) where the presence of private information

[8] Zipkin (1980) also provides an efficient algorithm to find the index $n^\star$ and the multiplier $U^\star$.

induces non-convexities in the objective function. The expressions in (i)–(iii) generalize the solution of the classic newsvendor problem to our setting with limited capacity, multiple retailers, and private information. The Lagrange multiplier $U^\star$ and the adjusted demand distribution $\widetilde{G}_i$ account for the capacity constraint and the presence of private information, respectively.

2.3.2 Distortions of the Optimal Allocations

This section examines distortions due to the presence of private information under limited capacity with multiple retailers. In particular, we compare the optimal allocations derived in Theorem 2.2 with the solutions of two standard benchmarks:[9] the classic newsvendor problem without capacity constraint,

$$\bar{q}_i(x_i) = \left[G_i^{-1}\left(\frac{h_s + b_i - c_i}{b_i + h_i}\right) - x_i\right]^+ \quad \text{for } \; x_i \in \mathcal{X}_i, \; i = 1, \ldots, N; \tag{2.14}$$

and the centralized distribution system in which a limited quantity x_s is allocated to minimize the overall cost of the supply chain

$$q_i^\circ(x_s, x) = \left[G_i^{-1}\left(\frac{U^\circ + h_s + b_i - c_i}{b_i + h_i}\right) - x_i\right]^+ \quad \text{for } \; x_s \in \mathcal{X}_s \;\; x \in \mathcal{X}, \; i = 1, \ldots, N, \tag{2.15}$$

where the Lagrange multiplier $U^\circ = U^\circ(x_s, x) \leq 0$ ensures that $\sum_{i=1}^N q_i^\circ(x_s, x) \leq x_s$.

The case in which the supplier has no private information and unlimited capacity collapses to the single retailer case studied in Zhang et al. (2009), under slightly different assumptions.[10] Indeed, once the capacity constraint is removed, the problem becomes separable across retailers and all results in Zhang et al. (2009) extend immediately to the multiple-retailer case.

[9] Both standard benchmarks have no private information, i.e. the inventory levels of retailers are known by the supplier.

[10] Zhang et al. (2009) also considered an interesting dynamic mechanism design problem which we do not consider here.

We begin by comparing the total quantities allocated in (2.9), (2.14) and (2.15).

Proposition 2.1. *In the SSMR model, under Assumptions 2.1 and 2.2, we have, for all* $x_s \in \mathcal{X}_s$ *and* $x \in \mathcal{X}$*,*

$$\sum_{i=1}^{N} q_i^\star(x_s, x) \leq \sum_{i=1}^{N} q_i^\circ(x_s, x) \leq \sum_{i=1}^{N} \bar{q}_i(x_i). \tag{2.16}$$

The first inequality is due to the presence of private information. This "friction" effect is familiar from standard mechanism design results. The second inequality follows directly from removing the capacity constraint in the centralized problem. The next proposition pertains to individual quantities.

Proposition 2.2. *In the SSMR model, under Assumptions 2.1 and 2.2, for all* $x_s \in \mathcal{X}_s$*,* $x \in \mathcal{X}$*, and* $i = 1, \ldots, N$*, we have*

$$q_i^\star(x_s, x) \leq \bar{q}_i(x_i) \quad \textit{and} \quad q_i^\circ(x_s, x) \leq \bar{q}_i(x_i).$$

As in Proposition 2.1, the first inequality is due to private information and the second to the removal of the capacity constraint.

The individual allocations $q_i^\star$ and q_i°, however, cannot be ranked, in general. This is because the simultaneous presence of private information and limited supplier's inventory creates two conflicting effects: (i) the "friction" effect described above, and (ii) a "redistribution" effect. Due to friction effect the supplier restricts quantities allocated to high inventory retailers. This allows the supplier to redistribute the released units to low inventory retailers. The redistribution effect is formalized in the next proposition.

Proposition 2.3. *In the SSMR model, under Assumptions 2.1 and 2.2, we have, for all* $x_s \in \mathcal{X}_s, x \in \mathcal{X}$*,*

$$U^\circ(x_s, x) \leq U^\star(x_s, x) \leq 0. \tag{2.17}$$

Proposition 2.3 implies that the adjusted fractiles are larger under private information, hence the capacity released by the friction effect is reallocated more profitably for the supplier. The lack of exact knowledge regarding the retailers inventory levels lowers the marginal value of additional capacity units for the supplier relative to the centralized distribution system.

2.4 Single Supplier with Multiple Symmetric Retailers

In this section, we restrict attention to the case where all the retailers are *ex ante* symmetric. This allows us to isolate the distortions arising purely due to private information about inventories.[11] By ex ante symmetric retailers we mean that the shipping/transaction costs for all retailers are the same $c = c_i$, and they have the same inventory cost function $C(\cdot) = C_i(\cdot)$ for $i = 1, \ldots, N$. In particular the holding and penalty costs are the same $h = h_i$, $b = b_i$ and all retailers face a stochastic demand with the same distribution $G = G_i$ and have the same belief $F_i = F$ about the others' inventory levels over the support $[\underline{x}, \bar{x}]$, $i = 1, \ldots, N$.

In this symmetric environment, it is well known that, without private information, the optimal allocation $q^\circ(x_s, x)$ for the centralized distribution system balances the final positions of the $n^\circ(x_s, x)$ retailers that are served, see Allen (1958). Relabeling retailers if necessary so that $x_1 \leq x_2 \leq \cdots \leq x_N < x_{N+1} = \infty$, we have

$$x_i + q_i^\circ = \frac{x_s + \sum_{i=1}^{n^\circ} x_i}{n^\circ} \quad \text{for all} \quad i \leq n^\circ := \min_{1 \leq n \leq N} \left\{ n : \frac{x_s + \sum_{i=1}^{n} x_i}{n} \leq x_{n+1} \right\}, \qquad (2.18)$$

if the capacity constraint is binding, and $q_i^\circ = [G^{-1}\left(\frac{h_s+b-c}{b+h}\right) - x_i]^+$ as in the newsvendor problem, otherwise. This yields the following balanced final positions among served retailers

$$x_1 + q_1^\circ \;=\; x_2 + q_2^\circ \;=\; \cdots \;=\; x_{n^\circ} + q_{n^\circ}^\circ \;\leq\; x_{n^\circ+1} \;\leq\; \cdots \;\leq\; x_N. \qquad (2.19)$$

[11] Asymmetry among retailers introduce additional distortions.

2.4.1 Reverse Ranking

In general, the balancing policy defined in (2.18) and (2.19) is no longer optimal under private information. For example, consider the unlimited capacity case with two retailers and F uniform. When $x_1 < x_2 < G^{-1}\left(\frac{h_s+b-c}{b+h}\right)$, we have $\widetilde{G}(y|x_1) < \widetilde{G}(y|x_2)$ which leads to the unbalanced final inventory positions $x_1 + q_1^* > x_2 + q_2^*$ by Theorem 2.3.

The next theorem generalizes this example and establishes that, for all served retailers, the order of their inventory levels is completely reversed by the optimal allocation to the problem (2.9).

Theorem 2.4 (Reverse Inventory Positions)**.** *In the SSMR model with symmetric retailers, for any $x_s \in \mathcal{X}_s$, $x \in \mathcal{X}$, relabel retailers, if necessary, so that*

$$x_1 \leq x_2 \leq \cdots \leq x_N,$$

and let $q^\star = q^\star(x_s, x)$ denote the solution to (2.9). If Assumptions 2.1 and 2.2 hold, there exists an integer $n^\star = n^\star(x_s, x)$ such that $q_i^\star > 0$ if and only if $i \leq n^\star$. Furthermore

(i) the final positions of all served retailers are reversed, i.e.

$$x_1 + q_1^\star \geq x_2 + q_2^\star \geq \cdots \geq x_{n^\star} + q_{n^\star}^\star \; . \tag{2.20}$$

(ii) fewer retailers are served relative to the centralized distribution system, i.e.

$$n^\star \leq n^\circ. \tag{2.21}$$

Under our assumptions, ranking retailers with respect to the indices μ_i is e-quivalent to ranking them by their inventory levels x_i. This allows us to use the characterization derived in Theorem 2.3.

2.4.2 Overshooting

In both allocations $q^\star$ and q^o, retailers with lower inventory levels receive larger shares of the available quantity x_s. The inequalities in (2.20) however imply that, under private information, this imbalance is larger. Nonetheless, as remarked in the previous section, the individual allocations $q_i^\star$ and q_i° cannot be ranked in general. This subsection will shed some light on this comparison.

The next result provides a characterization of a new type of distortion induced by the presence of private information in our setting, which is a consequence of the interplay between the "friction" effect and "redistribution" effect. We provide necessary and sufficient conditions for a retailer to be allocated *strictly more* under private information than it would be in the optimal allocation of the centralized distribution system (i.e., $q_i^\star > q_i^\circ$ for some i). We refer to this type of distortion as "overshooting".

Theorem 2.5 (Overshooting)**.** *In the SSMR model with symmetric retailers, let $N \geq 2$. Suppose that Assumptions 2.1 and 2.2 hold, and $F(x_i)/f(x_i)$ is strictly increasing over $[\underline{x}, \bar{x}]$. Fix $x_s \in \mathcal{X}_s$ and $x \in \mathcal{X}$, relabeling retailers, if necessary, by their inventory levels, with $x_1 < x_2$. Then, overshooting occurs*

$$q_i^\star(x_s, x) > q_i^\circ(x_s, x) \quad \textit{for some } i = 1, \ldots, N$$

if and only if

$$(i)\ x_s > x_2 - x_1 \qquad \textit{and} \qquad (ii)\ \ x_s \ < \ \sum_{i=1}^{n^\circ} \left[\widetilde{G}^{-1} \left(\frac{h_s + b - c}{h + b} \ \middle|\ x_1 \right) - x_i \right].$$

The proof also shows that overshooting occurs if and only if it occurs for the retailer with the lowest initial inventory, i.e. $q_1^\star(x_s, x) > q_1^\circ(x_s, x)$.

Condition (i) in Theorem 2.5 guarantees that it is optimal for the supplier to serve at least two retailers in the centralized distribution system, i.e. $n^\circ \geq 2$. Otherwise, by (2.16) and (2.21) we would have that $q_1^\star \leq q_1^\circ$, hence no overshooting.

Condition (ii) can be rewritten as $\widetilde{G}([x_s + \sum_{i=1}^{n^\circ} x_i]/n^\circ | x_1) < (h_s + b - c)/(h + b)$, in light of (2.18). Since $[x_s + \sum_{i=1}^{n^\circ} x_i]/n^\circ$ is non-decreasing in x_s, this shows that x_s cannot be too large which in turn implies that capacity is binding at the centralized allocation q°. With unlimited capacity, the problem would collapse to the single retailer case, studied in Zhang et al. (2009), in which overshooting cannot occur. However, overshooting can also occur when capacity is not binding for $q^\star$.

Deviations from the centralized solution similar to the ones discussed Proposition 2.1 and 2.2, and in particular the friction effect generated by private information that tends to reduce the total quantity allocated in the system, have been observed in many other supply chain environments, e.g. Cachon and Lariviere (1999), Corbett (2001), Deshpande and Schwarz (2005), and Zhang et al. (2009). However, in many of these models without any capacity constraint the problem decomposes into a family of independent single-agent problems. The overshooting distortion can arise only in the presence of multiple retailers and limited capacity. To the best of our knowledge, Theorem 2.5 is the first result that demonstrates that an agent (a retailer) can be allocated in the optimal mechanism more than it would in the absence of private information even within symmetric environments.

2.5 Numerical Examples

This section presents numerical simulations that illustrate the impact of private information in the SSMR model. To isolate the distortion effects due to the presence of private information about inventory, we focus on the symmetric case.

Consider the following numerical example: cost parameters $h_s = 0.3,\ c = 0.1,\ h = 0.5, b = 0.5$; each retailer's demand distribution is $D \sim \text{Unif}(0, 1)$; and identical independent probability assessment for the inventory level x_i given by $F_i \sim \mathrm{Unif}(0,1)$.

For each supplier's capacity level $x_s = 0.1, 0.2, \ldots, 3$ and number of retailers $N = 1, 5, 10, 50$ and each x_i drawn according to F, we compute the centralized solution $q^\circ(x_s, x)$ and the decentralized solution $\{q^\star(x_s, x), m^\star(x_s, x)\}$, based on which we computed several measures to illustrate the main results of Sections 3.2 and 2.4.

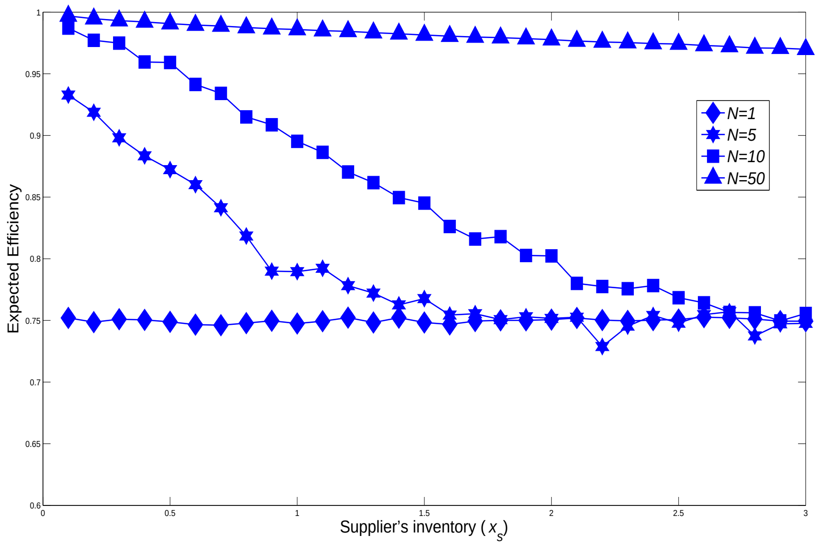


FIGURE 2.1: Expected efficiency ratios.

Letting $C(x_s, x, q) = C_s(x_s, 0) - C_s(x_s, q) + \sum_{i=1}^{N} [C_i(x_i, 0) - C_i(x_i, q_i)]$ denote the overall cost reduction of the supply chain, we define the expected efficiency of the decentralized solution as $\mathbb{E}_x \left[C(x_s, x, q^\star)\right] / \mathbb{E}_x \left[C(x_s, x, q^\circ)\right]$ (relative to the centralized solution). This is a measure of how detrimental private information about inventory is to the overall performance of the supply chain. As can be seen from Figure 2.1, the expected efficiency is always below 1 due to the "friction effect". The expected efficiency decreases with x_s and increases with N increases. Intuitively, as the capacity x_s becomes more stringent while the number of retailers becomes larger, the competition for the limited capacity becomes more intense. As a result, the "redistribution effect", which promotes efficiency, becomes larger relative to the "friction effect".

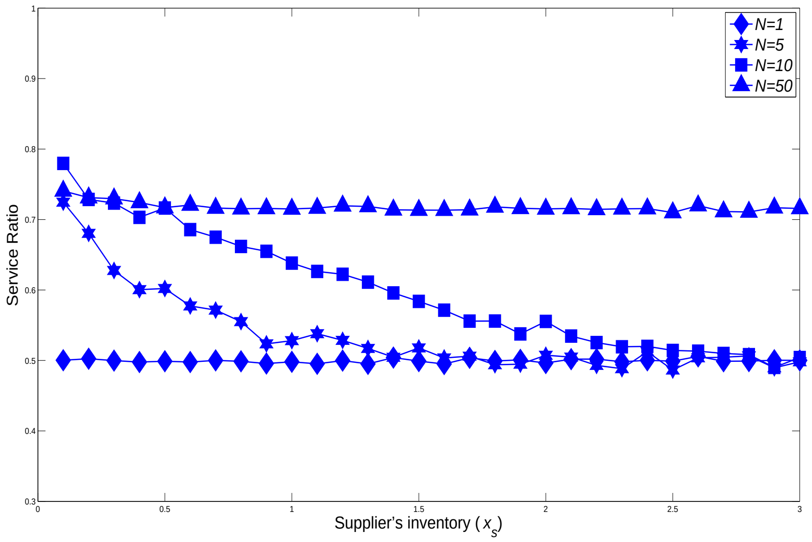


FIGURE 2.2: Service ratios.

We define the "service ratio" between the decentralized solution and the centralized solution as $\mathbb{E}_x\left[n^{\star}(x_s,x)\right]/\mathbb{E}_x\left[n^{\circ}(x_s,x)\right]$. As can be seen from Figure 2.2, the service ratio is always below 1, which illustrates part (ii) of Theorem 2.4.

In order to illustrate the presence of overshooting distortions, we compute the probability $\mathbb{P}_x\left[q_1^{\star}(x_s,x) > q_1^{\circ}(x_s,x)\right]$, and the expected number of overshooting retailers $\mathbb{E}_x\left[|\{i : q_i^{\star}(x_s,x) > q_i^{\circ}(x_s,x)\}|\right]$. Figures 2.3 and 2.4 illustrate these two measurements. The u-shape patterns in both figures indicate that Conditions (i) and (ii) in Theorem 2.5 are more likely to hold for intermediate values of capacity.

2.6 Concluding Remarks

In this chapter we investigated a single period two-echelon supply chain in which all parties have private information regarding their inventory level. Thus this work is complementary to most literature on decentralized supply chains with private information, which focuses on unilateral private information on costs.

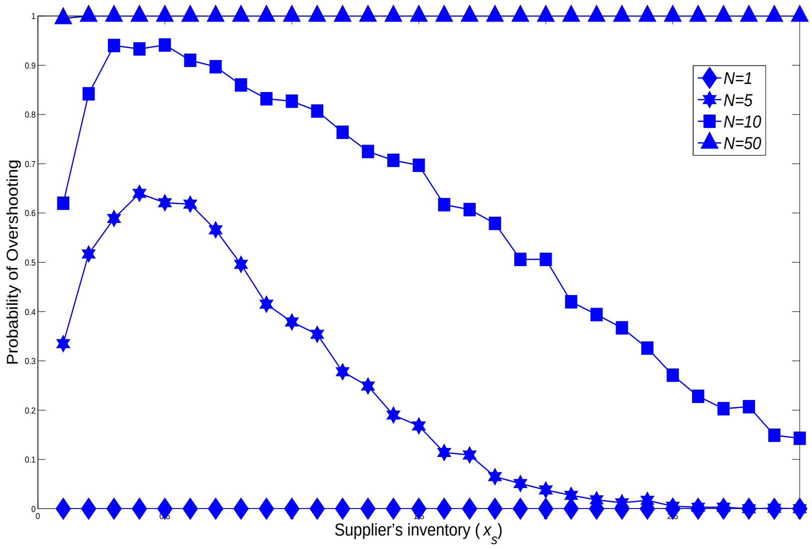


FIGURE 2.3: Probabilities of overshooting.

Two bargaining power configurations were studied. In the MSSR case, where the retailer bears the entire supply chain risk and has all the bargaining power, we find that the centralized allocation is optimal and allows for the full extraction of all cost reductions by the retailer. This results hinges critically on the assumption that the supplier's cost function is linear up to its initial inventory level. In our static environment, where production has already taken place, this is equivalent to the standard assumption of constant marginal handling cost.

In the SSMR case, where the supplier has all the bargaining power, the optimal allocation can be significantly different from the centralized solution. In particular, the total quantity sold tends to be lower. We characterize the contract via a ranking rule based on critical fractiles of a virtual demand distributions, which is adjusted to account for the incentives created by private information. Under the optimal allocation for symmetric environments, retailers with smaller initial inventory levels are brought to larger final inventory positions (reverse ranking), and in some cases final allocations exceed the ones prescribed by the centralized solution.

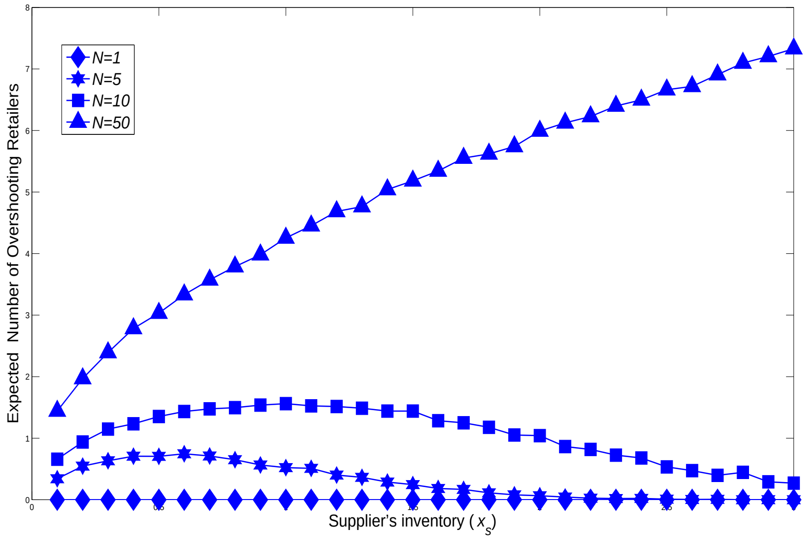


FIGURE 2.4: Expected number of overshooting retailers.

Several interesting questions remain beyond the scope of this chapter. Zhang et al. (2009) have studied a dynamic extension of our SSMR with one retailer and unlimited supplier's inventory. Extending both SSMR and MSSR models to dynamic settings appears to be the natural next step in this line of research. It would also be interesting to develop practical procedures for implementing the optimal contract in the SSMR case. We note however that even without a practical implementation, the characterization of the optimal mechanism provides useful insights, as well as a benchmark for evaluating the performance of any alternative mechanism. Finally, intermediate bargaining power structures should also be investigated.

3

Revenue Management with Multidimensional Screening

3.1 Motivation

This chapter revisits the classic problem of a monopolist offering a product that can be sold at different quality levels. The literature on this subject dates back to Mussa and Rosen (1978), who studied a model in which the seller's cost function is separable across buyers, and the buyers' private information is represented by a one-dimensional parameter (see also Maskin and Riley (1984)). In his seminal paper on optimal auctions (Myerson, 1981), Myerson dealt with the case in which the seller has a single object (hence the cost function is not separable across buyers), and showed that, under mild regularity assumptions, any auction in a large class that includes several popular formats, e.g. the English auction, the Dutch auction, and the first-price auction, maximizes the expected profit among all feasible selling procedures. Myerson's results immediately extend to the case in which the seller can choose the object's quality, as long as one maintains the assumption that each buyer's private information is represented by a one-dimensional variable.

This chapter considers the case in which all buyers have multidimensional types and the seller's cost function is not separable across buyers. We show that the associated infinite-dimensional optimization problem posed by the theoretical model can be approximated arbitrary well by a sequence of finite-dimensional linear programming problems. In addition we develop an efficient computational method to solve the resulting linear programming problem efficiently (in theory and practice).

The analysis of the multidimensional case has been proven significantly more complex, and has produced results, under the assumption that costs are separable across buyers, that seem unable to explain the use of popular selling procedures in terms of their optimality for the seller. Rochet and Choné (1998) provide general characterizations of optimal mechanisms in a multidimensional version of Mussa and Rosen's model, which had been previously investigated by Wilson (1993) and Armstrong (1996). Manelli and Vincent (2007) study the closely related problem of a monopolist selling several objects to a single buyer which has a privately known value for each.

The main technical difficulty that emerges in this stream of literature is that, for generic parameter values, "non-local" incentive compatibility constraints are binding, hence it is optimal for the seller to "bunch" types of consumers in various regions, i.e. to induce them to choose the same configuration of quality levels. Rochet and Choné (1998) deal with this issue by developing a "sweeping operator" which allows them to obtain an implicit characterization of optimal solutions. Manelli and Vincent (2007) show that a large variety of mechanisms can be optimal for the seller, depending on her prior beliefs about the buyer's type.

The main difference between the model studied in the present chapter and the models described above is that in our case the seller's cost function is not separable across buyers. In particular, we assume that the seller has just one object, as in Myerson's optimal auction problem, but can set the object's quality equal to any "grade"

Table 3.1: Summary of the relation with literature.

	One-dimensional types	Multidimensional types
Separable Cost	Mussa and Rosen (1978), Maskin and Riley (1984)	Wilson (1993), Armstrong (1996), Rochet and Choné (1998), Manelli and Vincent (2007)
Non-separable Cost	Myerson (1981)	This Chapter

within a finite set. Each buyer has a privately known willingness to pay for each quality level. A motivating example would be Federal Communications Commission (FCC) selling spectrum licenses with different levels of regulatory restrictions. To the best of our knowledge, this problem has not been studied in detail. Table 3.1 summarizes how this chapter relates to the previous literature.

Our contribution can be described as follows. First, we reformulate the seller's problem using "interim" variables only[1]. This reformulation also eliminates the dimensionality dependence on the number of buyers, a property which is desirable in practical applications where the number of buyers tends to be significantly larger than the number of quality levels. This approach hinges on the adaptation of results obtained by Border (1991) to our environment with multiple quality levels.

Next we establish that the infinite-dimensional optimization problem posed by the theoretical model can be approximated arbitrary well by a sequence of finite-dimensional linear programming problems, under standard assumptions on the buyers' valuations. Indeed this result cannot hold for arbitrary specifications of the

[1] In mechanism design terminology, "ex post" variables depend on all buyers' privately known types; while "interim" variable are obtained after integrating out all but one buyers' types.

buyers' valuation, because the set of all feasible mechanisms includes discontinuous and non-smooth functions, for which finite-dimensional approximations may perform poorly. Our proof hinges on regularity assumptions, such as smoothness conditions on the probability distribution that represents the seller's belief about the buyers' values, and consists in using an isoperimetric inequality to establish a notion of stability under small perturbations.

Having provided a theoretical justification for the finite-dimensional approximation, we approach the computational problem associated with such approximation. The proposed reformulation is based on interim variables and can still be cast as a linear programming problem. To deal with the exponentially large constraint sets, we resort to the implementation of a cutting-plane method which dynamically incorporates constraints as needed, and thus never deals with all of them at once. We provide an algorithm, which terminates in polynomial time in the problem size, to compute the separation oracle associated with the Border constraints and incentive compatibility constraints. As it is well-known, this implies that the finite-dimensional approximation is solvable in polynomial time (e.g., Grötschel et al. (1981, 1988); Padberg and Rao (1980)).

The computation of convergent approximations also provides an opportunity to detect structural patterns in the solution of finite-dimensional approximations. These patterns can help in formulating well-educated guesses about the nature of simple mechanisms that can approximate the optimal solution to the original infinite-dimensional problem.

The chapter is organized as follows. Section 3.2 introduces the theoretical model and poses the infinite-dimensional problem. In Section 3.3 we discuss the finite-dimensional approximation and establish the approximation results. Section 3.4 describes a cutting-plane algorithm for the finite-dimensional approximation. This section also shows that the separation oracles can be implemented in polynomial time, and finite termination of the overall method is achieved. The implementation and computational results are discussed in Section 3.5. Section 3.6 is devoted to the discussion of a particular example. Proofs and technical results are deferred to the appendices.

3.2 The Model

The seller of a single object faces N potential buyers. The seller can set the object's quality equal to any "grade" $j = 1, ..., J$. For notational convenience we identify the set of buyers and the set of quality grades with their cardinalities, that is $N = \{1, ..., N\}$ and $J = \{1, ..., J\}$. Each buyer $i \in N$ has utility function

$$u_i = \sum_{j \in J} v_j^i \; q_j^i - m_i,$$

where q_j^i denotes the probability that the object of grade j is awarded to buyer i, v_j^i denotes the buyer's willingness to pay for it, and m_i denotes his payment to the seller. Buyer i's type $v^i = \left(v_1^i, \ldots, v_j^i, \ldots, v_J^i\right)$ is obtained as the realization of a random variable, distributed independently of the other buyers' values $v^{-i} := \left(v^1, ..., v^{i-1}, v^{i+1}, .., v^N\right)$, with support $V := \prod_{j \in J} \left[\underline{v}_j, \overline{v}_j\right]$ and probability distribution function F. Buyer i observes the realization of his type v^i privately. Finally, the seller faces a cost c_j to provide the object at the quality level $j = 1, \ldots, J$. In the most general version of the model, c_j may depend on the entire type profile $v \in V$. The model can also be adapted, with the required modifications, to monopsony situations, e.g. procurement contexts, in which a single buyer faces N competing sellers.

The seller's problem is to select a pricing schedule that maximizes her expected profit. Invoking the revelation principle (see Myerson (1979)), we restrict attention to the set of all "direct" selling mechanisms in which the buyers simply report their types, and reporting truthfully is a Bayesian-Nash equilibrium of the corresponding type-reporting game. Formally, a direct mechanism (q, m) consists of an *assignment* rule

$$q : V^N \to \Delta(J \times N),$$

where $\Delta(J \times N)$ denotes the simplex on the set $J \times N$, and *payment rules*

$$m = \left(m^1, ..., m^N\right) : V^N \to \mathbb{R}^N,$$

specifying, for any profile of reported types $\omega := \left(v^1, ..., v^N\right) \in V^N$, the probability $q_j^i(\omega)$ that buyer i is awarded the object of grade j, $(i, j) \in N \times J$, and the payment $m^i(\omega)$ that buyer i makes to the seller. The restriction to deterministic payment rules is without loss of generality since all agents are risk-neutral.

A mechanism (q, m) satisfies:

i) *incentive compatibility,* if "truth-telling" is a Bayesian-Nash equilibrium, i.e.

$$\widehat{U}_i(v, v) \geq \widehat{U}_i(\widehat{v}, v) \qquad \text{for all } \widehat{v}, v \in V,\ i \in N; \tag{IC}$$

and

ii) *individual rationality,* if each buyer has no incentive to decline participation, i.e.

$$\widehat{U}_i(v, v) \geq 0 \qquad \text{for all } v \in V, \quad i \in N; \tag{IR}$$

where $\widehat{U}_i$ denotes buyer i's interim expected payoff function

$$\widehat{U}_i(\widehat{v}, v) \equiv \int_{V^{N-1}} \left[v \cdot q^i\left(\widehat{v}, v^{-i}\right) - m^i\left(\widehat{v}, v^{-i}\right)\right] \prod_{k \in N \setminus \{i\}} dF\left(v^k\right) \quad v, \hat{v} \in V.$$

The seller's problem can be formulated as:

$$
(P_0)\quad \begin{cases} \max\limits_{q,m} \displaystyle\int_{V^N} \sum_{i\in N}\left[m^i(\omega) - \sum_{j\in J} c_j(\omega)\, q^i_j(\omega)\right]\prod_{i\in N} dF\left(v^i\right) \\ \\ \text{subject to} \qquad (IC),\ (IR) \quad \text{and} \\ \\ \qquad (C) \begin{cases} q^i_j(\omega) \geq 0, & \text{for all } (i,j,\omega)\in N\times J\times V^N, \\ \sum\limits_{i\in N}\sum\limits_{j\in J} q^i_j(\omega) \leq 1, & \text{for all } \omega \in V^N. \end{cases} \end{cases}
$$

The linear program (P_0) contains functions with domain $V^N \subset \mathbb{R}^{JN}$. Therefore the size of any discrete approximation of (P_0) grows exponentially with both the number of buyers N and the number of quality levels J. Thus direct computation of solutions to finite approximations of (P_0) can only be done for relatively small grid sizes. An initial significant simplification comes directly from the observation that, since all buyers are *ex ante* identical (they draw their values from the same distribution), we can restrict attention, without loss of generality, to *symmetric* mechanisms.[2] However, even after imposing symmetry, the growth with N and J remains exponential.

We propose a reformulation of (P_0) in which the dimensionality is independent of the number of buyers - in particular, all variables have V as domain instead of V^N. The reformulation relies on *interim* variables (Q,U) obtained by integrating out the types of all but one buyer. The interim probability that a buyer is awarded

[2] A mechanism $\{q^i, m^i; i\in N\}$ is *symmetric* if

$$q^i(v) = q^{i'}(\widetilde{v}) \text{ and } m^i(v) = m^{i'}(\widetilde{v})$$

for all $i, i' \in N$, $v, \widetilde{v} \in V$, such that $v^i = \widetilde{v}^{i'}$, $\widetilde{v}^i = v^{i'}$ and $\widetilde{v}^k = v^k$ $k \neq i, i'$. If (q,m) is optimal and asymmetric, then its "mirror image" (q', m'), obtained by permuting buyers, is also optimal. Since the feasible set is convex, and the objective is linear, the mechanism $\frac{1}{2}(q,m) + \frac{1}{2}(q',m')$ is also optimal, and symmetric. The above argument mimics the steps in Maskin and Riley (1986) footnote 11.

the object of grade $j \in J$ is

$$Q_j\left(v\right) \equiv \int_{V^{N-1}} q_j^i\left(v, v^{-i}\right) \prod_{k \in N \setminus \{i\}} dF\left(v^k\right) \text{ for all } v \in V,$$

and the interim expected utility function of each buyer is defined as

$$U\left(v\right) \equiv \sum_{j \in J} v_j\ Q_j\left(v\right) - \int_{V^{N-1}} m^i\left(v, v^{-i}\right) \prod_{k \in N \setminus \{i\}} dF\left(v^k\right), \quad \text{for all } v \in V.$$

By appealing to an extension of the results in Border Border (1991) (see Lemma B.1), we can replace the constraints in (C) with the following "Border constraints" written only with the interim probabilities.

$$(B) \begin{cases} Q_j\left(v\right) & \geq & 0, & \text{for all } j \in J,\ v \in V \\ N \int_A \sum\limits_{j \in J} Q_j\left(v\right) dF\left(v\right) & \leq & 1 - \left(\int_{V \setminus A} dF\left(v\right)\right)^N, & \text{for all } A \subset V. \end{cases}$$

Moreover, we can rewrite (IC) and (IR) with interim variables as follows

$$U(v) - U(\hat{v}) \quad \geq \quad \langle Q(\hat{v}), v - \hat{v} \rangle \quad \text{for all } (v, \hat{v}) \in V \times V, \tag{IIC}$$

and

$$U(v) \geq 0 \quad \text{for all } v \in V; \tag{IIR}$$

and, by appealing to the following well-known characterization lemma (Manelli and Vincent, 2007), we eliminate the most of the (IIR) constraints.[3]

Lemma 3.1. *If (Q, U) satisfies (IIC), then U is convex, and $Q\left(v\right) \in \partial U\left(v\right)$ for all $v \in V$. Conversely, if U is convex, then $\left(Q, U\right)$, where $Q(v) \in \partial U(v)$ for all $v \in V$, satisfies (IIC).*

[3] Following standard notation in convex analysis Rockafellar (1970), we denote the subdifferential of a convex function $U : V \to \mathbb{R}$ at $v \in V$ by $\partial U(v) = \{s\ : U(\hat{v}) \geq U(v) + \langle s, \hat{v} - v \rangle \text{ for all } \hat{v} \in \mathbb{R}^n\}$.

Lemma 3.1 immediately implies that (IIR) holds if and only if $U(\underline{v}) \geq 0$, where $\underline{v} := \left(\underline{v}_1, ..., \underline{v}_j, ..., \underline{v}_J\right)$. Clearly, in any mechanism that maximizes the seller's expected revenue we must have $U(\underline{v}) = 0$.

Thus the seller's problem of finding a price schedule that maximizes her expected profit subject to incentive compatibility and individual rationality can be stated as

$$(P_*) \begin{cases} OPT_* = \max_{Q,U} \int_V \left[\sum_{j\in J} [v_j - c_j(v)]\, Q_j(v) - U(v)\right] dF(v) \\ \\ \text{subject to } (B),\ (IIC) \text{ and } U(\underline{v}) = 0. \end{cases}$$

As a final remark, note that the formulation (P_*) eliminates the dependence of the number of variables on N, since it is searching for variables that are functions on $V \subset \mathbb{R}^J$. However, this simplification comes at the cost of having to deal with an exponentially large class of Border constraints (B). This structure lends itself to the use of cutting-plane algorithms, which consider only a subset of the constraints at any iteration. By dynamically adjusting the subset of constraints, it is possible to efficiently solve the original problem (i.e. in polynomial time in the dimension of the problem).

3.3 Finite-Dimensional Approximation

In this section we study how to approximate the infinite-dimensional problem of interest (P_*) by a sequence of finite-dimensional problems. In particular, we provide an extension of any solution to a discretized problem for which we can guarantee (a rate of) convergence to the solutions of (P_*). Throughout the chapter we impose the following regularity conditions.

Assumption 3.1. *The data of the problem (P_*) satisfy: (i) the type space V is a compact subset of $\mathbb{R}^J_{++}$; (ii) the probability distribution F has a twice continuously* *differentiable density function f which is bounded away from zero; (iii) the mapping $w : V \to \mathbb{R}^J_+$, defined as $w_j(v) \equiv v_j - c_j(v)$, is twice continuously differentiable;*

In order to approximate an optimal solution of (P_*), we begin by discretizing the type space V. Let T denote a positive integer that will control the precision of the discretization. For each $j \in J$, let $V_T(j)$ denote the discretization of the interval $[\underline{v}_j, \overline{v}_j]$

$$V_T(j) = \{\underline{v}_j, \underline{v}_j + \epsilon, \underline{v}_j + 2\epsilon, \ldots, \overline{v}_j\}$$

where $\epsilon = \min_{j\in J} \left\{ \frac{\overline{v}_j - \underline{v}_j}{T} \right\}$. Our discretized version of V, parameterized by T, is defined as $V_T := \prod_{j\in J} V_T(j)$. The grid V_T is a set with $O(T^J)$ elements which defines a net over V such that

$$\max_{v\in V} \operatorname{dist}(v, V_T) \;\; \leq \;\; \epsilon\,\sqrt{J}.$$

Based on the probability density function f we can define a probability distribution function on V_T by setting $\hat{f}(v) = \frac{f(v)}{\sum_{t\in V_T} f(t)}$. Note that $\hat{f}(v)$ approximates the measure on the hypercube $\prod_{j\in J}[v_j, v_j + \epsilon]$.

For each $T > 0$, by replacing V with the grid V_T, we obtain the following (sequence of) finite-dimensional approximation for (P_*).

$$(P_T)\left\{\begin{array}{ll} OPT_T = \max\limits_{Q,U} & N \sum\limits_{v\in V_T} \left(\sum\limits_{j\in J} w_j(v)\; Q_j\,(v) - U\,(v) \right) \hat{f}(v) \\ & U(v) - U(\hat{v}) \geq \langle Q(\hat{v}), v - \hat{v}\rangle \quad \text{for all } (v,\hat{v}) \in V_T \times V_T, \\ & U(\underline{v}) = 0, \\ & N \sum\limits_{v\in A} \left[\sum\limits_{j\in J} Q_j\,(v) \right] \hat{f}(v) \leq 1 - \left(\sum\limits_{v\in V_T\setminus A} \hat{f}(v) \right)^N, \quad \text{for all } A \subset V_T, \\ & Q_j(v) \geq 0, \quad \sum\limits_{j\in J} Q_j(v) \leq 1, \quad 0 \leq U(v) \leq \overline{U}, \quad \text{for all } \; v \in V_T. \end{array}\right.$$

where $\overline{U}$ is an upper bound on U that can be derived from the fact that the set V is compact and Q is bounded from above. The "additional" upper bounds on Q are already implied by the Border constraints; they are included explicitly to highlight the fact that any feasible solution is bounded. We denote by (Q^T, U^T) a optimal solution of (P_T).

Since our "target" program (P_*) is infinite-dimensional, and allows discontinuous functions within its feasible set, the following issues must be confronted. First, in the present setting, the discretization affects not only the computation of the objective function value but also the constraint formulation. Therefore intuitive extensions of the solution to a discretized problem might not be near feasible for the continuous problem (P_*). Second, the presence of discontinuous functions Q_j in the feasible set introduces challenges to approximations based on grids. Intuitively, in order for an approximation scheme to be successful, the problem of interest must be stable under small perturbations.

We will work with the following approximation notions. Given a pair of mappings (Q, U), let $\delta_*(Q, U)$, respectively $\delta_T(Q, U)$, denote the supremum over all constraint violations in (P_*), respectively in (P_T), by (Q, U). Analogously, let $OPT_*(Q, U)$, respectively $OPT_T(Q, U)$, the objective function value obtained by (Q, U) in (P_*), respectively in (P_T). Note that (Q, U) is feasible for (P_*) only if $\delta_*(Q, U) \leq 0$.

The following notion of asymptotic optimality is standard: all violations converge uniformly to zero, and optimality is achieved in the limit.

Definition 3.1. *A sequence of mappings (Q^T, U^T) defined over V is said to be asymptotically optimal for (P_*), if $\lim_{T\to\infty} \delta_*(Q^T, U^T) = 0$ and $\lim_{T\to\infty} OPT_*(Q^T, U^T) = OPT_*$.*

Next, we propose an extension of the finite-dimensional solutions (Q^T, U^T) which has useful properties. Our first result is concerned with the feasibility guarantees of the extension.

Theorem 3.1. *Let (Q^T, U^T) be a solution to (P_T) associated with the grid V_T induced by $\epsilon = O(1/T)$. Consider the extensions $(\widetilde{Q}^T, \widetilde{U}^T)$ of (Q^T, U^T) to the set V defined as*

$$\begin{aligned} \widetilde{U}^T(v) &\equiv \max_{\hat{v} \in V_T} U^T(\hat{v}) + \left\langle Q^T(\hat{v}), v - \hat{v} \right\rangle, \quad \textit{and} \\ \widetilde{Q}_j^T(v) &\equiv Q_j^T(\hat{v}), \ j \in J, \ \hat{v} \in V_T, \ \hat{v} \le v < \hat{v} + \epsilon e. \end{aligned} \tag{3.1}$$

Under Assumption 3.1, we have:

(i) $\widetilde{U}^T(v)$ is convex, (ii) $\widetilde{Q}^T(v) \in \partial_{2\epsilon} \widetilde{U}^T(v), \forall v \in V$ and (iii) $\delta_*(\widetilde{Q}^T, \widetilde{U}^T) \le O(1/T)$.

Theorem 3.1 yields non-asymptotic bounds in the maximum violation of the extension. In particular it establishes that the extended $\widetilde{Q}^T(v)$ is a 2ϵ-subgradient of $\widetilde{U}^T(v)$ for all $v \in V$, namely for any $v' \in V$ we have

$$\widetilde{U}^T(v') \ge \widetilde{U}^T(v) + \left\langle \widetilde{Q}^T(v), v' - v \right\rangle - 2\epsilon.$$

Recall that the discretization also affects the Border constraints (B). Theorem 3.1 also establishes bounds for violation over all measurable subsets of V. However, it will be possible to satisfy all Border constraints of (P_*) by rescaling a feasible solution of (P_T), or vice-versa. A key step in the proof of this results is provided in the following technical, which establishes an isoperimetric inequality relating violations of Border constraints and the measure of the associated subset of the type space.

Lemma 3.2. *Consider an arbitrary probability measure F (possibly discrete) on V, and any measurable mapping $Q : V \to \mathbb{R}^J$ such that $0 \le \sum_{j \in J} Q_j(v) \le 1$. Then, for any measurable subset $A \subseteq V$, and any $\eta \ge 0$, the inequality*

$$N \int_A \sum_{j=1}^J Q_j(v) dF(v) \quad \ge \quad 1 - \left(\int_{A^c} dF(v) \right)^N + \eta,$$

implies that $\int_A dF(v) \geq \sqrt{2\eta}/N$.

Lemma 3.2 is instrumental in recovering feasible solutions for (P_*) or (P_T) based on near-feasible solutions that violate only Border constraints. Essentially it asserts that if a Border constraint is violated, the probability measure of the associated subset of the type space must be relatively large. In turn this ensures that the right-hand-side of the constraint is large (since the probability measure of the complement cannot be close to one). Corollary 3.1 below shows how all Border constraints can be satisfied without substantially reducing the corresponding objective function value.

Corollary 3.1. *Assume that a pair of mappings* (Q, U) *satisfies* $\delta(Q, U) \leq \frac{1}{2}$*, where* $\delta = \delta_*$ *or* $\delta = \delta_T$*. Then, the rescaled pair*

$$(Q^r, U^r) \equiv \left(1 - \frac{\sqrt{\delta(Q,U)}}{\sqrt{2} - \sqrt{\delta(Q,U)}}\right) \cdot (Q, U)$$

satisfies all Border constraints and $\delta(Q^r, U^r) \leq \delta(Q,U)\left(1 - \frac{\sqrt{\delta(Q,U)}}{\sqrt{2} - \sqrt{\delta(Q,U)}}\right)$.

By applying the previous corollary to the restriction of the optimal solution (Q^*, U^*) to the grid V_T and (P_T) we are able to bound the (unknown) optimal value of (P_*).

Theorem 3.2. *Let* (Q^*, U^*) *denote the optimal solution for* (P_*)*. For any finite* T *let* $\delta_T(Q^*, U^*)$ *denote the maximum violation of* (Q^*, U^*) *for* (P_T)*. Under Assumption 3.1 we have that* $\delta_T(Q^*, U^*) = O(1/T)$ *and*

$$OPT_* \leq OPT_T \left(1 - \frac{\sqrt{O(1/T)}}{\sqrt{2} - \sqrt{O(1/T)}}\right)^{-1} + O\left(\frac{1}{T}\right).$$

Theorem 3.2 provides a numerical bound for the true optimal value of (P_*) which can be used to evaluate the quality of any feasible solution for (P_*). Finally, we close this section by establishing its main result.

Corollary 3.2. *Under Assumption 3.1 the sequence of extensions* $(\widetilde{Q}^T, \widetilde{U}^T)$ *defined as in (3.1) is asymptotically optimal for* (P_*)*.*

The proof of the above corollary simply combines Theorems 3.1 and 3.2.

3.4 Algorithmic Structure

Having established the validity of the finite-dimensional approximations (3.1), we proceed to address the computational issues that arise when we solve (P_T) numerically.

To have a sense of how quickly the discrete program grows with T, note that the dimensionality of the variables, cardinality of (IIC) constraints and cardinality of Border constraints are respectively $O(JT^J)$, $O(T^{2J})$ and $O(2^{T^J})$. In order to obtain numerical solutions for relatively large values of T the particular structure of the problem (P_T) must be exploited. In particular, it is computationally intractable to simply enumerate all Border constraints for large values of T. We start by showing an equivalent characterization for the Border constraints.

Lemma 3.3. *Consider the Border constraints in* (P_*) *or* (P_T)*. For any given* $Q : V \to \mathbb{R}^J$ *the sets* $A \subset V$ *can be restricted to the form of*

$$E_\alpha(Q) = \left\{ v \in V : \sum_{j \in J} Q_j(v) \geq \alpha \right\} \quad \text{for all } \alpha \geq 0.$$

The version of Lemma 3.3 where J is a singleton was first established by Border (1991). In this chapter we fully exploit the computational consequences of this characterization. The following lemma plays a crucial role in our method. It establishes that both separation oracles associated with (IIC) and (B) can be implemented in polynomial time in T^J. That is, given a candidate solution (Q, U), it is possible to efficiently verify either that (Q, U) satisfies all (IC) and (B) constraints, or exhibit one constraint that is violated by (Q, U).

Lemma 3.4. *The separation oracle for the Border constraints (B) can be implemented in $O(JT^J \ln T)$ operations. The separation oracle for the (IC) constraints can be implemented in $O(JT^{2J})$ operations.*

The usefulness of the new representation of the Border constraints derived in (3.3) is notable. It enables an implementation of the separation oracle for (B) to be more efficient than the separation oracle for (IC).

In light of Lemma 3.4, the cutting-plane method can be efficiently implemented to solve (P_T) for any T as follows:

Algorithm 3.1. *Let S^k denote the set of constraints at iteration k.*

Step 1: *Let $k = 1$, $S^1 = \emptyset$, $\overline{OPT} = \infty$.*

Step 2: *Solve the linear program associated with S^k.*
Let (Q^k, U^k) denote a optimal solution and
let OPT^k denote the optimal value.

Step 3: *Select a subset $I^k \subset S^k$ of inactive (IC) and (B)*
constraints for (Q^k, U^k).

Step 4: *Solve the separation oracle for (IC) and (B).*
Let A^k denote a subset of the violated inequalities to be added.

Step 5: *If $OPT^k < \overline{OPT}$, set $S^{k+1} \leftarrow (S^k \backslash I^k) \cup A^k$ and $\overline{OPT} \leftarrow OPT^k$.*
Else set $S^{k+1} \leftarrow S^k \cup A^k$.

Step 6: *If $A^k = \emptyset$, stop.*
Else set $k \leftarrow k + 1$ and goto Step 2.

Algorithm 3.1 is a cutting-plane algorithm works with a subset of (IIC) and (B) constraints at each iteration. Step 5 is there to prevent the dimensionality of the program from growing unnecessarily, by dropping inactive constraints whenever the objective value decreases. The following lemma establishes that the algorithm terminates in a finite number of iterations.

Theorem 3.3. *Algorithm 3.1 terminates in finite steps with the optimal solution.*

Finally, invoking Grötschel et al. (1981, 1988) and Padberg and Rao (1980), since both separations oracles can be solved in polynomial time with respect to the dimension of (P_T), we have that the linear programming problem (P_T) is solvable in polynomial time (for instance, via the ellipsoid method).

3.5 Implementation and Computational Results

In this section we briefly discuss some aspects of our implementation and display our computational results in Table 3.2.

To properly solve any large-scale linear programming problem, considerable effort into the implementation is needed. Throughout the course of the implementation of Algorithm 3.1 we relied on many existent linear programming solvers including SeDuMi, SDPT3, LINSOL, and CPLEX. During an initial phase of the implementation we used MatLAB as the working environment for preliminary testing, and later we built a C++ environment to properly work with CPLEX for better memory management.

Table 3.2 summarizes the results of the implementation on a variety of instances for different configurations of the grid (T), buyers (N), and quality levels (J). As a by-product of the implementation, by solving instances without considering incentive compatibility constrains, we recover solutions for the "First Best Problem," whose optimal value is always above of the seller's maximization problem (P_T). In all instances the optimal solution was obtained in all instances. This illustrates the finite termination property established in Theorem 3.3.

The use of general cost functions and probability distributions is allowed by the implementation. Regarding the linear programming solver, as typical in cutting-plane methods, warm-start the solver with the previous solution can lead to signif-

Table 3.2: Computational results for various model specifications.

				Active		Total	
J	N	T	OPT VAL	(IC)	(B)	(IC)	(B)
Seller's Problem							
2	2	5	6.051008	56	9	$(25)^4$	2^{25}
2	2	10	5.946440	257	28	$(100)^4$	2^{100}
2	2	20*	5.896710	2355	118	$(400)^4$	2^{400}
2	2	30*	5.880743	2802	920	$(900)^4$	2^{900}
2	5	5	6.734230	54	9	$(25)^4$	2^{25}
2	5	10	6.605535	246	18	$(100)^4$	2^{100}
2	5	20*	6.5335155	1365	398	$(400)^4$	2^{400}
2	10	5	6.99129	40	8	$(25)^4$	2^{25}
2	10	10	6.88799	207	17	$(100)^4$	2^{100}
2	10	20*	6.8185167	1159	512	$(400)^4$	2^{400}
3	2	5	4.263083	641	139	$(125)^6$	2^{125}
3	5	5	7.82672	793	171	$(125)^6$	2^{125}
3	10	5	8.16429	549	134	$(125)^6$	2^{125}
First Best							
2	2	5	6.540664	–	25	–	2^{25}
2	2	10	6.487975	–	100	–	2^{100}
2	2	20	6.452313	–	400	–	2^{400}
2	5	5	6.901566	–	24	–	2^{25}
2	5	10	6.833399	–	86	–	2^{100}
2	5	20*	6.800695	–	2917	–	2^{100}
2	10	5	7.050792	–	20	–	2^{25}
2	10	10	6.991210	–	62	–	2^{100}
2	10	20*	6.956887	–	200	–	2^{100}
3	2	5	8.468104	–	181	–	2^{125}
3	5	5	9.035501	–	145	–	2^{125}
3	10	5	9.450985	–	135	–	2^{125}

The seller's maximization problem is the model considered in Section 3.3. The first-best problem ignores the incentive compatibility constraints. The instances marked with "*" were solved using a simplex method, while the others were solved using an IPM.

icant improvements in running time. Although interior point methods (IPMs) were usually faster than simplex-type methods to solve each LP from the scratch, simplex-type were able to use warm-starts and were more efficient in the larger instances that we considered.

Finally, it is worth pointing out that the number of active Border constraints is substantially smaller than the number of active incentive compatibility constraints, in all tested instances.

3.6 Insights and Discussion of a particular Example

The computation of discrete approximations also provides an opportunity to detect patterns in the qualitative nature of the optimal solutions that can help in formulating educated guesses about the functional form of the optimal solutions to the infinite-dimensional problem (P_*) posed by the theory.

In this section we focus on a simple example involving only two quality levels, where the buyers' types are drawn from a uniform distribution. Formally, we have $J = 2$, F uniform on a rectangular support $[\underline{v}_1, \overline{v}_1] \times [\underline{v}_2, \overline{v}_2]$. Figure 3.1 shows the graph of the buyer surplus U in (3.1), for a grid with $T = 30$.

A reasonable visual "approximation" for the function U is the maximum of two one-dimensional, strictly convex functions: one that depends only on v_1 and the other that depends only on v_2. Moreover, these two functions meet along a linear path, on which U is nondifferentiable.

The solutions obtained by considering various values for the cost vector c, and various grid sizes, exhibited patterns that have prompted us to formulate the following simple mechanism as a candidate for a good approximation of the optimal solution in the infinite-dimensional problem (P_*). Consider the following class of

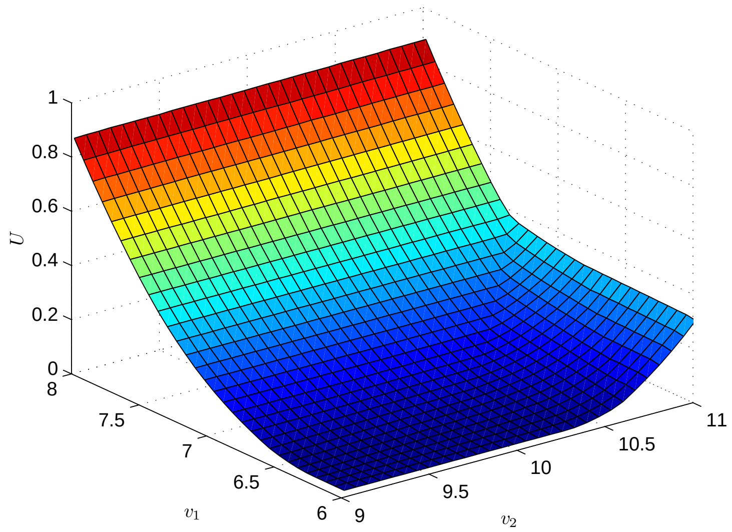


FIGURE 3.1: Buyer surplus function U for $T = 30$.

mechanisms, parameterized by the prescribed "add-on price" p:

$$Q_1^p(v_1, v_2) = F^{N-1}(v_1, v_1 + p)\ 1\{v_1 > v_2 - p\}$$

$$Q_2^p(v_1, v_2) = F^{N-1}(v_2 - p, v_2)\ 1\{v_1 < v_2 - p\}$$

and

$$U^p(v_1, v_2; p) = \max\left\{\int_{\underline{v}_1}^{v_1} Q_1^p(t, v_2)\, dt,\ \int_{\underline{v}_2}^{v_2} Q_2^p(v_1, t)\, dt\right\}.$$

For any p, the mechanism (Q^p, U^p) is feasible for the infinite-dimensional program (P_*), because it can be implemented by the following auction named "*exclusive buyer mechanism*", for the case with two quality levels (see Brusco et al. (2011)). The buyers compete in a second-price or ascending-bid auction, possibly augmented with a reserve price r, for the right to be the only buyer who gets to choose between buying the object of lower quality (grade 1) at no additional cost, or the object of higher quality (grade 2) for an additional payment of p.[4] Since the value for buyer

[4] For the case with $J > 2$ quality levels, the winner of the auction selects a price-quality pair from

$i \in N$ of winning the auction is $\beta_i \equiv \max\{v_1^i, v_2^i - p\}$, it is a dominant strategy for buyer i to bid β_i at the auction. Thus, in equilibrium, buyer i wins the auction only if $\beta_i \geq \max\{r, \max\{\beta_j; j \in N \backslash \{i\}\}\}$, is assigned the low-quality and high-quality object with probabilities $Q_1^p(v_1^i, v_2^i)$ and $Q_2^p(v_1^i, v_2^i)$, respectively, and his expected utility is $U^p(v_1, v_2; p)$.

Let p_* denote the value of p which maximizes the seller's expected profit within the class of exclusive buyers mechanisms, i.e. across all values of $p \in \mathbb{R}$. Table 3.3 shows how the value of the objective function in (P_*) evaluated at (Q^{p_*}, U^{p_*}) (last row) compares with the values obtained by the objective function in (3.1) for various grid sizes, and F uniform on $[\underline{v}_1, \overline{v}_1] \times [\underline{v}_2, \overline{v}_2] = [6, 8] \times [9, 11]$, $c_1 = 0.9$ and $c_2 = 5$.

Table 3.3: Optimal values of (P_T) and the objective function value of the exclusive buyer mechanism.

J	N	T	OPT VAL
2	2	5	6.051008
2	2	10	5.946440
2	2	20	5.896710
2	2	30	5.880743
2	2	1000*	5.838323

*obtained by numerical integration with $T = 1000$.

As Table 3.3 shows, the exclusive buyer mechanism performs quite well relative to the numerical optimal solutions, at least for the case in which F is uniform. Moreover, as Figure 3.2 illustrates, the difference between the U function shown in Figure 3.1 and the buyer surplus function in the optimal exclusive buyer mechanism, for a grid of size $T = 30$, is relatively small. (The largest deviations occur near the "manifold" of non-differentiable points of U.)

a menu of up to J choices.

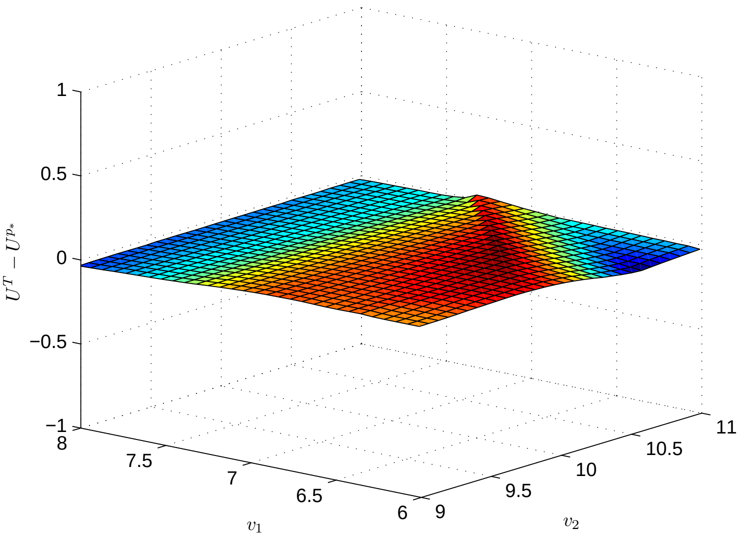


FIGURE 3.2: The difference of the buyer surplus functions between (P_T), for $T = 30$, and the optimal exclusive buyer mechanism.

It is interesting to contrast the nature of the exclusive buyer mechanism with the qualitative features of the optimal mechanisms in the single buyer models, studied in Armstrong (1996), Rochet and Choné (1998), and Manelli and Vincent (2007). A main result in Armstrong (1996) is that, in any optimal mechanism, it is optimal not to sell with positive probability. In contrast, in the optimal exclusive buyer mechanism with no reserve price, which performs quite well relative to the numerical optimal mechanisms in our numerical simulations, the object is always sold. Rochet and Choné (1998) establish that

> "bunching is robust in these multidimensional screening problems, even with very regular distributions of types. This comes from a strong conflict between participation constraints and second order incentive compatibility conditions." (Rochet and Choné, 1998, p. 783)

Manelli and Vincent (2007) reinforce the point that, in one buyer settings, optimal mechanisms are generically quite different in nature from the simple "take-it-or-leave-it" mechanisms, which are always optimal in the one-dimensional case:

> "We find that the set of extreme points contains, in addition to price-posting, many "novel" mechanisms. In particular, extreme points need not be simple functions (Example 2), and even when they are, they may randomly assign objects to consumers (Examples 1 and 3). In contrast to the one-good case, the form of the optimal mechanism is determined by the prior distribution of buyer valuations." (Manelli and Vincent, 2007, p. 154)

In contrast, our analysis suggests that the exclusive buyer mechanism with an optimally chosen 'add-on' price can perform quite well, and shares many of its defining features with its one-dimensional counterparts, i.e. Myerson's optimal auctions: there is no 'bunching', and is defined by simple rules that do not depend on the prior distribution of buyer valuations.

4

Decentralized Resource Allocation to Control an Epidemic

4.1 Motivation

A common strategy for containing the spread of an epidemic consists in reducing the number of susceptibles in a population so as to bring the effective reproduction ratio $\mathcal{R}$, (typically equal to the average number of infections caused by an infectious individual after countermeasures have been applied, also referred to as the "control reproductive ratio") below one (Smith, 1964; May and Anderson, 1984; Anderson and May, 1991). For instance, immunization through vaccinations has proven to be an effective technique for controlling many infectious diseases (see, for instance, Fenner et al. (1988)). In most situations, the amount of available resources to reduce the number of susceptibles is limited. This typically leads to an optimization problem where a central planner needs to allocate resources among different groups having various transmission rates in order to minimize the overall impact of the epidemic (Brandeau et al., 2003; Brandeau, 2004).

However, a central planner may not always exist. For example, the emerging threat of an influenza pandemic has revealed situations where available resources are divided among independent decision makers. Regions where transmissions of the H5N1 influenza A virus to humans have occurred thus far, such as South East Asia, do not maintain large inventories of antivirals or vaccines. Western Europe, North America and Japan, on the other hand, possess most of the world's stock of these medicines (Anonymous, 2005). Vaccines might need to be diverted to other countries in order to contain an epidemic (Monto, 2005), but no international organization exists to enforce a worldwide allocation rule. Rather, decision makers might distribute their limited resources so as to protect their own populations, possibly at the expense of others. Countries that have decentralized these allocation decisions, such as the United States, face a similar situation even within their borders.

This study examines how two selfish countries (or independent geographical areas) would allocate their resources selfishly at the onset of the epidemic when each seeks to protect its own population even at the expense of others. We model this situation as a game between countries where players strategically allocate their resources in order to minimize the total number of infected individuals (i.e. the final size of the epidemic) in their own population. We show in this framework that selfish countries always allocate their resources so as to bring the effective reproduction ratio below one, as long as there are enough resources in the system for a hypothetical central planner to do so. In other words, countries that act in their own self-interest avoid the outbreak of a pandemic when this is possible, as a central planner would do. Nonetheless, even when the effective reproduction ratio is below one, individuals in both populations are infected. In this case, the equilibrium allocation decisions in our model differ in general from the centralized optimal solution that would minimize the total number of removals in both countries. This means that at least one of the two countries will suffer from the decision of the other.

We describe the epidemic dynamics using a deterministic dynamic population SIR model, a heterogeneous version of Ball's modelBall (1991).[1] The total population is divided into two countries, in each of which the population mixes homogenously. A fraction of infectives also moves from one country to the other, infecting the new population. Each country possesses a given amount of resources. We assume, without loss of generality, that one unit of resource can immunize one susceptible. Allocation decisions are made at the beginning of the time horizon and their effects last until the end of the pandemic. Countries may not be symmetrical, i.e., they may have different population, transmission rates, and may hold different amounts of resources.

Given an initial number of infectives, a country needs to decide how much of its resources to use for its population and how much to allocate to the other country in order to limit infections from abroad. The objective of the country is to minimize the total number of removals from its population until the end of the epidemic. This situation gives rise to a game, and our goal is to characterize the Pareto Nash Equilibrium, when it exists. An important question is how the game equilibrium compares with the optimal decision that a central planner would make in order to minimize the total number of removals in the whole population. In particular, when the total amount of pooled resources is large enough, the central planner immunizes (possibly different) fractions of each population so as to bring the effective reproduction ratio below one and therefore avoid a major outbreak (Hill and Longini, 2003). We are interested in the conditions under which selfish countries achieve a similar result.

Given the allocation decisions and the initial number of infectives, the final size of the epidemic in each country is given by the solution of a system of highly non-linear

[1] Ball pointed out that a model incorporating movement of both susceptibles and infectives is preferred when describing human epidemics. This does not create any problem for our later results due to approximation. In fact, the Ball's model has also been adopted by other researchers (see Klein et al. (2007) for example) to model human epidemics.

equations. This makes the analysis of the game intractable. Hence, we consider first the case where the number of initial infectives is very small, a realistic assumption. More precisely, we study the system of equations as the initial number of infectives approaches zero. The solution at the limit corresponds to the players' response curves of the resource allocation game. In this framework, we show that a Pareto Nash Equilibrium exists, for which the effective reproduction ratio is below one. This result suggests that selfish decision makers are able to avoid major outbreaks whenever possible.

When an outbreak is avoidable, because we analyze the system at the limit when the number of initial infectives is null, the final size is also null at the equilibrium. Nonetheless, individuals are infected even when a major outbreak does not occur. To study this situation in more detail, we focus on the case where the effective reproduction ratio is below one, which is achievable in the Pareto Nash Equilibrium. In this setting, the number of infectives decreases over time and we make the classical approximation that the total number of susceptibles in a population remains constant over time. This approach can be thought of as a first order approximation of the original model. We can then derive analytical expressions for the final size of the epidemic, which depends on the initial number of infectives. We show the existence of a unique Pareto Nash Equilibrium that can be fully characterized. Finally, we solve the central planner's optimization problem, and show that in general it does not match the equilibrium of the decentralized game.

This chapter is one of the few studies where the resource allocation decision to contain the spread of an epidemic is decentralized. To our knowledge, only Sun, Yang, and de Véricourt (2009) have analyzed a similar problem. Their focus, however, is on containing a pandemic at its onset, while we are concerned with the final size of the epidemic. Besides this study, very few papers have formulated game theory problems in the context of epidemic dynamics. Notable exceptions are Bauch et al.

(2003) and Bauch and Earn (2004) where they studied the impact of individuals' decision to vaccinate according to self-interest on the eradication of a disease.

The chapter is organized as follows. In Section 4.2, we formally introduce our model and the notions of basic reproduction ratio and effective reproduction ratio. This allows us to derive necessary and sufficient conditions for allocation decisions for a major outbreak to occur. We consider in Section 4.3 the resource allocation game assuming the number of initial infectives is close to zero. In Section 4.4, we relax this assumption by considering that the size of the susceptible population remains constant over time. We conclude in Section 4.5.

4.2 Epidemic Dynamics and Reproduction Ratios

We consider a world with two countries, $i = 1, 2$, each with an original population size $\bar{n}_i$ when the epidemic first starts at time $t = 0$. For each country i at time $t \in [0, \infty)$, let $x_i(t), y_i(t)$ and $z_i(t)$ denote the total number of susceptibles, the total number of infectives, and the total number of removals, respectively. In particular, $z_i(\infty)$ represents the total number of final removals (henceforth referred to as the "final size") when the influenza deceases, and $z_i(0) = 0$. Further, let ϵ_i represent the number of initial infectives in country i such that $y_i(0) = \epsilon_i$. Let $\varepsilon = (\epsilon_1, \epsilon_2)'$ and assume that $0 \leq \epsilon_i \ll \bar{n}_i$ with $\epsilon_1 > 0$ or $\epsilon_2 > 0$, or both. Let β_i and γ_i denote the pairwise infectious contact rate and removal rate of infectives, respectively, within country i. Further let B denote the diagonal matrix diag$\{\beta_1, \beta_2\}$ and C the diagonal matrix diag$\{\gamma_1, \gamma_2\}$.

We model the temporal-spatial development of the epidemic following Ball (1991). Define q_{ij} the immigration rate of an infective from country i moving to country j, with $q_{ij} \geq 0$. Also denote $q_{ii} = -q_{ij}$ for $j \neq i$ and $Q = (q_{ij})$ the corresponding transition-rate matrix. The following system of ordinary differential equations (Ball,

1991) models the dynamics of the epidemic without any intervention:

$$\frac{dx_i}{dt} = -\beta_i x_i y_i \tag{4.1}$$

$$\frac{dy_i}{dt} = \beta_i x_i y_i + \sum_{j=1}^{2} q_{ji} y_j - \gamma_i y_i, \tag{4.2}$$

$$\frac{dz_i}{dt} = \gamma_i y_i \tag{4.3}$$

The total endowment of resources that country i possesses to protect susceptibles is equal to $\bar{u}_i$. Resources include antiviral drugs, vaccines, masks, etc. We calibrate the units such that one unit of resource fully protects one susceptible anywhere in the world from being infected by the pathogen. At time zero, country i allocates $u_i \in [0, \bar{u}_i]$ to country $j \neq i$ and keep $\bar{u}_i - u_i$ for its own population. Given each country's allocation decision we can define $a \equiv u_1 - u_2$, such that $a \in [-\bar{u}_2, \bar{u}_1]$. The quantity '$a$' corresponds to the net resources which are transferred from country 1 to country 2. Specifically,

$$a \begin{cases} > 0 & : \quad a \text{ units of resources transferred from country 1 to 2;} \\ = 0 & : \quad \text{no transfer of resources occurs;} \\ < 0 & : \quad -a \text{ units of resources transferred from country 2 to 1.} \end{cases} \tag{4.4}$$

(Similarly, it follows that $u_1 = \bar{u}_1 - a \geq 0$ and $u_2 = \bar{u}_2 + a \geq 0$.)

We also assume that the total resources in the world are not enough to fully remove all susceptibles in either of the countries. Otherwise, the game is degenerate when all susceptibles in one country are removed. Therefore, the initial number of susceptibles in country i is $x_i(0) = \bar{n}_i - u_i$, and is guaranteed to be positive for all feasible u_i.

For later ease of notation, we denote $X_0(a) = \text{diag}\{x_1(0), x_2(0)\}$. We also denote $n_1 = \bar{n}_1 - \bar{u}_1$ and $n_2 = \bar{n}_2 - \bar{u}_2$. Therefore $x_1(0) = n_1 + a$ and $x_2(0) = n_2 - a$.

Given the dynamics of an epidemic, a major outbreak occurs when the corresponding *basic reproduction ratio* $\mathcal{R}_0$ is above one. For heterogeneous populations, this ratio is defined as the radius spectrum of the next generation matrix (Diekmann and Heesterbeek, 2000), which depends on the initial number of susceptibles. In our framework, however, resource allocation 'a' decreases the number of susceptibles and therefore changes the conditions for a major outbreak. Hence, we refer to $\mathcal{R}(a)$ as the *effective reproduction ratio* given the allocation 'a'. When there is no transfer of resources, $\mathcal{R}(0) = \mathcal{R}_0$. One of the main questions this chapter seeks to answer is whether the allocation decisions countries make when they act in their own self interest would lead to a major outbreak ($\mathcal{R}(a) \geq 1$ at the equilibrium).

For our model, Ball (1991) discusses the conditions that give rise to a major outbreak for the symmetrical case, when $\beta_i = \beta$, $\gamma_i = \gamma$ and $n_i = n$ for all i. The basic reproduction ratio is then equal to $\mathcal{R}_0 = \beta n/\gamma$. In order to determine $\mathcal{R}(a)$ for asymmetrical countries, we first consider a correspondence between the deterministic model and the stochastic model following Ball (1991). Let p_{ij} be the probability that an infective from group i makes contact with an individual in group j. After a unit of time, one infective from group i will cause $p_{ij}\beta_j x_j(0)$ infectives in group j, while the removal rate of the infective in group i is γ_i. According to the corresponding birth and death processes, one infective from group i will produce $\gamma_i^{-1} p_{ij}\beta_j x_j(0)$ secondary infectious cases in expectation in group j. Therefore, the effective next generation matrix is given by

$$C^{-1}PBX_0(a),$$

where $P = (p_{ij})$. According to (Ball, 1991, Theorem 4.1) we have

$$P = C(C - Q)^{-1}.$$

Hence, the effective next generation matrix can be further reduced to

$$(C - Q)^{-1}BX_0(a).$$

Definition 4.1. *Given a strategy 'a' at the onset of the epidemic, the* effective reproduction ratio $\mathcal{R}(a)$ *is defined as the radius spectrum of the* effective next generation matrix

$$(C - Q)^{-1} B X_0(a).$$

Alternatively, we may identify a set $\mathcal{A}$ of allocations such that $\mathcal{R}(a) \leq 1$ if and only if $a \in \mathcal{A}$. Next, we first provide the definition of such a set $\mathcal{A}$ and then establish the equivalence.

Definition 4.2. *Define the* feasible allocation set $\mathcal{A}(\bar{u}_1, \bar{u}_2) \subset [-\bar{u}_2, \bar{u}_1]$ *to be the set of net transfers 'a' which satisfies the following three conditions:*

$$\gamma_1 + q_{12} - \beta_1(n_1 + a) \geq 0 \tag{4.5}$$

$$\gamma_2 + q_{21} - \beta_2(n_2 - a) \geq 0 \tag{4.6}$$

$$\{\gamma_1 + q_{12} - \beta_1(n_1 + a)\}\{\gamma_2 + q_{21} - \beta_2(n_2 - a)\} \geq q_{12}q_{21}. \tag{4.7}$$

The complement *of* $\mathcal{A}(\bar{u}_1, \bar{u}_2)$*, denoted by* $\mathcal{A}^C(\bar{u}_1, \bar{u}_2)$*, is defined as*

$$\mathcal{A}^C(\bar{u}_1, \bar{u}_2) := [-\bar{u}_2, \bar{u}_1] \setminus \mathcal{A}(\bar{u}_1, \bar{u}_2).$$

The next lemma provides equivalent conditions for $\mathcal{R}(a) \leq 1$.

Lemma 4.1. *The following statements are equivalent:*

1. $\mathcal{R}(a) \leq 1$*;*

2. *the real parts of all the eigenvalues of the matrix*

$$G(a) := BX_0(a) - C + Q = \begin{pmatrix} \beta_1(n_1 + a) - q_{12} - \gamma_1 & q_{12} \\ q_{21} & \beta_2(n_2 - a) - q_{21} - \gamma_2 \end{pmatrix}, \tag{4.8}$$

are less than or equal to 0;

3. $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$.

The proof is presented in the Appendix C.1.

Note that when $\gamma_i = \gamma$, $\beta_i = \beta$ and $n_1 + a = n_2 - a = n$, conditions (4.5)-(4.7) are equivalent to $\beta n/\gamma \leq 1$ in Ball (1991).

The next results provide conditions for the total amount of resources in the world such that $\mathcal{R}(a) < 1$ is achievable. Define $m_i = \bar{n}_i - (q_{ij} + \gamma_i)/\beta_i$ for $i \neq j, i, j = 1, 2$ and $q = (q_{12}q_{21})/(\beta_1\beta_2)$, the product of the migration rates between countries normalized by the within country transmission rates.

Proposition 4.1. *The total endowment of the resources from both countries* $\bar{u}_1$ *and* $\bar{u}_2$ *satisfy*

$$\bar{u}_1 + \bar{u}_2 \geq M, \tag{4.9}$$

if and only if $\mathcal{A}(\bar{u}_1, \bar{u}_2) \neq \emptyset$ *and*

$$\mathcal{A}(\bar{u}_1, \bar{u}_2) = [-\bar{u}_2, \bar{u}_1] \cap [s_1, s_2], \tag{4.10}$$

where

$$M = \begin{cases} m_2 - \frac{q}{m_1}, & \text{if } \; m_1 \leq m_2 \text{ and } m_1 < -\sqrt{q} \\ m_1 - \frac{q}{m_2}, & \text{if } \; m_2 \leq m_1 \text{ and } m_2 < -\sqrt{q} \\ m_1 + m_2 + 2\sqrt{q}, & \text{if } \; m_1, m_2 \geq -\sqrt{q} \end{cases}, \tag{4.11}$$

and

$$s_1 := \frac{1}{2}(m_2 - m_1 - \bar{u}_2 + \bar{u}_1) - \frac{1}{2}\sqrt{(m_1 + m_2 - \bar{u}_1 - \bar{u}_2)^2 - 4q} \tag{4.12}$$

$$s_2 := \frac{1}{2}(m_2 - m_1 - \bar{u}_2 + \bar{u}_1) + \frac{1}{2}\sqrt{(m_1 + m_2 - \bar{u}_1 - \bar{u}_2)^2 - 4q}. \tag{4.13}$$

For a proof, see Appendix C.2.

From Proposition 4.1, it is obvious that the feasible allocation set $\mathcal{A}(\bar{u}_1, \bar{u}_2)$ is either empty or a connected closed interval in $[-\bar{u}_2, \bar{u}_1]$, which can be characterized by a single threshold M of the world's total resource endowment $\bar{u}_1 + \bar{u}_2$.

4.3 The General Resource Allocation Game in the Limit

We consider a game where each country i tries to minimize the final size of the epidemic in its own population by making its allocation decision. Given the numbers of initial infectives ε and the allocation $\{u_1, u_2\}$, or, equivalently, the net transfer $a = u_1 - u_2$, the autonomous system (4.1), (4.2) and (4.3) has a unique global differentiable solution given the initial conditions and the corresponding assumptions (Capasso, 1993). Hence, each country's response curve is well-defined and we denote by $z_i(a, \varepsilon)$ the (unique) final size in country i as a function of the allocation decision 'a' and the initial number of infectives ε. The following proposition shows that $z_i(a; \varepsilon)$, $i = (1, 2)$ is the unique solution of a system of equations (usually referred to as *final size equations* in the literature, see for example (Anderson and May, 1991)).

Proposition 4.2. *Given an allocation 'a' and initial infectives $\varepsilon \neq 0$, the final sizes $z_1(a; \varepsilon)$ and $z_2(a; \varepsilon)$ are the unique positive solutions (z_1, z_2) to the following system of* final size equations*:*

$$z_2 = h_1(z_1; \epsilon_1, a) \tag{4.14}$$

$$z_1 = h_2(z_2; \epsilon_2, a), \tag{4.15}$$

in which functions h_1 and h_2 are defined as

$$h_1(z_1; \epsilon_1, a) := \frac{\gamma_2}{q_{21}} \left\{ -\epsilon_1 + \left[1 + \frac{q_{12}}{\gamma_1} \right] z_1 - (n_1 + a) \left[1 - e^{-\frac{\beta_1}{\gamma_1} z_1} \right] \right\} \ , \ \textit{and}$$

$$h_2(z_2; \epsilon_2, a) := \frac{\gamma_1}{q_{12}} \left\{ -\epsilon_2 + \left[1 + \frac{q_{21}}{\gamma_2} \right] z_2 - (n_2 - a) \left[1 - e^{-\frac{\beta_2}{\gamma_2} z_2} \right] \right\} \ .$$

The proof is presented in Appendix C.3. Notice that Equations (4.14)-(4.15) are transcendental equations, whose solutions cannot be written in an explicit form.

Equations (4.14)-(4.15) are highly non-linear and the corresponding resource allocation game is intractable. However, allocation decisions are made at the beginning of the pandemic when the initial number of infectives is small. Our approach consists then of replacing $z_i(a;\varepsilon)$ by its limit as ε approaches zero. The following lemma gives the solutions of (4.14)-(4.15) when $(\epsilon_1, \epsilon_2) = (0,0)$.

Lemma 4.2. *Assume* $(\epsilon_1, \epsilon_2) = (0,0)$. *The system of final size equations (4.14)-(4.15) has*

- *a unique non-negative solution at the origin,* $(0,0)$*, when* $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$*;*
- *two non-negative solutions,* $(0,0)$ *and* $(z_1^0(a), z_2^0(a)) > 0$*, when* $a \in \mathcal{A}^C(\bar{u}_1, \bar{u}_2)$*.*

The proof is presented in Appendix C.4.

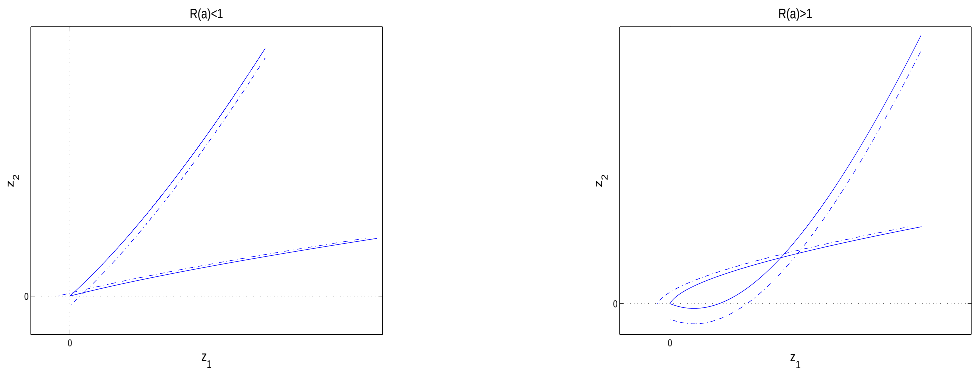
The only possible limit for $(z_1(a;\varepsilon_1), z_2(a;\varepsilon_1))$ is zero when ε approaches 0 from above, as shown in the next proposition. On the other hand, according to Lemma 4.2, two possible candidates for the limit exist in case of an outbreak. The next proposition shows that this limit corresponds then to the positive solution.

Proposition 4.3. *As* (ϵ_1, ϵ_2) *approaches* $(0,0)$ *from above, the solution* $(z_1(a;\varepsilon), z_2(a;\varepsilon))$ *to the final size equations (4.14)-(4.15) asymptotically converges to*

- $(0,0)$ *from above, for all* $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$*;*
- $(z_1^0(a), z_2^0(a))$ *from above, for all* $a \in \mathcal{A}^C(\bar{u}_1, \bar{u}_2)$*.*

See Appendix C.5 for a proof.

Figures 4.1(a)-4.1(b) present some geometric intuitions behind this result. Figure 4.1(a) depicts the case when $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$. The two dotted curves represent the final size equations (4.14)-(4.15), respectively. The two solid curves represent the same equations with $\varepsilon = 0$. As we can see from the graph, when the numbers of initial infectives $\varepsilon = 0$, the final sizes of the epidemic in the two countries, represented by the intersection of the two solid curves, is at the origin. On the other hand, the two curves intersect in the positive orthant when $\varepsilon > 0$. This intersection converges to the origin as ε approaches zero. Figure 4.1(b), on the other hand, represents the case when $a \in \mathcal{A}^C(\bar{u}_1, \bar{u}_2)$. The two solid curves, representing the final size equations (4.14)-(4.15) with $\varepsilon = 0$, have two intersections, one at the origin and the other positive. With $\varepsilon > 0$, the non-negative solution becomes unique and dominates the positive intersection of the two solid curves. This unique positive solution approaches the positive intersection of the two solid curves as ε approaches zero. Note that when $\varepsilon = 0$, the final size is also null. However the final sizes as ε approaches 0 from above do not converge to the origin but to the positive intersect. In other words, the solution of the final size equations is not continuous in ε at zero.



(a) Intersection of $h_1(z_1; \epsilon_1, a)$ and $h_2(z_2; \epsilon_2, a)$ when $\mathcal{R}(a) < 1$

(b) Intersection of $h_1(z_1; \epsilon_1, a)$ and $h_2(z_2; \epsilon_2, a)$ when $\mathcal{R}(a) > 1$

FIGURE 4.1: Illustration of the Final Size Equations.

Based on the previous analysis we introduce the resource allocation game where each country's objective is to minimize the asymptotic final size

$$\hat{z}_i(a) := \lim_{\varepsilon \downarrow 0} z_i(a; \varepsilon) = \begin{cases} 0, & a \in \mathcal{A}(\bar{u}_1, \bar{u}_2) \ , \\ z_i^0(a), & a \in \mathcal{A}^C(\bar{u}_1, \bar{u}_2) \ . \end{cases}$$

Since the final allocation 'a' is a result of decisions among the two countries, the solution concept to the game that we adopt here is the *Pareto Nash Equilibrium* (PNE). Formally, $a^* \in [-\bar{u}_2, \bar{u}_1]$ is a PNE, if

1. no country has an incentive to deviate from the allocation a^*, and

2. there does not exists another $a \in [-\bar{u}_2, \bar{u}_1]$ such that

   (1) no country has an incentive to deviate from 'a', and

   (2) $\hat{z}_i(a) \leq \hat{z}_i(a^*)$ for both $i = 1, 2$ and the inequality holds strictly for some i.

The PNE is more restrictive than Nash equilibrium as a solution concept to a game.

The previous analysis implies that such a Pareto Nash Equilibrium exists in our game, which constitutes the main result of this chapter.

Theorem 4.1. *Any strategy $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$ Pareto dominates strategy $a' \in \mathcal{A}^C(\bar{u}_1, \bar{u}_2)$, and is a PNE in the game.*

This is an immediate consequence of Proposition 4.3.

Theorem 4.1 implies that even without centralized control, decision makers have the incentive to allocate resources to prevent an outbreak.

4.4 The Resource Allocation Game in the Absence of a Major Outbreak

Theorem 4.1 suggests that as long as there are enough resources to bring $\mathcal{R}(a)$ down to one or below, the countries should agree upon such an allocation as $\varepsilon \to 0$. In this section we relax this asymptotic condition by considering positive ϵ_i, which is negligible compared to the population size n_i (or, more precisely, the number of susceptibles $n_i \pm a$). Using a common approximation scheme (Daley and Gani, 1994), we characterize and show the uniqueness of the Pareto Nash equilibrium.

The approximation follows the intuition that for an allocation $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$, the total number of infectives $y_i(t)$ tends to decrease from the already very small ϵ_i, compared to the large population size n_i. Therefore, the susceptible population $x_i(t)$ does not change much over time. As an approximation, we fix $x_i(t)$ at $x_i(0)$

throughout the time horizon. That is, we approximate the dynamics of the epidemic (4.1), (4.2) and (4.3) by the following system of linear ordinary differential equations with constant coefficients,

$$\frac{d\tilde{y}_i}{dt} = \beta_i x_i(0)\tilde{y}_i + \sum_{j=1}^{2} q_{ji}\tilde{y}_j - \gamma_i \tilde{y}_i, \quad j \neq i \tag{4.16}$$

$$\frac{d\tilde{z}_i}{dt} = \gamma_i \tilde{y}_i \tag{4.17}$$

with the initial boundary condition $\tilde{y}_i(0) = \epsilon_i \ll \bar{n}_i$ and $\tilde{z}_i(0) = 0$.

Similar to the previous analysis, we can define the final size $\tilde{z}_i(a;\varepsilon)$ for 'a' in the interior of $\mathcal{A}(\bar{u}_1, \bar{u}_2)$,

$$\tilde{z}_i(a;\varepsilon) := \tilde{z}_i(\infty) = \gamma_i \int_0^\infty \tilde{y}_i(t)dt \ . \tag{4.18}$$

as a function of the allocation $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$ and the number of initial infectives ε. (Note that when 'a' is on the boundary of $\mathcal{A}(\bar{u}_1, \bar{u}_2)$ or in $\mathcal{A}^C(\bar{u}_1, \bar{u}_2)$, the limit does not exist and $\tilde{z}_i(a;\varepsilon)$ tends to infinity. The approximation does not work in such a regime.) The approximation provides an upper-bound such that $z_i(t) < \tilde{z}_i(t)$ and $z_i(a,\varepsilon) < \tilde{z}_i(a,\varepsilon)$. Given 'a' and ε we derive the analytical expressions for $\tilde{z}_i(a;\varepsilon)$ in the following result,

Proposition 4.4. *For an allocation 'a' in the interior of $\mathcal{A}(\bar{u}_1, \bar{u}_2)$, $\tilde{z}_i(a;\varepsilon)$ is a linear function of ε.*

$$\tilde{z}_1(a;\varepsilon) = \frac{-\gamma_1(-\epsilon_1\beta_2 a + \epsilon_1\beta_2(m_2 - \bar{u}_2) - \epsilon_2 q_{21})}{\beta_1\beta_2[(a + m_1 - \bar{u}_1)(-a + m_2 - \bar{u}_2) - q]} \tag{4.19}$$

$$\tilde{z}_2(a;\varepsilon) = \frac{-\gamma_2(\epsilon_2\beta_1 a + \epsilon_2\beta_1(m_1 - \bar{u}_1) - \epsilon_1 q_{12})}{\beta_1\beta_2[(a + m_1 - \bar{u}_1)(-a + m_2 - \bar{u}_2) - q]}. \tag{4.20}$$

Furthermore,

$$\tilde{z}_i(a;\varepsilon) = z_i(a;\varepsilon) + o(\varepsilon), \tag{4.21}$$

i.e., $\tilde{z}_i(a;\varepsilon)$ *is the first-order approximation of* $z_i(a;\varepsilon)$ *for* $i = 1, 2$.

For a proof, see Appendix C.6.

4.4.1 Characterization of the Game Equilibrium

In the game we consider, each country i allocates its resources so as to minimize $\tilde{z}_i(a;\varepsilon)$, which admits a unique minimizer as shown by the next Lemma.

Lemma 4.3. *For any fixed nonzero* ε, $\tilde{z}_1(a;\varepsilon)$ *and* $\tilde{z}_2(a;\varepsilon)$ *are both positive convex functions on the interval* (s_1, s_2) *(*s_1 *and* s_2 *are defined in equations (4.12) and (4.13)), and* $\tilde{z}_i(a;\varepsilon)$ *achieves its global minimum at* $a = t_i$ *defined in terms of model parameters. Further more,*

$$s_1 \leq t_1 \leq t_2 \leq s_2 \ .$$

The proof is presented in Appendix C.7.

The expressions of t_i are given in Appendix C.7. It is worth mentioning that the thresholds t_1 and t_2 critically depend on all model parameters, especially on the ratio between ϵ_1 and ϵ_2. Moreover, if only one of ϵ_1 and ϵ_2 is positive, the minimizer t_i does not depend on ϵ_1 and ϵ_2 .

Figure 4.2 provides an illustration of the shapes of $\tilde{z}_1(a;\varepsilon)$ (the blue dashed line) and $\tilde{z}_2(a;\varepsilon)$ (the green dashed and dotted line), both of which are convex functions in 'a'. We also illustrate the total final size using the solid curve, which will be the subject of discussion in the next section. From the graph, it is clear that when the allocation 'a' is between s_1 and t_1, the final sizes of both countries are monotonically decreasing in 'a'; and they are increasing when the allocation 'a' is between t_2 and s_2. When $a \in [t_1, t_2]$, on the other hand, the two countries' interests conflict with each other. Since $a = u_1 - u_2$, this situation gives rise to a game, which admits a unique Pareto Nash Equilibrium.

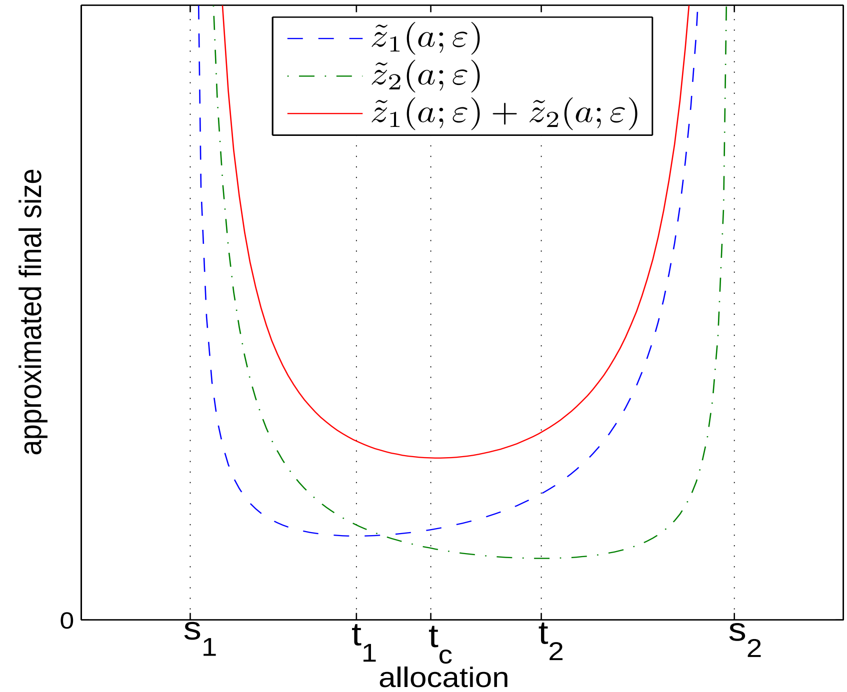


FIGURE 4.2: The first order approximation $\tilde{z}_i(x;\epsilon)$ of the TNFR $z_i(x;\epsilon)$ for $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$.

Theorem 4.2. *Assume each country's objective is to minimize $\tilde{z}_i(a,\varepsilon)$. If the world-wide resources are abundant ($\bar{u}_1 + \bar{u}_2 > M$) and the decentralized decisions are made at the onset of the epidemic ($(\epsilon_1, \epsilon_2) \ll (\bar{n}_1, \bar{n}_2)$), there exists a unique PNE allocation a^* such that $\mathcal{R}(a^*) < 1$. Furthermore, the allocation a^* is implemented through a two-threshold resource transfer structure: there exists $t_1 \le t_2$ determined in terms of the model parameters such that*

1. *if $t_1 > 0$, $\min\{\bar{u}_1, t_1\}$ units of resources are transferred from country 1 to country 2;*
2. *if $t_2 < 0$, $\min\{\bar{u}_2, -t_2\}$ units of resources are transferred from country 2 to country 1;*
3. *if $t_1 \le 0 \le t_2$, neither country gives anything to the other.*

See proof in Appendix C.8.

4.4.2 Central Planner Optimal Allocations

An important question our study seeks to answer is how the previous equilibrium compares with the optimal decision that a central planner (e.g., the World Health Organization) would make in order to minimize the total final size of the epidemic the world population as a whole, without distinguishing between countries. To address this point, we first derive the optimal allocation decision of the central planner. This allows us to derive necessary and sufficient conditions under which the game equilibrium corresponds to the centralized choice.

By Theorem 4.1, we may restrict our search within the set $\mathcal{A}(\bar{u}_1, \bar{u}_2)$. In order to obtain analytical solutions, we follow the same approximation as used in Section 4.4. Formally, we denote allocation $a^\star \in [-\bar{u}_2, \bar{u}_1]$ to be the *central planner's optimal allocation* (CPOA) that minimizes the approximated total final size $\tilde{z}_1(a;\varepsilon)+\tilde{z}_2(a;\varepsilon)$. In Figure 4.2, the solid line illustrates the world's final size $\tilde{z}_1(a;\varepsilon)+\tilde{z}_2(a;\varepsilon)$, which is again a convex function in 'a'. From the graph, we notice that the minimum is achieved between t_1 and t_2, where the two countries' interests conflict. With the help of this intuition, we obtain the main result of this section.

Theorem 4.3. *Assume the central planner's objective is to minimize $\tilde{z}_1(a;\varepsilon) + \tilde{z}_2(a;\varepsilon)$. If the worldwide resources are abundant ($\bar{u}_1 + \bar{u}_2 > M$) and the central planner's decision is made at the onset of the epidemic ($(\epsilon_1, \epsilon_2) \ll (\bar{n}_1, \bar{n}_2)$), there exists a unique CPOA $a^\star$ such that $\mathcal{R}(a^\star) < 1$. Furthermore, the allocation $a^\star$ is implemented through a single-threshold resource transfer structure: there exists $t_c \in [t_1, t_2]$ determined in terms of the model parameters such that the CPOA $a^\star$:*

1. *if $t_c > 0$, transfer $\min\{t_c, \bar{u}_1\}$ units of the resources from country 1 to country 2;*

2. *if* $t_c < 0$, *transfer* $\min\{-t_c, -\bar{u}_2\}$ *units of the resources from country 2 to country 1;*

3. *if* $t_c = 0$, *no resource transfer occurs between the two countries.*

See proof in Appendix C.9.

Notice that $t_c = 0$ occurs with very few sets of model parameters. So Case 3 in Theorem 4.3 may be perceived as degenerate.

Combining Theorem 4.2 and Theorem 4.3, we have the following corollary, which describes the relationship between the centralized optimal solution and the equilibrium.

Corollary 4.1. *Under the condition that the worldwide resources are abundant* ($\bar{u}_1 + \bar{u}_2 > M$) *and the central planner's decision is made at the onset of the epidemic* ($(\epsilon_1, \epsilon_2) \ll (\bar{n}_1, \bar{n}_2)$)*, the effective reproduction ratio is less than 1 under either the CPOA or the PNE; the exact allocations of resources under the CPOA and the PNE, however, coincide if and only if we are in one of the following three situations:*

1. $t_1 \geq \bar{u}_1$, *when all the resources of country 1 are transferred to country 2;*

2. $t_2 \leq -\bar{u}_2$, *when all the resources of country 2 are transferred to country 1;*

3. $t_c = 0$, *no net transfer of resources between these two countries occurs.*

In other words, selfish countries reach an allocation equilibrium that is also optimal for the whole world when they act either completely altruistically or completely selfishly. However, such situations are extreme cases. In general, decentralized decisions result in more infectives for at least one of the countries compared to the centralized choice. Alternatively, one country would need to incur more infectives in its population in order to achieve the best outcome for the overall population.

Furthermore, in a decentralized equilibrium, there exists a wide range of model parameters (such that $0 \in [t_1, t_2]$) under which both country hoard resources within their own territory, which is fundamentally different from the centralized optimal allocation, where such a scenario rarely happens.

4.5 Conclusion

This study constitutes the first analysis of a resource allocation problem to minimize the finial size of a pandemic when different regions allocate their resources selfishly, i.e., to protect their own population. Our main result suggests that the allocation decisions at the equilibrium bring the effective reproduction ratio below one. In other words, if decision makers act rationally to achieve their own best interest, a major outbreak in the world is avoided. In particular, no resource allocation precommitments need to be secured to contain an epidemic globally, and the total number of infected individuals remains small compared to the whole population size. However, we show that in general the final size of the epidemic resulting from decentralized decisions is still larger than that from the optimal centralized allocation rule. This raises difficult ethical questions as one country would need to increase its number of infectives if the final size of the epidemic in the world were to be minimized.

Our model assumes that the transmission parameters remain constant over time and that the population's behaviors do not change significantly. In particular, countries do not close their borders during the pandemic. When countries are able to isolate themselves from the rest of the world at the onset of the epidemic, a stochastic model studied by Sun, Yang, and de Véricourt (2009) is more appropriate. The more difficult situation where transmission rates vary over time remains an open problem for the decentralized as well as centralized cases.

Another limitation of our model is that we assume perfect knowledge of the parameters of the epidemic dynamics. In real circumstances, decision-makers may only have access to partial or even biased information. It is worth studying the impact of partial information, and extend our model to include information acquisition.

We also assume that the effects of the resources on the populations are immediate. An interesting research direction consists in describing the distribution of resources within and between countries that also considering logistic constraints. In particular, congestion effects should be expected in the delivery of drugs and a queueing system could be added to the model in the spirit of Kaplan et al. (2003).

Another natural extension of our model is the case of multiple (i.e., more than two) countries. We expect some preliminary results, such as the conditions for a major outbreak or the uniqueness of the final size to be easily generalized. However, the analysis of the equilibrium appears quite challenging and has proven intractable so far. Further research is needed to study this situation.

This chapter is not concerned with the production of the resources necessary to contain an epidemic. For instance, the world stock of mass prophylaxis is insufficient to tackle an influenza pandemic, (see Smolinski et al. (2003)). Rather, we focus on the actual allocation of these resources in case of an outbreak. However, we believe that considering strategic behaviors of the relevant decision makers can improve preparedness for a pandemic. From this perspective, our primary result suggests that countries should strive to reach a political agreement on the total number of resources available in the world, rather than on allocation rules.

Appendix A

Proofs in Chapter 2

A.1 Mechanism Design by an Informed Principal

Following the mechanism design nomenclature, we refer to the party with all the bargaining power (the retailer in the MSSR case, and the supplier in the SSMR case) as the principal. In both models the principal has private information about its inventory level. This changes the mechanism design problem into an "informed-principal problem." On one hand the principal could exploit its private information to achieve higher payoffs, on the other hand the proposed contract could signal its own private information to the other parties.

In order to deal with the incentives faced by the informed principal, Myerson (1983) introduces the concept of "strong solution" for environments with finite type spaces and finite outcome spaces. An extension to non-finite environments can be found in Mylovanov and Tröger (2008).

Definition A.1 (Safe, Dominated, and Strong Solution)**.** *A mechanism is said to be* safe *if it would remain incentive feasible (incentive compatible and individually rational) for the principal and all agents, even if all agents knew the principal's* *private information. An incentive feasible mechanism is said to be* dominated *if there exists another incentive feasible mechanism such that all types of the principal are at least as well off and a positive mass of types of the principal is strictly better off. A mechanism is said to be a* strong solution *if it is both safe and not dominated.*

Myerson (1983) (Theorem 2) proves that in any environment with finite type spaces and a finite outcome space, any strong solution is an equilibrium outcome of an informed-principal game where any finite simultaneous-move game form is a feasible mechanism. Moreover, Myerson (1983) also points out that any two strong solutions yield the same expected payoff for all types of the principal.

Mylovanov and Tröger (2008) provide an extension of this result to non-finite environments, which can be adopted in our setting. Formally Mylovanov and Tröger show that no type of the principal has an incentive to deviate by offering any finite (simultaneous-move or multi-stage) game. The restriction to finite deviating mechanisms is due to the need to guarantee that any feasible mechanism has an equilibrium.

In Appendices A.2 and A.3, we construct an incentive feasible mechanism that (i) solves the principal problem when its type is publicly known (full-information-optimal), and (ii) minimizes the principal's expected cost (ex-ante optimal). By Lemma 1 of Mylovanov and Tröger (2008), since we are in a private values model, this mechanism is a strong solution for the informed principal problem.

A.1.1 Specializing to SSMR

In this section we specialize the analysis of the informed principal problem to the SSMR scenario (the MSSR scenario is analogous so it is omitted).

The full information optimal for each inventory level x_s of the supplier is given by

$$\begin{aligned} \min_{q,m} \quad & \mathbb{E}_x \left[C_s\left(x_s, q(x_s,x)\right) - \sum\nolimits_{i=1}^{N} m_i(x_s,x) \right] \\ \text{s.t.} \quad & (\text{IC}_i)(x_s),\ (\text{IR}_i)(x_s),\ q_i(x_s,x) \geq 0 \quad i=1,\ldots,N \\ & \sum\nolimits_{i=1}^{N} q_i(x_s,x) \leq x_s \end{aligned} \tag{A.1}$$

where $(\text{IC}_i)(x_s)$ and $(\text{IR}_i)(x_s)$ represent the incentive constraints for the retailers knowing the supplier's inventory level x_s.

The ex-ante optimal is given by

$$\begin{aligned} \min_{q,m} \quad & \mathbb{E}_{x_s,x} \left[C_s\left(x_s, q(x_s,x)\right) - \sum\nolimits_{i=1}^{N} m_i(x_s,x) \right] \\ \text{s.t.} \quad & (\text{IC}_s),\ (\text{IR}_s),\ (\text{IC}_i),\ (\text{IR}_i),\ q_i(x_s,x) \geq 0 \quad i=1,\ldots,N \\ & \sum\nolimits_{i=1}^{N} q_i(x_s,x) \leq x_s. \end{aligned} \tag{A.2}$$

The supplier's problem (2.8) has the objective function of (A.1) and the constraints of (A.2). Under our assumptions, the optimal contract $q^\star, m^\star$ characterized in Theorem 2.2 solves (2.8), (A.1), and (A.2) simultaneously. In particular, by solving (A.1) and (A.2), since we have private values environment, Lemma 1 of Mylovanov and Tröger (2008) establishes that the optimal contract $q^\star, m^\star$ is a strong solution. Thus, the mechanism $q^\star, m^\star$ is also an equilibrium of the associated game, see Myerson (1983) Theorem 2.

A.2 Proofs in Section 2.2

Proof of Theorem 2.1. We divide the proof in two steps: i) the main arguments, and ii) technical results on the allocation q° used in i).

Step 1. Main Arguments. First observe that setting $m_i^\circ(x) = (c_i - h_i)\ q_i^\circ(x)$ implies that (IR_i) holds with equality for all $i = 1, \ldots, M$, thus the retailer extracts the entire cost reduction given by (2.4). It remains to show that the contract

$\{q_i^\circ(x_r,x), m_i^\circ(x_r,x)\}_{i=1}^M$ also satisfies (IC$_i$), (IC$_r$) and (IR$_r$). For any x_r, x, we have

$$C_i(x_i, q_i^\circ(x_r,x)) - m_i^\circ(x_r,x) = C_i(x_i,0) = h_i x_i, \quad \text{for } \ i = 1,\ldots,M,$$

which implies that (IC$_i$) holds with equality. Moreover, by (2.4) we have

$$C_r\left(x_r, \sum_{i=1}^M q_i^\circ(x_r,x)\right) + \sum_{i=1}^M C_i(x_i, q_i^\circ(x_r,x)) \le C_r(x_r,0) + \sum_{i=1}^M C_i(x_i,0)$$

and by the previous equality

$$\begin{aligned} & C_r\left(x_r, \sum_{i=1}^M q_i^\circ(x_r,x)\right) + \sum_{i=1}^M m_i^\circ(x_r,x) \\ = \ & C_r\left(x_r, \sum_{i=1}^M q_i^\circ(x_r,x)\right) + \sum_{i=1}^M \left[C_i(x_i, q_i^\circ(x_r,x)) - C_i(x_i,0)\right] \\ \le \ & C_r(x_r,0). \end{aligned}$$

The last inequality is (IR$_r$). Finally, to see that (IC$_r$) also holds, for each x_r, x_r', x, define

$$\Delta := C_r\left(x_r, \sum_{i=1}^M q_i^\circ(x_r,x)\right) + \sum_{i=1}^M m_i^\circ(x_r,x) - \left[C_r\left(x_r, \sum_{i=1}^M q_i^\circ(x_r',x)\right) + \sum_{i=1}^M m_i^\circ(x_r',x)\right]$$

$$=(h_r+b_r)\int_{x_r+\sum_{i=1}^N q_i^\circ(x_r',x)}^{x_r+\sum_{i=1}^N q_i^\circ(x_r,x)} G_r(z)dz - \sum_{i=1}^M (b_r+h_i-c_i)\left(q_i^\circ(x_r,x) - q_i^\circ(x_r',x)\right).$$

We will show that $\Delta \le 0$ hence (IC$_r$) holds by integrating x out.

Let $\eta_i = G_r^{-1}((h_i+b_r-c_i)/(h_r+b_r))$ for $i=1,\ldots,M$. Note that η_i is decreasing in $i=1,\ldots,M$ since $c_i - h_i$ is increasing. The index $i^\circ(x_r,x)$ is defined in Step 2 below.

If $\eta_1 \le x_r < x_r'$, then $q_i^\circ(x_r,x) = q_i^\circ(x_r',x) = 0$ by Step 2 $(***)$ and hence $\Delta = 0$.

If $x_r < x_r'$ and $x_r \le \eta_i \le x_r + \sum_{i=1}^M x_i$ for some $i=1,\ldots,M$, then by Step 2

$(**)$,

$$G_r\left(x_r+\sum_{j=1}^{M} q_j^\circ(x_r,x)\right)=\frac{b_r+h_{i^\circ(x_r,x)}-c_{i^\circ(x_r,x)}}{h_r+b_r}.$$

Furthermore, we have $q_i^\circ(x_r,x)\geq q_i^\circ(x_r',x)$ by Step 2 $(*)$. Hence,

$$\Delta \leq \sum_{i=1}^{i^\circ(x_r,x)}\left[(h_r+b_r)G_r\left(x_r+\sum_{i=1}^{N} q_i^\circ(x_r,x)\right)-(b_r+h_i-c_i)\right](q_i^\circ(x_r,x)-q_i^\circ(x_r',x))$$

$$=\sum_{i=1}^{i^\circ(x_r,x)}\underbrace{\left[(h_{i^\circ(x_r,x)}-c_{i^\circ(x_r,x)})-(h_i-c_i)\right]}_{\leq 0}\underbrace{(q_i^\circ(x_r,x)-q_i^\circ(x_r',x))}_{\geq 0}\leq 0.$$

If $x_r<x_r'$ and $\eta_M\geq x_r+\sum_{i=1}^{M} x_i$, then by Step 2 $(*\,*\,**)$,

$$G_r\left(x_r+\sum_{j=1}^{M} q_j^\circ(x_r,x)\right)\leq\frac{b_r+h_M-c_M}{h_r+b_r}.$$

Furthermore, we have $q_i^\circ(x_r,x)\geq q_i^\circ(x_r',x)$ by Step 2 $(*)$. Hence,

$$\Delta\leq\sum_{i=1}^{M}\left[(h_r+b_r)G_r(x_r+\sum_{i=1}^{N} x_i)-(b_r+h_i-c_i)\right](q_i^\circ(x_r,x)-q_i^\circ(x_r',x))$$

$$\leq\sum_{i=1}^{M}\underbrace{\left[(h_M-c_M)-(h_i-c_i)\right]}_{\leq 0}\underbrace{(q_i^\circ(x_r,x)-q_i^\circ(x_r',x))}_{\geq 0}\leq 0.$$

If $x_r>x_r'\geq\eta_1$, then $q_i^\circ(x_r,x)=q_i^\circ(x_r',x)=0$ by Step 2 $(*\,*\,*)$ and hence $\Delta=0$.

If $x_r>\eta_1\geq x_r'$, we have $0=i^\circ(x_r,x)<i^\circ(x_r',x)$ and $0=q_i^\circ(x_r,x)\leq q_i^\circ(x_r',x)$ for all $i=1,\ldots,M$ by Step 2 $(*\,*\,*)$

$$\Delta\leq\sum_{i=1}^{i^\circ(x_r',x)}\left[(h_r+b_r)G_r(x_r)-(b_r+h_i-c_i)\right](q_i^\circ(x_r,x)-q_i^\circ(x_r',x))$$

$$\leq\sum_{i=i^\circ(x_r,x)}^{i^\circ(x_r',x)}\underbrace{\left[(h_1-c_1)-(h_i-c_i)\right]}_{\geq 0}\underbrace{(q_i^\circ(x_r,x)-q_i^\circ(x_r',x))}_{\leq 0}\leq 0.$$

If $x_r > x'_r$, which implies $i^\circ(x_r, x) \leq i^\circ(x'_r, x)$ and $q_i^\circ(x_r, x) \leq q_i^\circ(x'_r, x)$ by Step 2 (*), and $x_r \leq \eta_i \leq x_r + \sum_{i=1}^M x_i$ for some $i = 1, \ldots, M$, which implies, by Step 2 (**),

$$G_r\left(x_r + \sum_{j=1}^M q_j^\circ(x_r, x)\right) = \frac{b_r + h_{i^\circ(x_r,x)} - c_{i^\circ(x_r,x)}}{h_r + b_r},$$

then we have

$$\Delta \leq \sum_{i=i^\circ(x_r,x)}^{i^\circ(x'_r,x)} \left[(h_r + b_r)G_r\left(x_r + \sum_{i=1}^N q_i^\circ(x_r, x)\right) - (b_r + h_i - c_i)\right](q_i^\circ(x_r, x) - q_i^\circ(x'_r, x))$$

$$= \sum_{i=i^\circ(x_r,x)}^{i^\circ(x'_r,x)} \underbrace{\left[(h_{i^\circ(x_r,x)} - c_{i^\circ(x_r,x)}) - (h_i - c_i)\right]}_{\geq 0} \underbrace{(q_i^\circ(x_r, x) - q_i^\circ(x'_r, x))}_{\leq 0} \leq 0.$$

If $x_r > x'_r$ but $\eta_M \geq x_r + \sum_{i=1}^M x_i$, we have $q_i^\circ(x_r, x) = q_i^\circ(x'_r, x) = x_i$ for all $i = 1, \ldots, M$. Hence, $\Delta = 0$.

Step 2. Properties of the allocation q°.

Recall the definition of $\eta_i = G_r^{-1}((h_i + b_r - c_i)/(h_r + b_r))$ for $i = 1, \ldots, M$. Note that η_i is decreasing in $i = 1, \ldots, M$ since $c_i - h_i$ is increasing. Because of the convexity of objective function in (2.4), KKT condition is the necessary and sufficient for the global optimality of q°, namely there exists $\nu_i, \lambda_i \geq 0$, $i = 1, \ldots, M$ such that the following (in)equalities are satisfied for $i = 1, \ldots, M$:

$$\begin{gathered}(h_r + b_r)G_r\left(x_r + \sum_{j=1}^M q_j\right) - b_r + c_i - h_i + \nu_i - \lambda_i = 0 \\ 0 \leq \ q_i \leq x_i, \quad \nu_i(x_i - q_i) = 0, \quad \lambda_i q_i = 0\end{gathered} \tag{A.3}$$

Hence, for given x, we consider the following greedy algorithm:

Set $i = 0$ and $q^\circ = 0$. If $x_r \geq \eta_1$, **terminate**.

Iteration Step: if $i \geq M$ **terminate**;

$i \leftarrow i+1$;

if $x_r + \sum_{j=1}^{i} x_j < \eta_i$, set $q_i^\circ = x_i$ and return to **Iteration Step**;

else set $q_i^\circ = \eta_i - x_r - \sum_{j=1}^{i-1} x_j$ and **terminate**.

Let i° denote the exiting index i and note that $q_i^\circ = 0$ for all $i > i^\circ$. Hence $q_i^\circ(x_r, x)$, $i = 1, \ldots, M$, is the one given in the statement of the theorem. Moreover, define

$$\nu_i = \begin{cases} (c_{i^\circ} - h_{i^\circ}) - (c_i - h_i), & \text{if} \quad i < i^\circ, \\ 0, & \text{if} \quad i \geq i^\circ, \end{cases}$$

and

$$\lambda_i = \begin{cases} 0, & \text{if} \quad i \leq i^\circ, \\ (c_i - h_i) - (c_{i^\circ} - h_{i^\circ}), & \text{if} \quad i > i^\circ. \end{cases}$$

Then $\nu_i, \lambda_i \geq 0$ for all i, together with $q_i^\circ(x_r, x)$, $i = 1, \ldots, M$ satisfy (A.3). Therefore, $q_i^\circ(x_r, x)$, $i = 1, \ldots, M$, indeed solves (2.4).

It follows that:

($*$) $i^\circ = i^\circ(x_r, x)$ and $q_i^\circ(x_r, x)$ is non-increasing in x_r for $i = 1, 2, \ldots, M$;

($**$) $x_r + \sum_{j=1}^{M} q_j^\circ(x_r, x) = \eta_{i^\circ}$ if $x_r \leq \eta_i \leq x_r + \sum_{i=1}^{M} x_i$ for some $i = 1, \ldots, M$;

($***$) $q_i^\circ(x_r, x) = 0$ for $i = 1, \ldots, M$ if $x_r \geq \eta_1$ (and hence $x_r \geq \eta_i$ for all $i = 1, \ldots, M$).

($****$) $q_i^\circ(x_r, x) = x_i$ for $i = 1, \ldots, M$ if $\eta_M \geq x_r + \sum_{i=1}^{M} x_i$ (and hence $\eta_i \geq x_r + \sum_{i=1}^{M} x_i$ for all $i = 1, \ldots, M$).

$\square$

A.3 Proofs of Section 3.2

Proof of Theorem 2.2. The proof consists in showing that the contract $q^\star, m^\star$ constitutes a strong solution (see Appendix A.1).

By Lemma A.4, we note that for any given x_s, $\{q^\star(x_s, x), m^\star(x_s, x)\}$ satisfies $(\mathrm{IC}_i)(x_s)$ and $(\mathrm{IR}_i)(x_s)$ and hence is incentive feasible for all the retailers even if they knew x_s. By Definition A.1, in order to show $\{q^\star(x_s, x), m^\star(x_s, x)\}$ is safe for the supplier, we need to verify (IC_s) and (IR_s). Since the contract $q^\star, m^\star$ minimizes

$$\mathbb{E}_x\left[C_s\left(x_s, q(x_s, x)\right) - \sum\nolimits_{i=1}^N m_i(x_s, x)\right]$$

over a set that contains the contract $q \equiv 0$ and $m \equiv 0$, we have

$$\mathbb{E}_x\left[C_s\left(x_s, q^\star(x_s, x)\right) - \sum_{i=1}^N m_i^\star(x_s, x)\right] \leq C_s\left(x_s, 0\right)$$

and (IR_s) holds. In order to verify (IC_s), by the supplier's cost structure and Lemma A.4, for any $x_s, x_s' \in \mathcal{X}_s$,

$$\mathbb{E}_x\left[C_s\left(x_s, q^\star(x_s', x)\right) - \sum\nolimits_{i=1}^N m_i^\star(x_s', x)\right] = \mathbb{E}_x\left[\pi^\star(x, q^\star(x_s', x))\right] + h_s x_s \tag{A.4}$$

where $\sum_{i=1}^N q_i^\star(x_s', x) \leq x_s$ for all $x \in \mathcal{X}$.

Thus, for any $x_s < x_s'$, we cannot have $\sum_{i=1}^N q_i^\star(x_s', x) > x_s$ for any x. If $\sum_{i=1}^N q_i^\star(x_s', x) \leq x_s < x_s'$ for all x, then x_s is enough to bring retailer i's final inventory position to the minimum $y_i^\star(x_i)$ of $V_i(\cdot|x_i)$ as defined in (A.18) for all i, i.e. for all x,

$$q_i^\star(x_s, x) = q_i^\star(x_s', x) = \left(y_i^\star(x_i) - x_i\right)^+.$$

Thus, $\pi^\star(x, q^\star(x_s, x)) = \pi^\star(x, q^\star(x_s', x))$ for all x. Therefore, (IC_s) holds (with equality) by (A.4).

For any $x_s > x_s'$, we must have $\sum_{i=1}^N q_i^\star(x_s', x) \le x_s' < x_s$ and $\sum_{i=1}^N q_i^\star(x_s, x) \le x_s$. Thus by Lemma A.4, since $q^\star$ solves (A.15), for any x,

$$\pi^\star(x, q^\star(x_s, x)) \le \pi^\star(x, q^\star(x_s', x)),$$

which immediately implies (IC_s) by (A.4).

On the other hand, because (A.6) minimizes the supplier's cost for any given x_s subject to only the retailers' incentive constraints (IC_i) and (IR_i) by ignoring (IC_s) and (IR_s). Hence, we must have for any other incentive feasible mechanism $\{\tilde{q}(x_s, x), \tilde{m}(x_s, x)\}$ such that, for all x_s,

$$\mathbb{E}_x \left[C_s\left(x_s, q(x_s, x)\right) - \sum_{i=1}^N m_i(x_s, x) \right] \le \mathbb{E}_x \left[C_s\left(x_s, \tilde{q}(x_s, x)\right) - \sum_{i=1}^N \tilde{m}_i(x_s, x) \right].$$

Thus, by Definition A.1, $\{q^\star(x_s, x), m^\star(x_s, x)\}$ is not dominated by any other incentive feasible mechanism. Therefore, by Definition A.1, $\{q^\star(x_s, x), m^\star(x_s, x)\}$ is a strong solution for the supplier's problem. □

Proof of Theorem 2.3. The result follows from Lemma A.7 and the observation that the fractile representation arises from $\mu_i(x_i + q_i^\star(x_s, x)|x_i) = U^\star$ being equivalent to $\widetilde{G}_i(x_i + q_i^\star(x_s, x)|x_i) = (U^\star + h_s + b_i - c_i)/(h_i + b_i)$. □

The classic centralized problem where there is no private information is:

$$\begin{aligned} \Pi_N^\circ(x_s, x) := \min_{q \ge 0} \quad & C_s\left(x_s, q\right) + \sum_{i=1}^N \left\{C_i\left(x_i, q_i\right) - C_i(x_i, 0)\right\} - C_s\left(x_s, 0\right) \\ & \textstyle\sum_{i=1}^N q_i \le x_s. \end{aligned} \tag{A.5}$$

As a corollary of Theorem 2.3, we can immediately characterize the solution $\{q^\circ, m^\circ\}$ to (A.5) in the next result, where $\phi_i(x_i) := (h_i + b_i)G_i(x_i) - (h_s + b_i - c_i)$ for $i = 1, \ldots, N$ and $\pi^\circ(x, q) := \sum_{i=1}^N \{C_i(x_i, q_i) - (h_s - c_i)q_i\}$:

Corollary A.1. *Fix an inventory profile* $x_s \in \mathcal{X}_s$, $x \in \mathcal{X}$, *and relabel retailers, if necessary, so that*

$$\phi_1(x_1) \leq \phi_2(x_2) \leq \cdots \leq \phi_N(x_N).$$

Then there exists a Lagrange multiplier $\min_{i=1,\ldots,N}\{c_i - h_s - b_i\} \leq U^\circ \leq 0$ *and an integer* n° *such that the allocation rule* q° *to* (A.5) *is determined by*

(A.1.1) for all $i \leq n^\circ$, $q_i^\circ > 0$ *and* $x_i + q_i^\circ = \left(\frac{U^\circ + h_s + b_i - c_i}{h_i + b_i}\right)$*-fractile of* $G_i(\cdot)$*;*

(A.1.2) for all $j > n^\circ$, $q_j^\circ = 0$ *and* $x_j \geq \left(\frac{U^\circ + h_s + b_j - c_j}{h_j + b_j}\right)$*-fractile of* $G_j(\cdot)$*; and*

(A.1.3) $\sum_{i=1}^N q_i^\circ \leq x_s$ *and* $U^\circ \cdot (\sum_{i=1}^N q_i^\circ - x_s) = 0$.

Proof. Following the same argument as in the proof of Lemma A.4 and Theorem 2.3 by substituting $\frac{F_i(x_i)}{f_i(x_i)}$ with 0, the relaxed problem (A.15) is reduced to the centralized problem (A.5). Notice that this is a convex program even without Assumption 2.1. Furthermore, there is no monotonicity constraints and hence we do not need Assumption 2.2. □

Proof of Proposition 2.2. By Theorem 2.3, if $q_i^\star > 0$ we have

$$q_i^\star = \widetilde{G}_i^{-1}\left(\frac{U^\star + h_s + b_i - c_i}{b_i + h_i} \,\middle|\, x_i\right) \leq G_i^{-1}\left(\frac{U^\star + h_s + b_i - c_i}{b_i + h_i}\right) \leq G_i^{-1}\left(\frac{h_s + b_i - c_i}{b_i + h_i}\right)$$

since $G_i(y) \leq \widetilde{G}_i(y|x_i)$ and $U^\star \leq 0$. □

Proof of Proposition 2.1. The second inequality is straightforward. Let us focus on the first inequality. Since $y_i^\star(x_i)$ is the unrestricted minimum for each function $V_i(\cdot|x_i)$ defined in (A.18). Note also that the unrestricted centralized allocation corresponds to $y^\star(\underline{x}_i)$, since the factor $F_i(\underline{x}_i)/f_i(\underline{x}_i) = 0$. Thus, we have

$$\sum_{i=1}^N q_i^\star(x) = \min\left\{x_s, \sum_{i=1}^N (y_i^\star(x_i) - x_i)^+\right\},$$

and

$$\sum_{i=1}^{N} q_i^{\circ}(x) = \min\left\{x_s, \sum_{i=1}^{N}(y^\star(\underline{x}_i) - x_i)^+\right\}.$$

Since $y_i^\star(\cdot)$ is non-increasing by Lemma A.6, we have

$$\min\left\{x_s, \sum_{i=1}^{N}(y_i^\star(x_i) - x_i)^+\right\} \le \min\left\{x_s, \sum_{i=1}^{N}(y_i^\star(\underline{x}_i) - x_i)^+\right\},$$

which immediately yields the result. □

Proof of Proposition 2.3. If $x_s > \sum_{i=1}^N (y_i^\star(x_i) - x_i)^+$, by Lemma 2.3, the lagrangian multiplier $U^\star(x_s, x) = 0$ and the result follows. Otherwise, $x_s \le \sum_{i=1}^N (y_i^\star(x_i) - x_i)^+$ and we have $\sum_{i=1}^N q_i^\star = \sum_{i=1}^N q_i^\circ = x_s$ by Proposition 2.1. Since $\widetilde{G}_i^{-1}(\alpha|x_i) \le \widetilde{G}_i^{-1}(\alpha|\underline{x}_i) = G_i^{-1}(\alpha)$ for $\alpha \in [0,1]$ and $i = 1, \ldots, N$, if $U^\star < U^\circ$ it follows by Lemma 2.3 that $x_i + q_i^\star < x_i + q_i^\circ$ for all i such that $q_i^\star > 0$. That yields $\sum_{i=1}^N q_i^\circ > x_s$ and therefore $U^\star < U^\circ$ cannot hold. □

A.4 Proofs of Section 2.4

Proof of Theorem 2.4. Under Assumption 2.2 for $i < j$ that $\frac{F(x_i)}{f(x_i)} \le \frac{F(x_j)}{f(x_j)}$. Hence by definition of $\mu_i(\cdot|x_i)$ in (A.17), $\mu_i(y|x_i) \le \mu_j(y|x_j)$ for all y. Moreover, for any i, $\mu_i(y|x_i) \le 0$ and non-decreasing for $y \in [0, y_i^\star(x_i)]$, and $\mu_i(y|x_i) \ge 0$ otherwise by Lemma A.5. Thus, if $\mu_i(x_i|x_i) > 0$, we have $x_i > y_i^\star(x_i)$ and we have correspondingly $q_i^\star = 0$ and the ith retailer is not served; so, without loss of generality, we may assume $\mu_i(x_i|x_i) \le 0$, i.e. $x_i \le y_i^\star(x_i)$. Thus, the condition $x_1 \le x_2 \le \cdots \le x_N$ implies that $\mu_1(x_1|x_1) \le \mu_2(x_2|x_2) \le \cdots \le \mu_N(x_N|x_N)$ since $\mu_i(x_i|x_i) \le \mu_{i+1}(x_i|x_{i+1}) \le \mu_{i+1}(x_{i+1}|x_{i+1})$.

By Lemma 2.3 we have two cases. Either $q^\star = 0$, in which case $n^\star = 0$ and both results (1) and (2) hold. Otherwise, for $i \le n^\star$ we have $q_i^\star > 0$, $\mu_i(x_i + q_i^\star|x_i) = U^\star$, and

$x_i+q_i^\star \le y_i^\star(x_i)$. Thus, $\mu_{i+1}(x_i+q_i^\star|x_{i+1}) \ge U^\star$ which implies that $x_i+q_i^\star \ge x_{i+1}+q_{i+1}^\star$ by the monotonicity of $\mu_{i+1}(\cdot|x_{i+1})$ up to $y_{i+1}^\star(x_{i+1}) \ge x_{i+1}+q_{i+1}^\star$.

To establish the second result we can assume $q^\star \neq 0$. Note that if for some $i_s \le n^\star$ we have $q_{i_s}^\star \ge q_{i_s}^\circ$, for any $i \le i_s$ we have $x_i + q_i^\star \ge x_i + q_i^\circ = x_{i_s} + q_{i_s}^\circ$ by the result (1) and the balancing property of the centralized supply chain solution. Thus, if $n^\star > n^\circ$, $q_{n^\star}^\star > 0 = q_{n^\star}^\circ$ which yields $\sum_{i=1}^N q_i^\star > \sum_{i=1}^N q_i^\circ$ which contradicts Proposition 2.1. Thus, $n^\star \le n^\circ$ and (2) follows. □

Proof of Theorem 2.5. Fix $x \in \mathcal{X}$ with $x_i \le x_{i+1}$ and $x_1 < x_2$, and let $q^\star = q^\star(x_s, x)$ and $q^\circ = q^\circ(x_s, x)$ be the optimal allocations for the asymmetric information and the centralized supply chain.

It suffices to show that $q_1^\star > q_1^\circ$. Indeed, if $q_i^\star > q_i^\circ$ we have $x_j + q_j^\star \ge x_i + q_i^\star > x_i + q_i^\circ = x_j + q_j^\circ$, hence $q_j^\star > q_j^\circ$ for all $j \le i$. Thus, if overshooting happens for retailer i, it must also happen for all retailers $j \le i$.

We first establish that Conditions (i) and (ii) are sufficient for overshooting.

First assume that $x_s \le \sum_{i=1}^N (y_i^\star(x_i) - x_i)^+$, so that $\sum_{i=1}^N q_i^\star = \sum_{i=1}^N q_i^\circ = x_s$. Therefore either $q^\star = q^\circ$ or there is an index i such that $q_i^\star > q_i^\circ$ since they are non-negative. Since $x_s > x_2 - x_1$, we have $q_2^\circ > 0$ by (2.19). Thus, if $q_2^\star = 0$ we are done. Otherwise, by Lemma 2.3, $\mu_1(x_1 + q_1^\star|x_1) = \mu_2(x_2 + q_2^\star|x_2)$ and since $F(x_1)/f(x_1) < F(x_2)/f(x_2)$ we have $x_1 + q_1^\star > x_2 + q_2^\star$. Therefore $q^\star \neq q^\circ$.

Second, if $\sum_{i=1}^N (y_i^\star(x_i) - x_i)^+ < x_s$, we have $q_i^\star = (y_i^\star(x_i) - x_i)^+$. Also, note that $x_1 + q_1^\circ = (x_s + \sum_{i=1}^{n^\circ} x_i)/n^\circ$ so that Condition (ii) is equivalent to $\widetilde{G}(x_1 + q_1^\circ|x_1) < (h_s + b - c)/(h + b)$ which implies that $x_1 + q_1^\circ < y_1^\star(x_1)$ by Lemma A.5. Hence we have $q_1^\circ < q_1^\star$, i.e. overshooting occurs.

Next we show the necessity of Conditions (i) and (ii). If $x_s \le x_2 - x_1$ we have $q_1^\circ = (G^{-1}(\frac{h_s+b-c}{h+b}) - x_1)^+ \wedge x_s$ and $q_i^\circ = 0$ for all $i = 2, \ldots, N$. By Theorem 2.4 part (ii), we also have $q_i^\star = 0$ for $i = 2, \ldots, N$, and, by Proposition 2.1, $q_1^\star \le q_1^\circ$. Thus, overshooting does not occur.

If $\widetilde{G}(x_1 + q_1^\circ | x_1) \geq \frac{h_s+b-c}{h+b}$, we have $y_1^\star(x_1) \leq x_1 + q_1^\circ$ by Lemma A.5. Therefore, $q_1^\star \leq (y_1^\star(x_1) - x_1)^+ \leq q_1^\circ$. Using the fact that q° balances the final inventory position of served retailers and Theorem 2.4, we have that $x_i + q_i^\circ \geq x_1 + q_1^\star \geq x_i + q_i^\star$ so that $q_i^\star \leq q_i^\circ$, i.e. overshooting does not occur. □

A.5 Characterization of (IC$_i$) and (IR$_i$) in the SSMR case

The following is the supplier's problem in Section 3.2 if its inventory/capacity level were publicly known:

$$\begin{aligned} \Pi_N^\star(x_s) := \min_{q,m} \quad & \mathbb{E}_x \left[C_s\left(x_s, q(x_s, x)\right) - \textstyle\sum_{i=1}^N m_i(x_s, x) \right] \\ \text{subject to} \quad & \textstyle\sum_{i=1}^N q_i(x_s, x) \leq x_s \\ & (\mathrm{IC}_i)(x_s),\ (\mathrm{IR}_i)(x_s),\ q_i(x_s, x) \geq 0 \text{ for } i = 1, \ldots, N \end{aligned} \tag{A.6}$$

where (IC$_i$)(x_s) and (IR$_i$)(x_s) are the incentive constraints for the retailers when x_s is publicly known.

In this section, we abbreviate $m_i(x_s, x)$ as $m_i(x)$, $q_i(x_s, x)$ as $q_i(x)$. For any q and m let

$$W_i(x_i) := \mathbb{E}_{x_{-i}}\left[C_i(x_i, q_i(x)) + m_i(x)\right], \quad i = 1, \ldots, N. \tag{A.7}$$

Lemma A.1. *Fix any $i \in \{1, \ldots, N\}$, suppose that $q_i(x)$ is non-increasing in $x_i \in \mathcal{X}_i$ for all $x_{-i} \in \mathcal{X}_{-i}$. Then the pair $q_i(x), m_i(x)$, where*

$$m_i(x) = C_i(x_i, 0) - C_i(x_i, q_i(x)) - (h_i + b_i) \int_{x_i}^{\bar{x}_i} \{G_i(z + q_i(z, x_{-i})) - G_i(z)\}\, dz, \quad x \in \mathcal{X}.$$

satisfies the following inequality

$$C_i(x_i, q_i(x)) + m_i(x) \leq C_i(x_i, q_i(x_i', x_{-i})) + m_i(x_i', x_{-i}), \quad \forall x_i' \in \mathcal{X}_i, x \in \mathcal{X}, \tag{A.8}$$

and hence satisfies (IC$_i$).

Proof. Notice that

$$C_i(x_i, q_i) - C_i(x_i', q_i) = (h_i + b_i) \int_{x_i'}^{x_i} G_i(z + q_i) dz - b_i(x_i - x_i').$$

Thus, we have

$$\begin{aligned}
&[C_i(x_i, q_i(x_i', x_{-i})) + m_i(x_i', x_{-i})] - [C_i(x_i, q_i(x)) + m_i(x)] \\
=&[C_i(x_i, q_i(x_i', x_{-i})) - C_i(x_i', q_i(x_i', x_{-i}))] \\
&+ [C_i(x_i', q_i(x_i', x_{-i})) + m_i(x_i', x_{-i})] - [C_i(x_i, q_i(x)) + m_i(x)] \\
=&[C_i(x_i, q_i(x_i', x_{-i})) - C_i(x_i', q_i(x_i', x_{-i}))] \\
&+ [C_i(x_i', 0) - C_i(x_i, 0)] - (h_i + b_i) \int_{x_i'}^{x_i} [G_i(z + q_i(z, x_{-i})) - G_i(z)]\, dz \\
=&(h_i + b_i) \int_{x_i'}^{x_i} [G_i(z + q_i(x_i', x_{-i})) - G_i(z + q_i(z, x_{-i}))]\, dz \geq 0,
\end{aligned}$$

where the last inequality is due to the fact that if $x_i > x_i'$, $q_i(x_i', x_{-i}) \geq q_i(z, x_{-i})$ for all $z \in [x_i, x_i']$ because of monotonicity of $q_i(x)$ in x_i for any given x_{-i}. (Similarly if $x_i < x_i'$ we have $q_i(x_i', x_{-i}) \leq q_i(z, x_{-i})$ for all $z \in [x_i, x_i']$.) This shows (A.8) and taking the expectation of (A.8) with respect to x_{-i} yields (IC$_i$). $\square$

Lemma A.2. *For any $i = 1, \ldots, N$, if q_i and m_i satisfy* (IC$_i$)*, then for any $x_i \in [\underline{x}_i, \bar{x}_i]$, we have*

$$W_i(x_i) = W_i(\bar{x}_i) - \mathbb{E}_{x_{-i}} \left[\int_{x_i}^{\bar{x}_i} \{(h_i + b_i) G_i(z + q_i(z, x_{-i})) - b_i\}\, dz \right], \tag{A.9}$$

or equivalently,

$$\mathbb{E}_{x_{-i}} [m_i(x)] = W_i(\bar{x}_i) - \mathbb{E}_{x_{-i}} \left[C_i(x_i, q_i(x)) + \int_{x_i}^{\bar{x}_i} \{(h_i + b_i) G_i(z + q_i(z, x_{-i})) - b_i\}\, dz \right]. \tag{A.10}$$

Proof. Define

$$\widehat{W}_i(x_i, x_i') := \mathbb{E}_{x_{-i}}\left[C_i(x_i, q_i(x_i', x_{-i})) + m_i(x_i', x_{-i})\right].$$

Then (IC_i) can be formulated as

$$W_i(x) = \min_{x_i' \in [\underline{x}_i, \bar{x}_i]} \widehat{W}_i(x_i, x_i').$$

We have

$$\frac{\partial}{\partial x_i}\widehat{W}_i(x_i, x_i') = \mathbb{E}_{x_{-i}}\left[\frac{\partial}{\partial x_i}C_i(x_i, q_i(x_i', x_{-i}))\right] = \mathbb{E}_{x_{-i}}\left[(h_i + b_i)\ G_i(x_i + q_i(x_i', x_{-i})) - b_i\right],$$

because the differentiation and expectation operators commute (Rosenthal, 2000, Proposition 9.2.1). Thus $\left|\frac{\partial}{\partial x_i}\widehat{W}_i(x_i, x_i')\right| \leq \max\{h_i, b_i\}$ for any (x_i, x_i'), and (A.9) follows from the *envelope theorem* of Milgrom and Segal (2002). The equality in (A.10) follows from the definition of W_i in (A.7). □

Corollary A.2. *For $i = 1, \ldots, N$, if q_i and m_i satisfies* (IC_i)*, then* (IR_i) *reduces to*

$$W_i(\bar{x}_i) \leq C_i(\bar{x}_i, 0). \tag{A.11}$$

Proof. By Lemma A.2, we have

$$\begin{aligned} W_i(x_i) - C_i(x_i, 0) &= W_i(\bar{x}_i) - C_i(\bar{x}_i, 0) \\ &\quad -(h_i + b_i)\mathbb{E}_{x_{-i}}\left[\int_{x_i}^{\bar{x}_i} \underbrace{\{G_i(z + q_i(z, x_{-i})) - G_i(z)\}}_{\geq 0} dz\right] \\ &\leq W_i(\bar{x}_i) - C_i(\bar{x}_i, 0). \end{aligned}$$

Thus the (IR_i) constraint, $W_i(x) - C_i(x_i, 0) \leq 0$ for all x_i, is equivalent to imposing only (A.11). □

Lemma A.3. *Let* $\{q_i^\star(\cdot)\}_{i=1}^N$ *denote the solution to the following program:*

$$\min_{q(\cdot)} \quad \mathbb{E}_x\left[\sum_{i=1}^N \left\{(c_i - h_s)q_i(x) + C_i(x_i, q_i(x)) + (h_i + b_i)\tfrac{F_i(x_i)}{f_i(x_i)} G_i(x_i + q_i(x))\right\}\right]$$

$$\text{s.t.} \quad \textstyle\sum_{i=1}^N q_i(x) \le x_s, \quad q_i(x) \ge 0, \quad x \in \mathcal{X}. \tag{A.12}$$

If $q_i^\star(x)$ *is also non-increasing in* x_i *for all* $x_{-i} \in \mathcal{X}_{-i}$ *and* $i = 1, \ldots, N$, *then the allocation rule* $\{q_i^\star(\cdot)\}_{i=1}^N$ *together with the payment rule*

$$m_i^\star(x) \;=\; C_i(x_i, 0) - C_i(x_i, q_i^\star(x)) - (h_i + b_i)\int_{x_i}^{\bar{x}_i} \{G_i(z + q_i^\star(z, x_{-i})) - G_i(z)\}\, dz,$$
$$i = 1, \ldots, N, \; x \in \mathcal{X}. \tag{A.13}$$

solves (A.6).

Proof. Replace the term $\{\mathbb{E}_{x_{-i}}[m_i(x)]\}_{i=1}^N$ in (A.6) with the expression in (A.10). In light of Corollary A.2 and Lemma A.1, we can focus on the following relaxed minimization problem:

$$\min_{\substack{q_i(\cdot), \ldots, q_N(\cdot) \\ W_1(\bar{x}_1), \ldots, W_N(\bar{x}_N)}} \quad \mathbb{E}_x\left[C_s\left(x_s, q(x)\right)\right] + \sum_{i=1}^N \mathbb{E}_x\left[C_i(x_i, q_i(x))\right]$$
$$+ \sum_{i=1}^N (h_i + b_i)\mathbb{E}_x\left[\int_{x_i}^{\bar{x}_i} G_i(z + q_i(z, x_{-i}))dz\right] - \sum_{i=1}^N W_i(\bar{x}_i) \tag{A.14}$$
$$\text{s.t.} \sum_{i=1}^N q_i(x) \le x_s, \; q_i(x) \ge 0, \text{ and } W_i(\bar{x}_i) \le C_i(\bar{x}_i, 0) \; \forall x \in \mathcal{X}, i = 1, \ldots, N.$$

Since the objective function is decreasing in $W_i(\bar{x}_i)$ for $i = 1, \ldots, N$, it is optimal to set,

$$W_i^\star(\bar{x}_i) \;=\; C_i(\bar{x}_i, 0), \quad i = 1, \ldots, N.$$

Furthermore, we have,

$$\begin{aligned}
\mathbb{E}_x\left[\int_{x_i}^{\bar{x}_i} G_i(z+q_i(z,x_{-i}))dz\right] &= \mathbb{E}_{x_{-i}}\left[\mathbb{E}_{x_i}\left[\int_{x_i}^{\bar{x}_i} G_i(z+q_i(z,x_{-i}))dz\right]\right] \\
&= \mathbb{E}_{x_{-i}}\left[\mathbb{E}_{x_i}\left[\frac{F_i(x_i)}{f_i(x_i)}G_i(x_i+q_i(x))\right]\right] \\
&= \mathbb{E}_x\left[\frac{F_i(x_i)}{f_i(x_i)}G_i(x_i+q_i(x))\right], \quad i=1,\ldots,N
\end{aligned}$$

where the second equality follows from integration by parts. Substituting the last expression into (A.14) yields (A.12). The result follows from Lemma A.1. □

Our approach of solving the problem reformulated in Lemma A.3 is to solve the following pointwise optimization problem by ignoring the monotonicity constraints.

$$\begin{aligned}
\min_{q\geq 0} \quad & \pi^\star(x,q) \\
& \textstyle\sum_{i=1}^N q_i \leq x_s,
\end{aligned} \tag{A.15}$$

where $\pi^\star(x,q) := \sum_{i=1}^{N}\Big[(c_i-h_s)q_i + C_i(x_i,q_i) + (h_i+b_i)\frac{F_i(x_i)}{f_i(x_i)}G_i(x_i+q_i)\Big]$.

In particular, one needs to account for the lack of global convexity in the objective function. We circumvent by establishing that the function is (globally) quasi-convex and convexity holds on the relevant region so that KKT conditions can still be used to characterize the global solution, which turns out to satisfy the neglected monotonicity conditions under our mild conditions on the distribution functions.

Lemma A.4. *Under Assumptions 2.1 and 2.2, the solution $q^\star(x_s,x)$ given in* (A.15) *and the payment $m^\star(x_s,x)$ defined by* (A.13) *solves (A.6).*

Proof. Combining Lemma A.3, Lemma A.7 and Lemma A.8 we see that the relaxed solution to the pointwise optimization problem (A.15) indeed solves the problem (A.6) under Assumptions 2.1 and 2.2. □

A.6 Quasi-convexity and Monotonicity in the SSMR Case

Lemma A.5 below will be applied through out the paper with $\alpha = (h_s + b_i - c_i)/(h_i + b_i)$, $G = G_i$, $g = g_i$, $\zeta = [F_i(x_i)/f_i(x_i)] \geq 0$.

Lemma A.5. *Let G be a probability distribution with a upper-semi continuous probability density function g such that $1 - G$ is log-concave, $\zeta \geq 0$, and $\alpha \in (0, 1)$. Define $y^\star = \inf\{y \in [0, +\infty) | -\alpha + G(y) + \zeta g(y) \geq 0\}$. Then*

(i) $-\alpha + G(y) + \zeta g(y)$ is non-decreasing for $y \in [0, y^\star]$;

(ii) $-\alpha + G(y) + \zeta g(y) \geq 0$, for all $y \in [y^\star, +\infty)$.

Proof. It is equivalent to show that $1 - G(y) - \zeta g(y)$ is non-increasing in $y \in [0, y^\star]$ and for $y \in [y^\star, \infty)$

$$1 - G(y) - \zeta g(y) \leq 1 - \alpha. \tag{A.16}$$

By definition of $y^\star$, if $1 - G(y) - \zeta g(y) \leq 0$, we must have $y \in [y^\star, \infty)$ and (A.16) holds for this case. Otherwise note that

$$\log\left(1 - G(y) - \zeta g(y)\right) = \log(1 - G(y)) + \log\left(1 - \zeta \frac{g(y)}{1 - G(y)}\right).$$

By the monotonicity of G we have $\log(1-G(y))$ is non-increasing and by log-concavity of $1 - G$ we have that $-g(y)/[1 - G(y)]$ is non-increasing in y. Therefore both terms are non-increasing in y, establishing (i) and (ii) since $1 - G(y^\star) - \zeta g(y^\star) \leq 1 - \alpha$. □

Lemma A.6 below will typically be applied to $\alpha_i = (h_s + b_i - c_i)/(h_i + b_i)$, $a = 1$, and $\zeta_i(x_i) = F_i(x_i)/f_i(x_i)$ as a non-decreasing function of x_i for fixed x_{-i}.

Lemma A.6. *Assume that ζ_i is non-decreasing (non-increasing) in x_i, and let $\alpha_i \in (0, 1)$, $a > 0$, and G_i be a probability distribution with density function g_i. Then $y_i^\star(x_i) = \inf\{y \in [0, +\infty) : -\alpha_i + G_i(y) + a\zeta_i(x_i)g_i(y) \geq 0\}$ is non-increasing (non-decreasing) in x_i.*

Proof. Pick $x_i < x_i'$ so that $\zeta_i(x_i) \leq \zeta_i(x_i')$. For any y

$$0 \leq -\alpha_i + G_i(y) + a\zeta_i(x_i)g_i(y) \leq -\alpha_i + G_i(y) + a\zeta_i(x_i')g(y)$$

so that any y considered in the infimum problem for $y_i^\star(x_i)$ is also considered in the infimum problem for $y_i^\star(x_i')$. Thus, $y_i^\star(x_i) \geq y_i^\star(x_i')$. Similar arguments apply when $\zeta_i(x_i)$ is non-increasing in x_i. □

In what follows we need the definition

$$\mu_i(y|x_i) := (h_i + b_i)\widetilde{G}_i(y|x_i) - (h_s + b_i - c_i) \tag{A.17}$$

which generalizes (2.12).

Lemma A.7. *Under Assumption 2.1, fix $x \in \mathcal{X}$, relabeling retailers if necessary so that*

$$\mu_1(x_1|x_1) \leq \mu_2(x_2|x_2) \leq \cdots \leq \mu_N(x_N|x_N),$$

the solution $q^\star(\cdot)$ to (A.15) *can be characterized as follows: there is $U^\star = U^\star(x_s, x) \in [\min_{j=1,\ldots,N}\{c_j - h_s - b_j\}, 0]$ and index $n^\star$ such that*

1. *$\mu_i(x_i + q_i^\star(x_s, x)|x_i) = U^\star$ for those i such that $q_i^\star(x_s, x) > 0, \;\; i \leq n^\star$.*

2. *$\mu_j(x_j|x_j) \geq U^\star$ for those j such that $q_j^\star(x_s, x) = 0$, $i > n^\star$.*

3. *$\sum_{i=1}^N q_i^\star(x_s, x) \leq x_s$ and $U^\star \cdot \left(\sum_{i=1}^N q_i^\star(x_s, x) - x_s\right) = 0$.*

Proof. Let

$$V_i(y|x_i) := (c_i - h_s)y + C_i(y, 0) + (h_i + b_i)\frac{F_i(x_i)}{f_i(x_i)}G_i(y). \tag{A.18}$$

The objective function in problem (A.15) can be written as $\sum_{i=1}^{N} V_i(x_i + q_i|x_i)$. Straightforward computation reveals that $\frac{d}{dy}V_i(y|x_i) = \mu_i(y|x_i)$ defined in (A.17). Lemma

A.5 establishes that $V_i(\cdot|x_i)$ is quasi-convex over the whole line, reach its minimum at $y_i^\star(x_i)$ and convex over $[-\infty, y_i^\star(x_i)]$. Thus it entails that $q_i^*(x_s, x) = 0$ for all those i such that $\mu_i(x_i|x_i) \geq 0$. For those i such that $\mu_i(x_i|x_i) < 0$, KKT conditions are necessary and sufficient to characterize the solution. Let $U^\star = U^\star(x_s, x)$ be the Lagrangian multiplier for the capacity constraint $\sum_{i=1}^N q_i \leq x_s$. Thus we have $\mu_i(x_i + q_i^\star(x_s, x)|x_i) = U^\star$ for those i such that $q_i^\star(x_s, x) > 0$ and $\mu_j(x_j|x_j) \geq U^\star$ for those j such that $q_j^\star(x_s, x) = 0$.

If for all $j = 1, \ldots, N$, $q_j^\star(x_s, x) = 0$, set $U^\star = \min\{\mu_j(x_j|x_j) : j = 1, \ldots, N\} \leq 0$ and hence $U^\star \geq \min\{-(h_s + b_j - c_j) : j = 1, \ldots, N\}$. If, for some i, $q_i^\star(x_s, x) > 0$, we have $U^\star = \mu_i(x_i + q_i^\star(x_s, x)|x_i) \geq -(h_s + b_i - c_i) \geq \min\{-(h_s + b_j - c_j) : j = 1, \ldots, N\}$ and $U^\star = \mu_i(x_i + q_i^\star(x_s, x)|x_i) \leq 0$ because $x_i + q_i^\star(x_s, x) \in [-\infty, y_i^\star(x_i)]$, over which $\mu_i(\cdot|x_i) \leq 0$ by quasi-convexity (Lemma A.5). The complementarity condition follows from the fact that if $\sum_{i=1}^N q_i^\star(x_s, x) < x_s$, then we must have $\mu_i(x_i + q_i^\star(x_s, x)|x_i) \geq \mu_i(y_i^\star(x_i)|x_i) = 0$ and it suffices to set $U^\star = 0$.

Finally, since $\mu_i(x_i|x_i)$ are increasing, $n^\star = \max\{i \geq 0 : q_i^\star(x_s, x) > 0\}$.

□

Lemma A.8. *Under Assumptions 2.1 and 2.2, let $q^\star$ be the solution to* (A.15) *characterized in Lemma A.7. Then for $i = 1, \ldots, N$ we have that $q_i^\star(x_s, x)$ is non-increasing in x_i. Moreover, $q_j^\star(x_s, x)$ is non-decreasing in x_i for any $j \neq i$.*

Proof. For fixed x_{-i} and $x_i \leq x_i'$ consider $q = q^\star(x_s, x_i, x_{-i})$ and $q' = q^\star(x_s, x_i', x_{-i})$. We can assume $q_i' > 0$ otherwise we are done. Moreover, since $\mu_i(x_i|x_i) \leq \mu_i(x_i'|x_i')$, we have that $q_i' > 0$ implies that $q_i > 0$. By Lemma A.7, q and q' are characterized by the existence of multipliers U and U', and integers n and n' such that

$$\begin{aligned} j \leq n, j \neq i,\ \mu_j(x_j + q_j|x_j) = U, q_j > 0 \qquad & j \leq n', j \neq i,\ \mu_j(x_j + q_j'|x_j) = U', q_j' > 0 \\ \mu_i(x_i + q_i|x_i) = U \qquad & \mu_i(x_i' + q_i'|x_i') = U' \\ j > n, j \neq i,\ \mu_j(x_j|x_j) \geq U, q_j = 0 \qquad & j > n', j \neq i,\ \mu_j(x_j|x_j) \geq U', q_j' = 0 \\ U \cdot \left(\sum_{j=1}^N q_j - x_s\right) = 0 \qquad & U' \cdot \left(\sum_{j=1}^N q_j' - x_s\right) = 0. \end{aligned}$$

Since $\mu_i(y|x_i) \leq \mu_i(y|x_i')$ for all y, it follows that $U' \geq U$. Moreover, note that $x_j + q_j \leq y_j^*(x_j)$ if $j \leq n$ and $x_j + q_j' \leq y_j^*(x_j)$ if $j \leq n'$. Therefore, since $\mu_j(\cdot|x_j)$ is non-decreasing for $y \in [0, y_j^*(x_j)]$ by Lemma A.5 (i), it follows that $q_j' \geq q_j$ for $j \leq \max\{n, n'\}$, $j \neq i$.

Thus, note that if $U < 0$ the conditions above imply $\sum_{i=1}^N q_i = x_s$ and we have

$$q_i' \leq x_s - \sum_{j \neq i} q_j' \leq x_s - \sum_{j \neq i} q_j = q_i$$

and the result follows. On the other hand, if $U = 0$, we also have $U' = 0$ and $q_i' = \max\{0, y_i^*(x_i') - x_i'\} \leq \max\{0, y_i^*(x_i) - x_i\} = q_i$ since $y_i^*(x_i) - x_i$ is non-increasing in x_i by Lemma A.6. $\square$

Appendix B

Proofs in Chapter 3

B.1 Auxiliary Technical Results

Lemma B.1. *The inequalities in (C) are equivalent to the inequalities in (B).*

Proof. Using our notation, Proposition 3.1 in Border (1991) states: *Let $X : V \to [0,1]$ be measurable. Then X is implementable by a symmetric auction if and only if for each measurable set $A \subset T$, the following inequality is satisfied:*

$$\int_A X(v)\, dF(v) \leq \frac{1 - \left(\int_{V \setminus A} dF(v)\right)^N}{N}.$$

By "X is implementable" Border means that there exists a symmetric function $x = \left(x^1, ..., x^N\right) : V^N \to [0,1]^N$ such that

$$X\left(v^i\right) \equiv \int_{V^{N-1}} x^i\left(v^i, v^{-i}\right) \prod_{k \in N \setminus \{i\}} dF\left(v^k\right).$$

The equivalence between (C) and (B) follows immediately from Border's Proposition 3.1, by letting $x^i(v) = \sum_{j \in J} q^i_j(v)$ and $X(v) = \sum_{j \in J} Q_j(v)$. □

Lemma B.2. *Let h be a Lipschitz function on V, with constant L, and let*

$$\hat{h}(v) = \{h(\hat{v}) : \text{ for some } \hat{v} \text{ such that } \|v - \hat{v}\| \leq \varepsilon\}$$

be a piecewise constant function. For any measurable function $g : V \to \mathbb{R}$ we have that

$$\int_D \left| h(v) - \hat{h}(v) \right| g(v) dv \leq L\varepsilon \int_D |g(v)| dv \quad \text{for any } D \subset V.$$

Proof. Since h is Lipschitz, by definition of $\hat{h}$ we have that $|h(v) - \hat{h}(v)| \leq L\varepsilon$. The result follows by integration. □

Lemma B.3. *For some $T > 0$, consider the grid V_T defined by $\epsilon = O(1/T)$. Let $\widetilde{U}^T$ be defined as in (3.1). For any $v \in V$ we have that*

$$0 \leq \widetilde{U}^T(v) - \widetilde{U}^T(v') \leq \epsilon, \quad v' \in V_T, \; v' \leq v < v' + \epsilon e.$$

Proof. Note that $\widetilde{U}^T$ is a convex function. By its definition, any subgradient of $\widetilde{U}^T$ belongs to the set $\text{conv}\{Q^T(v) : v \in V_T\}$. It follows that any subgradient $s \in \text{conv}\{Q^T(v) : v \in V_T\}$ is such that $s \geq 0$ and

$$\begin{aligned} \|s\|_1 &\leq \max_{\alpha_v \geq 0, \sum_{v \in V_T} \alpha_v = 1} \left\| \sum_{v \in V_T} \alpha_v Q^T(v) \right\|_1 \\ &\leq \sum_{v \in V_T} \alpha_v \sup_{v \in V_T} \|Q^T(v)\|_1 \leq 1. \end{aligned}$$

Thus,

$$\begin{aligned} \widetilde{U}^T(v) - \widetilde{U}^T(v') &\leq \sup_{\alpha \in [0,1], s \in \partial U(\alpha v + (1-\alpha)v')} |\langle s, v - v' \rangle| \\ &\leq \sup_{\alpha \in [0,1], s \in \partial U(\alpha v + (1-\alpha)v')} \|s\|_1 \|v - v'\|_\infty \leq \epsilon. \end{aligned}$$

Moreover, because $s \geq 0$ and $v - v' \geq 0$, we have

$$\widetilde{U}^T(v) \geq \widetilde{U}^T(v') + \langle s, v - v' \rangle \geq \widetilde{U}^T(v').$$

□

Lemma B.4. *Under Assumption 3.1 we have that*

$$\sum_{v\in V_T}\left|\hat{f}(v)-\int_{v\le v'\le v+\epsilon e} f(v')dv'\right| \le \epsilon 2\sqrt{J}\sup_{v'\in V}\|\nabla f(v')\|\mathrm{vol}(V)$$

Proof. For convenience let $C_v = \{v' \in V : v \le v' < v+\epsilon e\}$ where $\epsilon = O(1/T)$. By the mean value theorem, $\int_{C_v} f(v')dv' = f(\tilde{v})\epsilon^J$ where $\tilde{v} \in C_v$. The triangular inequality yields

$$\begin{aligned}\sum_{v\in V_T}\left|\hat{f}(v)-\int_{C_v} f(v')dv'\right| &\le \sum_{v\in V_T}\left(|\hat{f}(v)-f(v)\epsilon^J| + \left|\epsilon^J f(v) - \int_{C_v} f(v')dv'\right|\right)\\ &\le \sum_{v\in V_T}\left(|\hat{f}(v)-f(v)\epsilon^J| + \epsilon^J\,|f(v)-f(\tilde{v})|\right)\end{aligned}$$

Note that for the first term we have

$$\begin{aligned}\sum_{v\in V_T}|\hat{f}(v)-f(v)\epsilon^J| &= \sum_{v\in V_T}\frac{f(v)}{\sum_{v'\in V_T} f(v')}\left|1-\epsilon^J\sum_{v'\in V_T} f(v')\right| = \left|1-\sum_{v'\in V_T} f(v')\epsilon^J\right|\\ &= \left|\sum_{v'\in V_T} f(\tilde{v}')\epsilon^J - \sum_{v'\in V_T} f(v')\epsilon^J\right| \le \sum_{v'\in V_T}\epsilon^J|f(\tilde{v}')-f(v')|\end{aligned}$$

Finally, we bound

$$\sum_{v\in V_T}\epsilon^J\,|f(v)-f(\tilde{v})| \le \sum_{v\in V_T}\epsilon^J\sup_{v'\in C_v}|\langle\nabla f(v'), v-\tilde{v}\rangle| \le \epsilon\sqrt{J}\sup_{v\in V}\|\nabla f(v)\|\mathrm{vol}(V).$$

□

B.2 Proofs of Section 3.3

Proof of Theorem 3.1. Within this proof, we omit the superscript T for notational convenience. First note that $\widetilde{U}$ is convex (since it is the maximum of affine functions) and non-negative in V by construction because $\widetilde{U}(v) \ge U(\underline{v}) = 0$.

To show (ii), consider any two points $v, v' \in V$ and denote by $\hat{v}, \hat{v}' \in V_T$ the points that $\hat{v} \leq v < \epsilon e$, and $\hat{v}' \leq v' < \hat{v}' + \epsilon e$. Next recall that all (IIC) are satisfied for all $\hat{v}, \hat{v}' \in V_T$. Because $\widetilde{Q}(v) = Q(\hat{v})$, $\widetilde{U}(\hat{v}) = U(\hat{v})$, $\widetilde{U}(\hat{v}') = U(\hat{v}')$, we have

$$\begin{aligned}
&\widetilde{U}(v') - \widetilde{U}(v) - \left\langle \widetilde{Q}(v), v' - v \right\rangle \\
&= \widetilde{U}(v') - U(\hat{v}') + U(\hat{v}) - \widetilde{U}(v) + U(\hat{v}') - U(\hat{v}) - \langle Q(\hat{v}), v' - v\rangle \\
&\geq -\epsilon + \langle Q(\hat{v}), v' - \hat{v}' + \hat{v} - v\rangle + U(\hat{v}') - U(\hat{v}) - \langle Q(\hat{v}), \hat{v}' - \hat{v}\rangle \\
&\geq -\epsilon - \|Q(\hat{v})\|_1 \|\hat{v} - v\|_\infty \geq -2\epsilon,
\end{aligned}$$

by Lemma B.3, $\widetilde{U}(v') \geq \widetilde{U}(\hat{v}')$, $\langle Q(\hat{v}), v' - \hat{v}'\rangle \geq 0$, and $\|Q(\hat{v})\|_1 \leq 1$. Thus $\widetilde{Q}(v) \in \partial_{2\epsilon}\widetilde{U}(v)$.

Since violations to (IIC) we bounded by (ii) to show (iii) we need to control the violation among all Border constraints. By Lemma 3.3 it suffices to consider sets of the form $\widetilde{E}_\alpha := \{v \in V : \sum_{j=1}^J \widetilde{Q}_j^T(v) \geq \alpha\}$. The definition of $\widetilde{Q}$ implies that these sets are of the form unions of cubes of side ϵ (and therefore volume ϵ^J). Let $A = \cup_{v \in A_T \subset V_T}\{v' \in V : v \leq v' < v + \epsilon e\}$, then we have

$$\begin{aligned}
\left| \int_A \sum_{j\in J} \widetilde{Q}(v) dF(v) - \sum_{v \in A_T}\sum_{j \in J} Q(v)\hat{f}(v)\right| &\leq \sum_{v\in A_T}\sum_{j\in J} Q(v) \left| \int_{v \leq v' < v + \epsilon e} f(v')dv' - \hat{f}(v)\right| \\
&\leq \sum_{v \in V_T} \left| \int_{v \leq v' < v+\epsilon e} f(v')dv' - \hat{f}(v)\right| \\
&\leq \epsilon 2\sqrt{J} \sup_{v\in V} \|\nabla f(v')\| \mathrm{vol}(V)
\end{aligned}$$

by Lemma B.4. A similar bound applies to the right-hand-side of the Border constraint. Therefore the maximum Border violation is bounded by $O(\epsilon) = O(1/T)$.

□

Proof of Lemma 3.2. Let $\alpha = \int_A dF(v)$. First note that $N \int_A \sum_{j\in J} Q_j(v) dF(v) \leq N\alpha$. This implies that $N\alpha \geq 1 - (1-\alpha)^N + \eta$.

Using

$$(1-\alpha)^N \leq e^{-N\alpha} \leq 1 - N\alpha + \frac{N^2\alpha^2}{2} \quad \text{for } \alpha \in (0,1)$$

we have that

$$N\alpha \geq 1 - 1 + N\alpha - \frac{N^2\alpha^2}{2} + \eta = N\alpha - \frac{N^2\alpha^2}{2} + \eta.$$

This implies $\alpha \geq \frac{\sqrt{2\eta}}{N}$. □

Proof of Corollary 3.1. Consider an arbitrary violated Border constraint such that

$$N \int_A \sum_{j\in J} Q_j(v)dF(v) = 1 - \left(1 - \int_A dF(v)\right)^N + \eta$$

where we have $\eta \leq \delta(Q,U)$ by definition of δ. For notational convenience let $\alpha = \int_A dF(v)$ denote the probability measure of A.

Next consider scaling the mappings (Q,U) by

$$\frac{1-(1-\alpha)^N}{1-(1-\alpha)^N+\eta} = \frac{1}{1+\eta/(1-(1-\alpha)^N)} \geq 1 - \frac{\eta}{1-(1-\alpha)^N}$$

which would make it satisfy border for this particular set A.

By Lemma 3.2 we have that $\alpha \geq \sqrt{2\eta}/N$ which implies that

$$1-(1-\alpha)^N \geq N\alpha - \frac{N^2\alpha^2}{2} \geq \sqrt{2\eta} - \eta.$$

This yields a bound on the scaling

$$\frac{1-(1-\alpha)^N}{1-(1-\alpha)^N+\eta} \geq 1 - \frac{\eta}{\sqrt{2\eta}-\eta} = 1 - \frac{\sqrt{\eta}}{\sqrt{2}-\sqrt{\eta}}.$$

Since this last quantity is monotone in $\eta \leq \delta(Q,U)$ it suffices to scale (Q,U) by

$$t := 1 - \frac{\sqrt{\delta(Q,U)}}{\sqrt{2} - \sqrt{\delta(Q,U)}}$$

so that (tQ, tU) satisfies all Border constraints.

Regarding (IIC) if for every v, v' in the domain we have

$$U(v) - U(v') \geq \langle Q(v'), v - v' \rangle - \delta(Q, U),$$

since $t \geq 0$ it holds that

$$tU(v) - tU(v') \geq \langle tQ(v'), v - v' \rangle - t\delta(Q, U).$$

Finally, for (IIR) note that if $U(\underline{v}) \geq 0$ we have $tU(\underline{v}) \geq 0$. On the other hand if $U(\underline{v}) < 0$, multiplying by t reduces the violation as claimed.

Therefore, $\delta(tQ, tU) \leq t\delta(Q, U)$ and all Border constraint are satisfied by (tQ, tU).

$\square$

Proof of Theorem 3.2. Let $(\hat{Q}^*, \hat{U}^*)$ denote the restriction of (Q^*, U^*) to V_T and $(\tilde{Q}^*, \tilde{U}^*)$ the extension of $(\hat{Q}^*, \hat{U}^*)$ to V (as in (3.1)). First note that Lemma 3.2 and Corollary 3.1 can be applied to $(\hat{Q}^*, \hat{U}^*)$ in order to obtain a feasible solution for (P_T). Let $t = 1 - \frac{\sqrt{\delta_T(\hat{Q}^*, \hat{U}^*)}}{\sqrt{2} - \sqrt{\delta_T(\hat{Q}^*, \hat{U}^*)}}$ so that $(t\hat{Q}^*, t\hat{U}^*)$ is feasible for (P_T) since all (IIC)s are already satisfied.

In order to bound $\delta_T(\hat{Q}^*, \hat{U}^*)$, note that by convexity for any $h \geq 0$ we have

$$\frac{U^*(v) - U^*(v - he)}{h} \leq \sum_{j \in J} Q_j^*(v) \leq \frac{U^*(v + he) - U^*(v)}{h} \quad \text{for all} \quad v \in V,$$

where e denotes the vector of all ones. Next we apply Lemma 3.3 to restrict attention to the class of Border constraints with sets E_α. Without loss of generality we can restrict E_α to be a compact set.

Once again by convexity we have that $\langle Q^*(\hat{v}) - Q^*(v), \hat{v} - v \rangle \geq 0$. Applying this to $\hat{v} = v + he$ for $h \geq 0$ we have that $\sum_{j \in J} Q_j^*(v) \leq \sum_{j \in J} Q_j^*(v + he)$. Therefore, if $v \in E_\alpha$ we have that $v + he \in E_\alpha$ for any $h \geq 0$ such that $v + he \in V$.

Define $\mathcal{L}(E_\alpha) = \{v \in E_\alpha : \nexists h > 0 \text{ such that } v - he \in E_\alpha\}$. For each $v \in \mathcal{L}(E_\alpha)$ let $h(v) = \max_{h \geq 0}\{h : v + he \in V\}$. Therefore we can rewrite Border constraint as

$$
\begin{aligned}
& \textstyle\int_{E_\alpha} \sum_{j\in J} Q_j^*(v) f(v) dv \\
= & \int_{\mathcal{L}(E_\alpha)} \int_0^{h(v)} \sum_{j\in J} Q_j^*(v + he) f(v + hv) dh \; dv \\
= & \int_{\mathcal{L}(E_\alpha)} \left[U^*(v + he) f(v + he)|_0^{h(v)} - \int_0^{h(v)} U^*(v + he) e' \nabla f(v + he) dh \right] dv \\
= & \int_{\mathcal{L}(E_\alpha)} \left[\widetilde{U}^*(v + he) f(v + he)\Big|_0^{h(v)} - \int_0^{h(v)} \widetilde{U}^*(v + he) e' \nabla f(v + he) dh \right] dv + O(1/T)
\end{aligned}
$$

by Lemma B.2 since U^* is Lipschitz, V is compact, and f has bounded first and second derivatives. Convergence is uniform over all sets E_α since U is Lipschitz (constant at most 1), V is compact, , by using Lemma B.2. Combining this result with Lemma B.4 yields the rate of convergence $\delta_T(Q^*, U^*) = O(1/T)$ and

$$
\begin{aligned}
OPT_T & \geq \sum_{v\in V_T} \left(\left\langle t\widehat{Q}^*(v), v \right\rangle - t\widehat{U}^*(v) \right) \hat{f}(v) \\
& = \textstyle\int_V \left\langle t\widetilde{Q}^*(v), v \right\rangle - t\widetilde{U}^*(v) \; dF(v) + O(t/T) \\
& = t \int_V \langle tQ^*(v), v\rangle - tU^*(v) \; dF(v) + O(t/T) \\
& = tOPT_* + O(t/T).
\end{aligned}
$$

□

Proof of Corollary 3.2. Theorem 3.1 showed that $(\widetilde{Q}^T, \widetilde{U}^T)$ is asymptotically feasible. Now we <u>claim</u> that the objective value $OPT_*(\widetilde{Q}^T, \widetilde{U}^T)$ corresponding to $(\widetilde{Q}^T, \widetilde{U}^T)$ can be approximated by the objective value corresponding to $OPT_T(Q^T, U^T)$ corresponding to (Q^T, U^T). Then by Theorem 3.2, $OPT_*(\widetilde{Q}^T, \widetilde{U}^T)$ is asymptotically approximating the true optimal OPT_*. Thus, the two conditions in Definition 3.1 are satisfied by $(\widetilde{Q}^T, \widetilde{U}^T)$ and we obtain the corollary.

Now we are going to prove the claim. The objective function in the continuum is given by

$$\int_V \left[\langle Q(v), w(v)\rangle - U(v)\right] f(v)dv. \tag{B.1}$$

First we have

$$\begin{aligned}
&\left|\int_V \tilde{U}^T(v)f(v) - \sum_{v'\in V_T} U^T(v')\hat{f}(v')\right| \\
\leq\ & \left|\sum_{v\in V_T}\int_{v\leq v'\leq v+\epsilon e}\left(\tilde{U}^T(v') - U^T(v)\right) f(v')dv'\right| + \bar{U}\sum_{v\in V_T}\left|\hat{f}(v) - \int_{v\leq v'\leq v+\epsilon e} f(v')dv'\right| \\
\leq\ & \epsilon\sum_{v\in V_T}\int_{v\leq v'\leq v+\epsilon e} f(v')dv' + \epsilon 2\sqrt{J}\sup_{v'\in V}\|\nabla f(v')\|\text{vol}(V) \quad \text{(by Lemma B.3 and B.4)} \\
\leq\ & (2\sqrt{J}\sup_{v'\in V}\|\nabla f(v')\| + 1)\text{vol}(V)\epsilon
\end{aligned}$$

Let $\bar{W} = \max_{v\in V}\sum_{j\in J} w(v) < \infty$ by Assumption 3.1. Then we have

$$\begin{aligned}
&\left|\int_V \langle\tilde{Q}^T(v'), w(v')\rangle f(v')dv' - \sum_{v\in V_T}\langle Q^T(v), w(v)\rangle\hat{f}(v)\right| \\
\leq\ & \left|\sum_{v\in V_T}\int_{v\leq v'\leq v+\epsilon e}\left(\langle\tilde{Q}^T(v'), v'\rangle - \langle Q^T(v), v\rangle\right) f(v')dv'\right| \\
& +\bar{W}\sum_{v\in V_T}\left|\hat{f}(v) - \int_{v\leq v'\leq v+\epsilon e} f(v')dv'\right| \quad \text{(because)} \\
\leq\ & (2\sqrt{J}\sup_{v'\in V}\|\nabla f(v')\| + 1)\text{vol}(V)\epsilon \quad \text{(by Lemma B.4)}
\end{aligned}$$

Comparing the above two inequalities with (B.1) immediately yields our claim.

□

B.3 Proofs of Section 3.4

Proof of Lemma 3.3. We will show that for any measurable set $A \subseteq V$ there is a set of the form E_α whose violation of Border is at least as large the violation associated with A. Let α be the largest value such that

$$\int_{E_\alpha} dF(v) \geq \int_A dF(v)$$

which exists since the mapping $\alpha \mapsto \int_{E_\alpha} dF(v)$ is continuous from the left. We have that

$$\begin{aligned}
\int_A \left[\sum_{j \in J} Q_j(v)\right] dF(v) &= \int_{A \setminus E_\alpha} \left[\sum_{j \in J} Q_j(v)\right] dF(v) + \int_{A \cap E_\alpha} \left[\sum_{j \in J} Q_j(v)\right] dF(v) \\
&\leq \int_{A \setminus E_\alpha} \alpha dF(v) + \int_{A \cap E_\alpha} \left[\sum_{j \in J} Q_j(v)\right] dF(v) \\
&= \alpha \int_{A \setminus E_\alpha} dF(v) + \int_{A \cap E_\alpha} \left[\sum_{j \in J} Q_j(v)\right] dF(v).
\end{aligned}$$

Note that if we can take $\int_{E_\alpha} dF(v) = \int_A dF(v)$, we also have $\int_{E_\alpha \setminus A} dF(v) = \int_{A \setminus E_\alpha} dF(v)$ and E_α leads to a larger violation than A (since the respective right-hand sides are equal). Therefore we can assume $\int_{E_\alpha} dF(v) > \int_A dF(v)$. In turn this implies that

$$H_\alpha = \left\{ v \in V : \sum_{j \in J} Q_j(v) = \alpha \right\}$$

has positive measure with respect to F, that is

$$\int_{H_\alpha} dF(v) > \int_{E_\alpha} dF(v) - \int_A dF(v) > 0.$$

(Otherwise we would be able to choose a larger α.)

Consider the violation of the Border constraint associated with A

$$\delta_A := \int_A \left[\sum_{j \in J} Q_j(v) \right] dF(v) - 1 + \left(1 - \int_A dF(v) \right)^N \quad \text{and}$$

$$\delta(\mu) := \alpha\mu + \int_{E_\alpha \setminus H_\alpha} \left[\sum_{j \in J} Q_j(v) \right] dF(v) - 1 + \left(1 - \int_{E_\alpha \setminus H_\alpha} dF(v) - \mu \right)^N$$

where $\mu = \int_H dF(v)$ with $H \subset H_\alpha$. The latter is a convex function in the scalar $\mu \in [0, \int_{H_\alpha} dF(v)]$ so it is maximized at one of the extremes (i.e., the extremes have a violation at least as big as δ_A). Note that there might be no set H associated with intermediate values of μ (say $\mu = \int_A dF(v) - \int_{E_\alpha \setminus H_\alpha} dF(v)$) but the maximum is achieved in the extreme. This allows us to obtain at least the same violation as δ_A either with $H = \emptyset$ or $H = H_\alpha$. These cases correspond to set of the form E_α. □

Proof of Lemma 3.4. Regarding the separation oracle for (IIC) note that we can simply enumerate them. Each of the $O(T^{2J})$ constraint requires $O(J)$ operations.

Next we turn to Border constraints. Let (Q, U) be the point to be considered by the oracle. First construct the $O(T^J)$-dimensional vector $\overline{Q}$ defined as

$$\overline{Q}(v) = \sum_{j \in J} Q_j(v)$$

which accounts for $O(JT^J)$ operations. Then we sort the vector $\overline{Q}$ in descending order using $O(T^J \ln(T^J))$ operations to create $\widehat{Q}$. Next note that we can generate all the $O(T^J)$ sets E_α associated with Q by considering the following sets of components

$$\{\{1\}, \{1, 2\}, \{1, 2, 3\}, \ldots, \{1, 2, \ldots, O(T^J)\}\}$$

of $\widehat{Q}$. Since this can be done incrementally, the computational cost to search over all these sets is $O(T^J)$. □

Proof of Lemma 3.3. Let S^k denotes the set of indices of included Border and IC constraints at the k-th iteration. Since there are a finite number of constraints, there is a finite number of possible sets S^k each of which containing a finite number of extreme points or faces. Therefore there is only a finite number of different objective function values that can be achieved at any iteration. Since inequalities are discarded only if an improvement on the objective function is observed, either we keep accumulating inequalities or a improvement in the objective function is made. $\square$

Appendix C

Proofs in Chapter 4

C.1 Proof of Lemma 4.1

Proof. We show the following directions:

(1)⇔(2) First notice that $C - Q$ is a diagonally dominant M-matrix, and hence has inverse-positivity property, i.e. $(C-Q)^{-1}$ exists and has nonnegative entries (Seneta, 1981, Exercise 2.4). Let $\lambda_{\max}$ and $\mu_{\max}$ be the eigenvalue of $(C-Q)^{-1}BX_0$ (non-negative matrix) and $BX_0 - C + Q$ (ML-matrix) respectively, which have the largest real part among all the eigenvalues. Then by (Seneta, 1981, Theorem 2.6), $\lambda_{\max}$ and $\mu_{\max}$ are both real. Hence, by definition 4.1,

$$\mathcal{R}(a) = \lambda_{\max}.$$

Suppose $\mu_{\max} \leq 0$. Let $\mathbf{v}$ be the associated positive right eigenvector, then $[BX_0 - C + Q]\mathbf{v} = \mu_{\max}\mathbf{v}$. Hence, we have

$$BX_0\mathbf{v} = (C-Q)\mathbf{v} + \mu_{\max}\mathbf{v} \preceq (C-Q)\mathbf{v}. \tag{C.1}$$

Due to the positivity of $(C-Q)^{-1}$, we can multiply $(C-Q)^{-1}$ on the both sides of (C.1), and get

$$(C-Q)^{-1}BX_0\mathbf{v} \preceq \mathbf{v},$$

i.e. $\mathbf{v}$ is a subinvariant for $(C-Q)^{-1}BX_0$. Then by (Seneta, 1981, Theorem 1.6), $\lambda_{\max} \leq 1$, which, by definition, means $\mathcal{R}(a) \leq 1$.

For the inverse, suppose $\lambda_{\max} \leq 1$. First notice that $C(C-Q)^{-1}$ is the transition probability matrix and hence $(C-Q)^{-1}BX_0$ is primitive. By (Seneta, 1981, Theorem 1.1), let $\mathbf{v}$ be the associated positive right eigenvector, i.e.

$$(C-Q)^{-1}BX_0\mathbf{v} = \mathbf{v}.$$

Or equivalently,

$$[BX_0 - C + Q]\mathbf{v} = 0,$$

which implies, by (Seneta, 1981, Theorem 2.6), $\mu_{\max} \leq 0$.

(2)⇔(3) Notice that $G(a)$ is so-called ML-matrix or quasi-monotone matrix. A necessary and sufficient condition for an ML-matrix to have the spectral radius less than 0 (and hence the real part of all the eigenvalues) is that all the principal minors of $G(a)$ is positive. The closure of the all such 'a' is exactly the set $\mathcal{A}(\bar{u}_1, \bar{u}_2)$. (Gantmacher, 2005; Seneta, 1981).

□

C.2 Proof of Proposition 4.1.

Proof. Conditions (4.5) and (4.6) can be further reduced to (just by rearranging the terms)

$$a \leq -m_1 + \bar{u}_1, \tag{C.2}$$

$$a \geq m_2 - \bar{u}_2, \tag{C.3}$$

whose consistency requires

$$m_1 + m_2 - \bar{u}_1 - \bar{u}_2 \leq 0. \tag{C.4}$$

After some algebra, condition (4.7) can be converted to

$$-a^2 + (m_2 - m_1 - \bar{u}_2 + \bar{u}_1)a + (m_1 - \bar{u}_1)(m_2 - \bar{u}_2) - q \geq 0. \tag{C.5}$$

The discriminant of the quadratic form on the left-hand-side of (C.5) is given by

$$\Delta = (m_1 + m_2 - \bar{u}_1 - \bar{u}_2)^2 - 4q. \tag{C.6}$$

Due to (C.4), we only need to consider the following two possibilities:

- if $-2\sqrt{q} < m_1 + m_2 - \bar{u}_1 - \bar{u}_2 \leq 0$, then $\Delta < 0$ and hence (C.5) is infeasible;
- if $m_1 + m_2 - \bar{u}_1 - \bar{u}_2 \leq -2\sqrt{q}$, the $\Delta \geq 0$ and hence the feasible a's are between the two roots (may coincide) of the quadratic form (C.5):

$$\begin{aligned} s_1 &:= \frac{1}{2}(m_2 - m_1 - \bar{u}_2 + \bar{u}_1) - \frac{1}{2}\sqrt{\Delta} \\ s_2 &:= \frac{1}{2}(m_2 - m_1 - \bar{u}_2 + \bar{u}_1) + \frac{1}{2}\sqrt{\Delta} \end{aligned}$$

Notice that $[s_1, s_2] \subset [m_2 - \bar{u}_2, -m_1 + \bar{u}_1]$ and (C.2), (C.3) and (C.5) are consistent if $m_1 + m_2 - \bar{u}_1 - \bar{u}_2 \leq -2\sqrt{q} < 0$ or equivalently,

$$\bar{u}_1 + \bar{u}_2 \geq m_1 + m_2 + 2\sqrt{q}. \tag{C.7}$$

On the other hand, in order for $[-\bar{u}_2, \bar{u}_1] \cap [s_1, s_2] \neq \emptyset$, we need

$$\bar{u}_1 \geq s_1 \tag{C.8}$$

$$-\bar{u}_2 \leq s_2 \tag{C.9}$$

Therefore, if $\bar{u}_1$ and $\bar{u}_2$ satisfies the conditions (C.7), (C.8) amd (C.9), then $\mathcal{A}(\bar{u}_1, \bar{u}_2) \neq \emptyset$ and (4.10) holds. Now we show that the (C.7), (C.8) amd (C.9) are precisely characterized by (4.9) and (4.11).

(C.8) can be converted to

$$\sqrt{(m_1 + m_2 - \bar{u}_1 - \bar{u}_2)^2 - 4q} \geq m_2 - m_1 - \bar{u}_1 - \bar{u}_2 \tag{C.10}$$

If $m_1 \geq -\sqrt{q}$**,** $m_1+m_2+2\sqrt{q} \geq m_2-m_1$, which implies (C.7) $\Rightarrow m_2-m_1-\bar{u}_1-\bar{u}_2 \leq 0$ $\Rightarrow$ (C.10). Thus, in this case, we only need to consider conditions (C.7) and (C.9).

If $m_1 < -\sqrt{q}$**,** $m_1 + m_2 + 2\sqrt{q} < m_2 - m_1$. If still $\bar{u}_1 + \bar{u}_2 \geq m_2 - m_1$, then $m_2 - m_1 - \bar{u}_1 - \bar{u}_2 \leq 0 \Rightarrow$ (C.10), in which case conditions (C.7) and (C.8) are substituted by $\bar{u}_1 + \bar{u}_2 \geq m_2 - m_1$; if $m_1 + m_2 + 2\sqrt{q} \leq \bar{u}_1 + \bar{u}_2 < m_2 - m_1$, then squaring both sides of (C.10) (and rearranging terms) yields $\bar{u}_1 + \bar{u}_2 \geq m_2 - \frac{q}{m_1} > m_1 + m_2 + 2\sqrt{q}$, in which case conditions (C.7) and (C.8) are substituted by $\bar{u}_1+\bar{u}_2 \geq m_2-\frac{q}{m_1}$. By noticing $m_2-m_1 > m_2-\frac{q}{m_1}$ in this case, we obtain that $\bar{u}_1 + \bar{u}_2 \geq m_2 - \frac{q}{m_1}$ covers (C.7) and (C.8) when $m_1 < -\sqrt{q}$.

By symmetry, we can similarly deal with the condition (C.9). Putting them together, we obtain the proposition. ☐

C.3 Proof of Proposition 4.2

Proof. First, from (4.3) and the initial boundary condition $z_i(0) = 0$, we can get

$$z_i(\infty) = \gamma_i \int_0^\infty y_i(t)dt \tag{C.11}$$

Notice that (4.1) can be reformulated as

$$d \ln x_i = -\beta_i y_i dt. \tag{C.12}$$

Solve it and plug in (C.11), we will get:

$$x_i(\infty) = x_i(0)e^{-\beta_i/\gamma_i z_i(\infty)}. \tag{C.13}$$

For fixed i, summing up (4.1), (4.2) and (4.3) yield

$$\frac{d}{dt}(x_i + y_i + z_i) = \sum_{k=1}^{2} q_{ki} y_k. \tag{C.14}$$

Integrating (C.14), we get

$$[x_i(\infty) + y_i(\infty) + z_i(\infty)] - [x_i(0) + y_i(0) + z_i(0)] = \sum_{j=1}^{2} q_{ji} \int_0^\infty y_j dt, \tag{C.15}$$

where we substitute $x_i(0) = \bar{n}_i - u_i$, $y_i(0) = \epsilon_i$,$z_i(0) = 0$, $y_i(\infty) = 0$, $x_i(\infty) = x_i(0)e^{-\beta_i/\gamma_i z_i(\infty)}$ (C.13) and $\int_0^\infty y_j dt = \frac{z_i(\infty)}{\gamma_i}$ by (C.11). For each $i = 1, 2$, (C.15) turns out to be (4.14) and (4.15) respectively, i.e. $z_1(a; \varepsilon)$ and $z_2(a; \varepsilon)$ must satisfy the system (4.14) and (4.15). In addition, (C.11) implies that $z_i(\infty) > 0$.

Now we claim that the system (4.14) and (4.15) has exactly one positive solution provided that $(0, 0) \neq (\epsilon_1, \epsilon_2) \in \mathbb{R}_+^2$. As a result, the final sizes $z_1(x; \varepsilon) > 0$ and $z_2(x; \varepsilon) > 0$ are completely characterized by the system (4.14) and (4.15).

Existence Substituting z_2 in (4.15) with (4.14) immediately implies that $z_1(a; \varepsilon)$ is the zeros of the following function:

$$g(z_1) := h_2(h_1(z_1; \epsilon_1, a); \epsilon_2, a) - z_1. \tag{C.16}$$

Obviously, function $h_i(\cdot; \epsilon_i, a)$ and $g(\cdot)$ are continuous on $\mathbb{R}$. Now, notice that

$$h_1(0; \epsilon_1, a) = -\frac{\gamma_2}{q_{21}} \epsilon_1 \leq 0, \qquad h_1(+\infty; \epsilon_1, a) = +\infty.$$

The Intermediate Value Theorem immediately implies that there exists a $z_1^\# \geq 0$ such that $h_1(z_1^\#) = 0$, where $z_1^\# = 0$ iff $\epsilon_1 = 0$. Notice again, since $(\epsilon_1, \epsilon_2) \neq (0, 0)$

$$g(z_1^\#) = -\frac{\gamma_1}{q_{12}} \epsilon_2 - z_1^\# < 0 \qquad g(+\infty) = +\infty.$$

Again, by the Intermediate Value Theorem, there exists a $z_1^* > z_1^\# \geq 0$ such that $g(z_1^*) = 0$, i.e. the zeros of function $g(\cdot)$ exists in $(0, +\infty)$, so does $z_1(a; \varepsilon) \in (0, +\infty)$. Exactly similarly, we can prove the existence of $z_2(a; \varepsilon) \in (0, +\infty)$.

Uniqueness We prove this by contradiction. Suppose that there are two distinct positive solutions to the system (4.14) and (4.15), say $\mathbf{z}^{(1)} = (z_1^{(1)}, z_2^{(1)})$ and $\mathbf{z}^{(2)} = (z_1^{(2)}, z_2^{(2)})$. Now consider the line passing through these two points ($\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$) in the z_1-z_2 plane, namely

$$\ell_{\mathbf{z}^{(1)}\mathbf{z}^{(2)}} := \{(z_1, z_2) \in \mathbb{R}^2 : (z_2^{(2)} - z_2^{(1)})(z_1 - z_1^{(1)}) - (z_1^{(2)} - z_1^{(1)})(z_2 - z_2^{(1)}) = 0\}.$$

Without loss of generality, assume $z_1^{(1)} > z_1^{(2)}$. We first demonstrate that $z_2^{(1)} > z_2^{(2)}$. In fact, by the strict convexity of $h_1(z_1; \epsilon_1, a)$ and the fact that $h_1(0; \epsilon_1, a) < 0$ and $h_1(+\infty; \epsilon_1, a) = +\infty$, there is a single crossing point $z_1^\#$ such that $h_1(z_1^\#; \epsilon_1, a) = 0$ and $h_1(z_1; \epsilon_1, a) < 0$ for $z_1 \in [0, z_1^\#)$ and $h_1(z_1; \epsilon_1, a) > 0$, $h_1'(z_1; \epsilon_1, a) > 0$ for $z_1 \in (z_1^\#, \infty)$. Since we are consider positive solutions, it must be the case that $z_1^{(1)} > z_1^{(2)} > z_1^\#$, i.e. the solutions must lie on the (strictly) increasing part of $h_1(z_1; \epsilon_1, a)$. Therefore, we immediately get

$$z_2^{(1)} = h_1(z_1^{(1)}; \epsilon_1, a) > h_1(z_1^{(2)}; \epsilon_1, a) = z_2^{(2)}.$$

Consider the following two cases:

1. If $\vartheta := (z_2^{(2)} - z_2^{(1)})(0 - z_1^{(1)}) - (z_1^{(2)} - z_1^{(1)})(0 - z_2^{(1)}) \leq 0$, then by the strict convexity of $h_1(z_1; \epsilon_1, a)$, we have

$$\begin{aligned} h_1(0; \epsilon_1, a) &> \text{the value of } z_2\text{-coordinate of the line } \ell_{\mathbf{z}^{(1)}\mathbf{z}^{(2)}} \text{ when } z_1 = 0 \\ &= \frac{\vartheta}{z_1^{(2)} - z_1^{(1)}} \geq 0, \end{aligned}$$

which is a contradiction since $h_1(0; \epsilon_1, a) = -\frac{\gamma_2}{q_{21}}\epsilon_1 \leq 0$.

2. If $\vartheta := (z_2^{(2)} - z_2^{(1)})(0 - z_1^{(1)}) - (z_1^{(2)} - z_1^{(1)})(0 - z_2^{(1)}) > 0$, consider the function $h_2(z_2; \epsilon_2, a)$ on $[0, +\infty)$. Then, by the strict convexity of $h_2(z_2; \epsilon_2, a)$, we have

$$\begin{aligned} h_2(0; \epsilon_2, a) &> \text{the value of } z_1\text{-coordinate of the line } \ell_{\mathbf{z}^{(1)}\mathbf{z}^{(2)}} \text{ when } z_2 = 0 \\ &= \frac{\vartheta}{z_2^{(1)} - z_2^{(2)}} > 0, \end{aligned}$$

which is a contradiction since $h_2(0; \epsilon_2, a) = -\frac{\gamma_1}{q_{12}} \epsilon_2 \leq 0$.

To sum up, we must have a unique positive solution to the system (4.14) and (4.15).

□

C.4 Proof of Lemma 4.2

Proof. Our proof consists of the following parts [1]:

$a \in \mathcal{A}(\bar{u}_1, \bar{u}_2) \Rightarrow$ **the unique solution to** (4.14) **and** (4.15)**:** It is very easy to verify that $(0, 0)$ is a solution to (4.14) and (4.15) when $\epsilon_1 = \epsilon_2 = 0$. Now we just need to demonstrate its uniqueness. Since $h_1(z_1)$ is a strictly convex function in z_1, hence the points $(z_1^{(1)}, z_2^{(1)})$ such that $z_2^{(1)} = h_1(z_1^{(1)})$ must satisfy

$$z_2^{(1)} \geq h_1'(0) z_1^{(1)}, \quad \text{and "=" holds when } (z_1^{(1)}, z_2^{(1)}) = (0, 0), \tag{C.17}$$

where $h_1'(0) = \frac{\gamma_2}{q_{21}}[(1 + \frac{q_{12}}{\gamma_1}) - \frac{\beta_1}{\gamma_1}(n_1 + a)]$. Similarly, if the points $(z_1^{(2)}, z_2^{(2)})$ are such that $z_1^{(2)} = h_1(z_2^{(2)})$

$$z_1^{(2)} \geq h_2'(0) z_2^{(2)}, \quad \text{and "=" holds when } (z_1^{(1)}, z_2^{(1)}) = (0, 0), \tag{C.18}$$

[1] In the following arguments, we abbreviate $h_i(z_i; \epsilon_i, a)$ as $h_i(z_i)$ in the absence of confusion for any given ε and 'a'.

where $h_2'(0) = \frac{\gamma_1}{q_{12}}[(1 + \frac{q_{21}}{\gamma_2}) - \frac{\beta_2}{\gamma_2}(n_2 - a)]$.

Now notice that if $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$, $h_1'(0) > 0$, $h_2'(0) > 0$ and $h_1'(0) \geq \frac{1}{h_2'(0)}$. Let k be such that $h_1'(0) \geq k \geq \frac{1}{h_2'(0)}$, then, by (C.17) and (C.18), we have $z_2^{(1)} \geq k z_1^{(1)}$ with equality at $(0,0)$ and $z_2^{(2)} \leq k z_1^{(2)}$ with equality only at $(0,0)$. Hence, $(0,0)$ is the unique intersection point in this case.

$a \in \mathcal{A}^C(\bar{u}_1, \bar{u}_2) \Rightarrow$ **two solutions to** (4.14) **and** (4.15)**:** $(0,0)$ and $(z_1^0(a), z_2^0(a))$ with $z_i^0(a) > 0$. Clearly, $(0,0)$ is a solution, and therefore, we just need to show the existence and uniqueness of the other solution $(z_1^0(a), z_2^0(a))$ with $z_i^0(a) > 0$. Let $g(z_1)$ be as defined in (C.16).

We first demonstrate the existence. If $a \in \mathcal{A}^C(\bar{u}_1, \bar{u}_2)$, then one of the conditions (4.5), (4.6) and (4.7) is violated. By symmetry, we just need to consider the following two cases:

If (4.7) **is violated:** When $\epsilon_1 = \epsilon_2 = 0$, we have

$$\begin{aligned} h_1(0) &= 0, h_1(+\infty) = +\infty, \\ h_1'(0) &= \frac{\gamma_2}{q_{21}}[1 + \frac{q_{12}}{\gamma_1} - \frac{\beta_1}{\gamma_1}(n_1 + a)] \\ g(0) &= 0, g(+\infty) = +\infty, \\ g'(0) &= \frac{1}{q_{21}\gamma_1}[(\gamma_1 + q_{12} - \beta_1(n_1 + a))(\gamma_2 + q_{21} - \beta_2(n_2 - a)) - q_{12}q_{21}] < 0, \end{aligned}$$

which implies $g(\delta) < 0$ for some small enough $\delta > 0$. Therefore, the Intermediate Value Theorem, implies a zero point $z_1^0(a) > \delta > 0$ of $g(z_1)$. Similarly, $z_2^0(a) > 0$.

If (4.5) **is violated** then we have

$$h_1(0) = 0, h_1(+\infty) = +\infty, \quad h_1'(0) = \frac{\gamma_2}{q_{21}}[1 + \frac{q_{12}}{\gamma_1} - \frac{\beta_1}{\gamma_1}(n_1 + a)] < 0,$$

which, by the strict convexity of $h_1(z_1)$, implies that there exists a unique $z_1^{\#} > 0$ such that $h_1(z_1^{\#}) = 0$ (because the strictly convex function can at most have two zeros), $h_1(z_1) < 0$ for $z_1 \in (0, z_1^{\#})$ and $h_1(z_1) > 0$ for $z_1 > z_1^{\#}$. Therefore,

$$g(z_1^{\#}) = -\frac{q_{12}}{\gamma_1} z_1^{\#} < 0, g(+\infty) = +\infty,$$

which, by the Intermediate Value Theorem, implies that there a zero point $z_1^0(a) > z_1^{\#} > 0$ of $g(z_1)$, and hence $z_2^0(a) = h_1(z_1^0(a)) > 0$.

We now demonstrate the uniqueness. Notice that if (z_1^*, z_2^*) is a nonnegative solution of the system (4.14) and (4.15), then (z_1^*, z_2^*) must satisfy

$$g(z_1^*) = 0.$$

If (4.6) **holds,** $h_2'(0) \geq 0$ and hence $h_2(z_2)$ is increasing in z_2 and $g(z_1)$ is an increasing function for $z_1 > z_1^{\#}$. Thus, we can conclude that $(z_1^0(a), z_2^0(a))$ is the only solution to $g(z_1) = 0$ such that $z_i^0(a) > 0$.

If (4.6) **is violated,** similarly we can show that there exists a unique $z_2^{\#} > 0$ such that $h_2(z_2^{\#}) = 0$, $h_2(z_2) < 0$ for $z_2 \in (0, z_2^{\#})$ and $h_2(z_2) > 0$ for $z_2 > z_2^{\#}$. Now let $z_1^{\#\#}$ be the unique solution to $h_1(z_1) = z_2^{\#}$ by the convexity of $h_1(\cdot)$. The it is easy to see that $z_1^{\#\#} > z_1^{\#}$ and $h_1(z_1) > z_2^{\#}$ for $z_1 > z_1^{\#\#}$ and the solution $z_1^* > z_1^{\#\#}$. On the other hand, we can easily see that $g(z_1)$ is increasing for $z_1 > z_1^{\#\#}$. Thus, we can conclude that $(z_1^0(a), z_2^0(a))$ is the only solution to $g(z_1) = 0$ such that $z_i^0(a) > 0$.

□

C.5 Proof of Proposition 4.3

Proof. Our proof consists of the following two parts.

1. For $i = 1, 2$ and $\epsilon_i \geq 0$,

$$z_i(a; \epsilon_1, \epsilon_2) \geq z_i(a; \epsilon_1, 0) \geq z_i(a; 0, 0) ,$$

in which $z_i(a; \epsilon_1, \epsilon_2)$, $z_i(a; \epsilon_1, 0)$ and $z_i(a; 0, 0)$ are solutions to the system of final size equations (4.14)-(4.15) when ε takes values (ϵ_1, ϵ_2), $(\epsilon_1, 0)$ and $(0, 0)$, respectively.

We only demonstrate the first inequality $z_i(a, \epsilon_1, \epsilon_2) \geq z_i(a, \epsilon_1, 0)$ and the other one ($z_i(a, \epsilon_1, 0) \geq z_i(a, 0, 0)$) follows from the similar logic.

By definition, $(z_1(a; \epsilon_1, \epsilon_2), z_2(a; \epsilon_1, \epsilon_2))$ is the solution to the following system:

$$\begin{aligned} z_2 &= h_1(z_1; \epsilon_1, a) = -\frac{\gamma_2}{q_{21}}\epsilon_1 + h_1(z_1; 0, a) \\ z_1 &= h_2(z_2; \epsilon_2, a) = -\frac{\gamma_1}{q_{12}}\epsilon_2 + h_2(z_2; 0, a). \end{aligned}$$

And similarly, $(z_1(a; \epsilon_1, 0), z_2(a; \epsilon_1, 0))$ is the solution to the following system:

$$\begin{aligned} z_2 &= h_1(z_1; \epsilon_1, a) = -\frac{\gamma_2}{q_{21}}\epsilon_1 + h_1(z_1; 0, a) \\ z_1 &= h_2(z_2; 0, a). \end{aligned}$$

Thus if we define

$$\tilde{g}(z_1) := h_2(h_1(z_1; \epsilon_1, a); 0, a) - z_1,$$

then, according to (C.16),

$$g(z_1) = \tilde{g}(z_1) - \frac{\gamma_1}{q_{12}}\epsilon_2.$$

By definition, $z_1(a;\epsilon_1,0)$ is the solution of $\tilde{g}(z_1) = 0$ and $z_1(a;\epsilon_1,\epsilon_2)$ is the solution of $g(z_1) = 0$.

By Proposition 4.2 and Lemma 4.2, there are at most two nonnegative solutions to the equation $\tilde{g}(z_1) = 0$: 0 and/or a nonnegative number ($z_1(a;\epsilon_1,0)$), which may collapse into one. By the continuity of $\tilde{g}(\cdot)$, we must have $\tilde{g}(z_1) \le 0$ for $z_1 \in [0, z_1(a;\epsilon_1,0)]$ with $=$ holds at the end points. Now notice that $g(z_1)$ is a downward shifting of $\tilde{g}(z_1)$, which implies that the solutions to $g(z_1) = 0$ must lie outside of the interval $[0, z_1(a;\epsilon_1,0)]$. Thus, by the nonnegativity of $z_1(a;\epsilon_1,\epsilon_2)$ (Proposition 4.2) $z_1(a;\epsilon_1,\epsilon_2) \ge z_1(a;\epsilon_1,0)$, and the strict inequality holds if $\epsilon_2 > 0$.

Because the nonnegative intersection of $z_1 = h_2(z_2;\epsilon_2,a)$ and $z_2 = h_1(z_1;\epsilon_1,a)$ must lie on the increasing branch of $h_1(z_1;\epsilon_1,a)$, $h_1(z_1;\epsilon_1,a)$ is increasing in z_1 for all $z_1 \ge z_1(a;\epsilon_1,0)$. Hence,

$$z_2(a;\epsilon_1,\epsilon_2) = h_1(z_1(a;\epsilon_1,\epsilon_2);\epsilon_1,a) \ge h_1(z_1(a;\epsilon_1,0);\epsilon_1,a) = z_2(a;\epsilon_1,0)$$

by definition.

2. As $(\epsilon_1,\epsilon_2) \to (0,0)$,

$$z_i(a;\varepsilon) \to \begin{cases} 0 & : \text{ if } a \in \mathcal{A}(\bar{u}_1,\bar{u}_2), \\ z_i^0(a) & : \text{ if } a \in \mathcal{A}^C(\bar{u}_1,\bar{u}_2) \end{cases} \tag{C.19}$$

We first give some notations used in the following analysis. For any fixed 'a', (4.14) and (4.15) can be rewritten as:

$$\epsilon_1 = (1+\frac{q_{12}}{\gamma_1})z_1 - (n_1+a)[1-e^{-\frac{\beta_1}{\gamma_1}z_1}] - \frac{q_{21}}{\gamma_2}z_2 \tag{C.20}$$

$$\epsilon_2 = -\frac{q_{12}}{\gamma_1}z_1 + [1+\frac{q_{21}}{\gamma_2}]z_2 - (n_2-a)[1-e^{-\frac{\beta_2}{\gamma_2}z_2}], \tag{C.21}$$

which can be regarded as a map from $(z_1, z_2) \in \mathbb{R}^2$ to $(\epsilon_1, \epsilon_2) \in \mathbb{R}^2$. The Jacobian of this mapping is given by

$$
\begin{aligned}
J(z_1, z_2) &= \begin{pmatrix} \frac{\partial \epsilon_1}{\partial z_1} & \frac{\partial \epsilon_1}{\partial z_2} \\ \frac{\partial \epsilon_2}{\partial z_1} & \frac{\partial \epsilon_2}{\partial z_2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\gamma_1 + q_{12} - \beta_1 (n_1 + a) e^{-\frac{\beta_1}{\gamma_1} z_1}}{\gamma_1} & -\frac{q_{21}}{\gamma_2} \\ -\frac{q_{12}}{\gamma_1} & \frac{\gamma_2 + q_{21} - \beta_2 (n_2 - a) e^{-\frac{\beta_2}{\gamma_2} z_2}}{\gamma_2} \end{pmatrix}.
\end{aligned}
$$

Now, we consider the following cases:

(a) Fix $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$ with (4.7) being strict. At the *unique* intersection point $(0, 0)$, we have

$$
\det J(0, 0) = \frac{[\gamma_1 + q_{12} - \beta_1 (n_1 + a)][\gamma_2 + q_{21} - \beta_2 (n_2 - a)] - q_{12} q_{21}}{\gamma_1 \gamma_2} > 0.
$$

The Inverse Function Theorem implies that there is a neighborhood around $(0, 0)$ in the z_1-z_2 space such that the unique positive solution $z_i(a; \varepsilon)$ is smooth function of (ϵ_1, ϵ_2) for ε in a neighborhood of $(0, 0)$ in the ϵ_1-ϵ_2 space for $i = 1, 2$. Therefore, as $(\epsilon_1, \epsilon_2) \to (0, 0)$, $(z_1(a; \varepsilon), z_2(a; \varepsilon)) \to (0, 0)$.

(b) Fix $a \in \mathcal{A}^C(\bar{u}_1, \bar{u}_2)$ with (4.7) being strict. At the *unique* intersection point $(z_1^0(a), z_2^0(a))$, we claim that

$$
\det J(z_1^0(a), z_2^0(a)) = \frac{[\gamma_1 + q_{12} - \beta_1 (n_1 + a) e^{-\frac{\beta_1}{\gamma_1} z_1^0(a)}][\gamma_2 + q_{21} - \beta_2 (n_2 - a) e^{-\frac{\beta_2}{\gamma_2} z_2^0(a)}] - q_{12} q_{21}}{\gamma_1 \gamma_2} > 0.
$$

The Inverse Function Theorem implies that there is a neighborhood around (z_1^0, z_2^0) in the z_1-z_2 space such that the unique positive solution $z_i(a; \varepsilon)$ is smooth function of (ϵ_1, ϵ_2) for ε in a neighborhood of $(0, 0)$ in the ϵ_1-ϵ_2 space for $i = 1, 2$. Here notice that $(z_1(a; \varepsilon), z_2(a; \varepsilon))$ must be in the neighborhood of (z_1^0, z_2^0) rather than $(0,0)$ because we have show that $z_i(a;\varepsilon) \geq z_i^0$ Therefore, as $(\epsilon_1, \epsilon_2) \to (0,0)$, $(z_1(a;\varepsilon), z_2(a;\varepsilon)) \to (z_1^0, z_2^0)$. Now we show the above claim above. Let us define

$$g_0(z_1) := h_2(h_1(z_1; 0, a); 0, a) - z_1.$$

Then $z_1^0(a)$ is the positive solution of $g_0(z_1) = 0$ (0 is another solution) and, by the strict convexity of $g_0(z_1)$, we must have $g_0'(z_1^0) > 0$ and notice that $z_2^0 = h_1(z_1^0; 0, a)$, then

$$h_2'(z_2^0; 0, a) h_1'(z_1^0; 0, a) - 1 > 0,$$

which is, when written explicitly, exactly our claim.

(c) If $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$ with (4.7) being active, i.e.

$$h_1'(0; \epsilon_1, a) h_2'(0; \epsilon_2, a) = 1. \tag{C.22}$$

By definition, $(z_1(a;\varepsilon), z_2(a;\varepsilon))$ is the solution to the following system:

$$\begin{aligned} z_2 &= h_1(z_1; \epsilon_1, a) = -\frac{\gamma_2}{q_{21}}\epsilon_1 + h_1(z_1; 0, a) && \text{(C.23)} \\ z_1 &= h_2(z_2; \epsilon_2, a) = -\frac{\gamma_1}{q_{12}}\epsilon_2 + h_2(z_2; 0, a). && \text{(C.24)} \end{aligned}$$

Now consider an alternative system:

$$\begin{aligned} z_2 &= h_1(z_1; \epsilon_1, a) = -\frac{\gamma_2}{q_{21}}\epsilon_1 + h_1(z_1; 0, a) && \text{(C.25)} \\ z_1 &= -\frac{\gamma_1}{q_{12}}\epsilon_2 + \underbrace{h_2(0; 0, a)}_{=0} + h_2'(0; \epsilon_2, a) z_2. && \text{(C.26)} \end{aligned}$$

Denote the solution to this system as $(z_1^\#(a;\varepsilon), z_2^\#(a;\varepsilon))$. Then we have the following two claims:

Claim 1: $0 \le z_1(a;\varepsilon) \le z_1^\#(a;\varepsilon)$. It is obvious that $z_1^\#(a;\varepsilon)$ is the zero point of the following strict convex functions:

$$f_\varepsilon(z_1) := -\frac{\gamma_1}{q_{12}}\epsilon_2 - h_2'(0;\epsilon_2,a)\frac{\gamma_2}{q_{21}}\epsilon_1 + h_2'(0;\epsilon_2,a)h_1(z_1;0,a) - z_1;$$

while, due to the strict convexity of $h_2(z_2;\epsilon_2,a)$, we have $z_1(a;\varepsilon) = h_2(h_1(z_1(a;\varepsilon);\epsilon_1,a);\epsilon_2,a) \ge h_2(0;\epsilon_2,a) + h_2'(0;\epsilon_2,a)h_1(z_1(a;\varepsilon);\epsilon_1,a)$, i.e.

$$f_\varepsilon(z_1(a;\varepsilon)) < 0 = f_\varepsilon(z_1^\#(a;\varepsilon)). \tag{C.27}$$

On the other hand, the strict convexity of $h_1(z_1;0,a)$ implies for all $z_1 > 0$, we have

$$f_\varepsilon'(z_1) = h_2'(0;\epsilon_2,a)h_1'(z_1;\epsilon_1,a) - 1 > h_2'(0;\epsilon_2,a)h_1'(0;\epsilon_1,a) - 1 = 0,$$

by (C.22). That is to say $f_\varepsilon(z_1)$ is an increasing function. Therefore, (C.27) immediately implies our first claim.

Claim 2: $z_1^\#(a;\varepsilon) \to 0$ as $(\epsilon_1,\epsilon_2) \to (0,0)$.

It is very easy to see from the above analysis that $y = f_{(0,0)}(z_1)$ is a strictly convex function tangent to the horizontal axis $y = 0$ at $z_1 = 0$, and for $(\epsilon_1,\epsilon_2) \gneq (0,0)$, $f_\varepsilon(z_1)$ is a parallel downward shift of function $f_{(0,0)}$ with a positive solution (the larger intersection point of $y = f_\varepsilon(z_1)$ and $y = 0$), namely $z_1^\#(a;\varepsilon)$. Obviously, as $(\epsilon_1,\epsilon_2) \to (0,0)$, this positive solution $z_1^\#(a;\varepsilon)$ approaches to 0.

Those two claims together implies that, as $(\epsilon_1, \epsilon_2) \to (0,0)$, $z_1(a;\varepsilon) \to 0$. Exactly similar arguments can be applied to show that $z_2(a;\varepsilon) \to 0$ as $(\epsilon_1, \epsilon_2) \to (0,0)$.

□

C.6 Proof of Proposition 4.4

Proof. Denote $\tilde{\mathbf{y}}(t) = (\tilde{y}_1(t), \tilde{y}_2(t))'$, then (4.16) can be written in the following matrix form:

$$\frac{d\tilde{\mathbf{y}}}{dt} = G(a)^T \tilde{\mathbf{y}}, \tag{C.28}$$

where $\mathbf{y}(0) = (\epsilon_1, \epsilon_2)'$ and $G(a)$ is as defined in Lemma 4.1. Notice (C.28) is a homogenous linear system with constant coefficients. Hence, for the given initial condition, we have a unique closed form solution

$$\tilde{\mathbf{y}}(t) = \exp(tG(a)^T)\varepsilon$$

Integrating this expression yields

$$\begin{aligned} \int_0^\infty \tilde{\mathbf{y}}(t)dt &= (\int_0^\infty \exp(tG(a)^T)dt)\varepsilon \\ &= -G(a)^{-T}\varepsilon \end{aligned}$$

The last integration is because all the eigenvalues of $G(a)$ have negative real parts for 'a' in the interior of $\mathcal{A}(\bar{u}_1, \bar{u}_2)$ thanks to Lemma 4.1.

Hence, the definition of $\tilde{z}(a;\varepsilon)$ (4.18) implies

$$\begin{pmatrix} \tilde{z}_1(a;\varepsilon) \\ \tilde{z}_2(a;\varepsilon) \end{pmatrix} = \begin{pmatrix} \tilde{z}_1(\infty) \\ \tilde{z}_2(\infty) \end{pmatrix} = -\begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix} G(a)^{-T}\varepsilon$$

$$= -\begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix} \begin{pmatrix} \beta_1(n_1+a) - q_{12} - \gamma_1 & q_{21} \\ q_{12} & \beta_2(n_2-a) - q_{21} - \gamma_2 \end{pmatrix}^{-1} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

$$= -\frac{1}{\det G(a)} \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix} \begin{pmatrix} \beta_2(n_2-a) - q_{21} - \gamma_2 & -q_{21} \\ -q_{12} & \beta_1(n_1+a) - q_{12} - \gamma_1 \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{-\gamma_1(-\epsilon_1\beta_2 a + \epsilon_1\beta_2(m_2-\bar{u}_2) - \epsilon_2 q_{21})}{\beta_1\beta_2[(a+m_1-\bar{u}_1)(-a+m_2-\bar{u}_2) - q]} \\ \dfrac{-\gamma_2(\epsilon_2\beta_1 a + \epsilon_2\beta_1(m_1-\bar{u}_1) - \epsilon_1 q_{12})}{\beta_1\beta_2[(a+m_1-\bar{u}_1)(-a+m_2-\bar{u}_2) - q]} \end{pmatrix},$$

which yields the expressions (4.19) and (4.20).

Now we show (4.21). Through the proof of Proposition 4.3, we see that for any given 'a' in the interior of $\mathcal{A}(\bar{u}_1, \bar{u}_2)$, $z_i(a;\varepsilon)$ is a C^1 function of (ϵ_1, ϵ_2) in the neighborhood of $(\epsilon_1, \epsilon_2) = (0,0)$ and we also have $z_i(a;\mathbf{0}) = 0$. Hence, we can assume for any given 'a'

$$z_i(a;\varepsilon) = c_i\epsilon_1 + d_i\epsilon_2 + o(\varepsilon), \tag{C.29}$$

where c_i and d_i are to determine.

Substituting $z_i(a;\varepsilon)$ in (4.14) and (4.15) with (C.29), and expanding the exponential in its Taylor series yields

$$\begin{aligned} \frac{q_{21}}{\gamma_2}(c_2\epsilon_1 + d_2\epsilon_2 + o(\varepsilon)) &= -\epsilon_1 + [1 + \frac{q_{12}}{\gamma_1}](c_1\epsilon_1 + d_1\epsilon_2 + o(\varepsilon)) \\ &\quad -(n_1+a)\frac{\beta_1}{\gamma_1}(c_1\epsilon_1 + d_1\epsilon_2 + o(\varepsilon)) \qquad \text{(C.30)} \\ \frac{q_{12}}{\gamma_1}(c_1\epsilon_1 + d_1\epsilon_2 + o(\varepsilon)) &= -\epsilon_2 + [1 + \frac{q_{21}}{\gamma_2}](c_2\epsilon_1 + d_2\epsilon_2 + o(\varepsilon)) \\ &\quad -(n_2-a)\frac{\beta_2}{\gamma_2}(c_2\epsilon_1 + d_2\epsilon_2 + o(\varepsilon)). \qquad \text{(C.31)} \end{aligned}$$

Comparing the coefficients of ϵ_1 and ϵ_2 on both sides of the above equations, we obtain the linear system that c_i and d_i satisfy:

$$\begin{aligned}
[\frac{\beta_1}{\gamma_1}(n_1+a)-(1+\frac{q_{12}}{\gamma_1})]c_1+\frac{q_{21}}{\gamma_2}c_2 &= -1,\\
\frac{q_{12}}{\gamma_1}c_1+[\frac{\beta_2}{\gamma_2}(n_2-a)-(1+\frac{q_{21}}{\gamma_2})]c_2 &= 0;
\end{aligned}$$

$$\begin{aligned}
[\frac{\beta_1}{\gamma_1}(n_1+a)-(1+\frac{q_{12}}{\gamma_1})]d_1+\frac{q_{21}}{\gamma_2}d_2 &= 0,\\
\frac{q_{12}}{\gamma_1}d_1+[\frac{\beta_2}{\gamma_2}(n_2-a)-(1+\frac{q_{21}}{\gamma_2})]d_2 &= -1.
\end{aligned}$$

For $a \in \mathcal{A}(\bar{u}_1, \bar{u}_2)$, we can invoke Cramer's rule to solve the system and get

$$\begin{aligned}
c_1 &= \frac{-\gamma_1(-\beta_2 a+\beta_2(m_2-\bar{u}_2))}{\beta_1\beta_2[(a+m_1-\bar{u}_1)(-a+m_2-\bar{u}_2)-q]},\\
c_2 &= \frac{\gamma_2\epsilon_1 q_{12}}{\beta_1\beta_2[(a+m_1-\bar{u}_1)(-a+m_2-\bar{u}_2)-q]},\\
d_1 &= \frac{\gamma_1 q_{21}}{\beta_1\beta_2[(a+m_1-\bar{u}_1)(-a+m_2-\bar{u}_2)-q]},\\
d_2 &= \frac{-\gamma_2(\beta_1 a+\beta_1(m_1-\bar{u}_1))}{\beta_1\beta_2[(a+m_1-\bar{u}_1)(-a+m_2-\bar{u}_2)-q]}
\end{aligned}$$

Plugging the above expression of coefficients back to (C.29) and comparing the expressions (4.19) and (4.20), we obtain the desired results. □

C.7 Proof of Lemma 4.3

Proof. Since $\varepsilon \neq 0$, the positivity of $\tilde{z}_i(a;\varepsilon)$ immediately follows from Proposition 4.2. The convexity of $\tilde{z}_i(a;\varepsilon)$ can be directly read from their expressions (4.19) and (4.20). In fact, the denominators of (4.19) and (4.20) are quadratic forms in 'a' and the numerators are linear in 'a'. Hence, the global minimum of $\tilde{z}_i(a;\varepsilon)$ is achieved at a certain t_i, which can be calculated through the first order condition and we obtain:

$$t_1 = \begin{cases} \frac{1}{2}(m_2 - m_1 + \bar{u}_1 - \bar{u}_2) & : \text{ if } \epsilon_1 = 0, \epsilon_2 > 0 \\ (m_2 - \bar{u}_2) + \sqrt{q} & : \text{ if } \epsilon_1 > 0, \epsilon_2 = 0 \\ (m_2 - \bar{u}_2) - \frac{\epsilon_2 q_{21}}{\beta_2 \epsilon_1} & \\ +\sqrt{-\frac{\epsilon_2 q_{21}}{\beta_2\epsilon_1}(m_1 + m_2 - \bar{u}_2 - \bar{u}_1 - \frac{\epsilon_1 q_{12}}{\beta_1 \epsilon_2} - \frac{\epsilon_2 q_{21}}{\beta_2 \epsilon_1})} & : \text{ if } \epsilon_1 > 0, \epsilon_2 > 0 \end{cases},$$

$$t_2 = \begin{cases} -(m_1 - \bar{u}_1) - \sqrt{q} & : \text{ if } \epsilon_1 = 0, \epsilon_2 > 0 \\ \frac{1}{2}(m_2 - m_1 + \bar{u}_1 - \bar{u}_2) & : \text{ if } \epsilon_1 > 0, \epsilon_2 = 0 \\ -(m_1 - \bar{u}_1) + \frac{\epsilon_1 q_{12}}{\beta_1 \epsilon_2} & \\ -\sqrt{-\frac{\epsilon_1 q_{12}}{\beta_1\epsilon_2}(m_1 + m_2 - \bar{u}_2 - \bar{u}_1 - \frac{\epsilon_1 q_{12}}{\beta_1 \epsilon_2} - \frac{\epsilon_2 q_{21}}{\beta_2 \epsilon_1})} & : \text{ if } \epsilon_1 > 0, \epsilon_2 > 0 \end{cases}.$$

Now we show that $s_1 \le t_1 \le t_2 \le s_2$.We discuss the following cases:

If $\epsilon_1 = 0$ and $\epsilon_2 > 0$: $t_2 - t_1 = -\frac{1}{2}(m_1 - \bar{u}_1 + m_2 - \bar{u}_2) - \sqrt{q} \ge 0$, according to (C.7). And $t_1 - s_1 = \frac{1}{2}\sqrt{(m_1 + m_2 - \bar{u}_2 + \bar{u}_1)^2 - 4q} \ge 0$, obviously. Finally, $s_2 - t_2 = \frac{1}{2}[(m_1 - \bar{u}_1 + m_2 - \bar{u}_2) + 2\sqrt{q} + \sqrt{(m_1 + m_2 - \bar{u}_2 + \bar{u}_1)^2 - 4q}] \ge 0$, because $(m_1 + m_2 - \bar{u}_2 + \bar{u}_1)^2 - 4q - [(m_1 - \bar{u}_1 + m_2 - \bar{u}_2) + 2\sqrt{q}]^2 = 4\sqrt{q}[-(m_1 - \bar{u}_1 + m_2 - \bar{u}_2) - 2\sqrt{q}] \ge 0$, still by (C.7).

If $\epsilon_1 > 0$ and $\epsilon_2 = 0$: Similar to the above case by symmetry.

If $\epsilon_1 > 0$ and $\epsilon_2 > 0$: Let $\Psi := -(m_1 + m_2 - \bar{u}_2 - \bar{u}_1 - \frac{\epsilon_1 q_{12}}{\beta_1 \epsilon_2} - \frac{\epsilon_2 q_{21}}{\beta_2 \epsilon_1}) \ge 0$, then we have

$$t_2 - t_1 = \sqrt{\Psi}[\sqrt{\Psi} - (\sqrt{\frac{\epsilon_1 q_{12}}{\beta_1 \epsilon_2}} + \sqrt{\frac{\epsilon_2 q_{21}}{\beta_2 \epsilon_1}})] \ge 0,$$

because $\Psi - (\sqrt{\frac{\epsilon_1 q_{12}}{\beta_1 \epsilon_2}} + \sqrt{\frac{\epsilon_2 q_{21}}{\beta_2 \epsilon_1}})^2 = -(m_1 + m_2 - \bar{u}_2 - \bar{u}_1) - 2\sqrt{q} \ge 0$ still by (C.7).

In order to prove that $t_1 \ge s_1$, we denote $a = \frac{\epsilon_1 q_{12}}{\beta_1 \epsilon_2}$, $b = \frac{\epsilon_2 q_{21}}{\beta_2 \epsilon_1}$ and $\phi = m_1 + m_2 - \bar{u}_2 - \bar{u}_1$. Then, by (C.7), we have $\phi^2 \ge 4ab$. Under this notation, $t_1 - s_1 = \frac{1}{2}[\phi - 2b + 2\sqrt{b}\sqrt{a + b - \phi} + \sqrt{\phi^2 - 4ab}]$., where we claim $\phi - 2b +$

$2\sqrt{b}\sqrt{a+b-\phi} \geq 0$, and hence $t_1 - s_1 \geq 0$. In fact we have $\phi - 2b < 0$ and $(2\sqrt{b}\sqrt{a+b-\phi})^2 - (\phi - 2b)^2 = \phi^2 - 4ab \geq 0$. Exactly similarly, we can show that $s_2 \geq t_2$.

□

C.8 Proof of Theorem 4.2

Proof. By Lemma 4.3, $\tilde{z}_i(a;\varepsilon)$ is decreasing when $a \leq t_i$ and increasing when $a \geq t_i$ for $i = 1, 2$, and we also know that $s_1 \leq t_1 \leq t_2 \leq s_2$. Hence, the equilibrium depends on the order of the five critical values: t_1, t_2, $-\bar{u}_2$, 0 and $\bar{u}_1$. For example, if $-\bar{u}_2 < 0 < \bar{u}_1 \leq t_1 \leq t_2$, i.e. the feasible region is below t_1 where both countries' $\tilde{z}_i(a;\varepsilon)$ is decreasing in 'a'. Therefore, in order to minimize $\tilde{z}_i(a;\varepsilon)$, country 1 is willing to give up all its resources $\bar{u}_1$; while country 2 has no incentive to return any of these resources, i.e. she will accepts the offer. Obviously, such an allocation is also Pareto. Similarly, we can analysis all the combinations of the five critical values and summarize the corresponding PNE in the following table:

Table C.1: PNEs corresponding to different orders of the critical values.

Parameter Order Relations	Equilibrium a^*
$-\bar{u}_2 < 0 < \bar{u}_1 \leq t_1 \leq t_2$	$\bar{u}_1$
$-\bar{u}_2 < 0 \leq t_1 < \bar{u}_1 \leq t_2$	t_1
$-\bar{u}_2 < 0 \leq t_1 \leq t_2 < \bar{u}_1$	t_1
$-\bar{u}_2 \leq t_1 < 0 < \bar{u}_1 \leq t_2$	0
$t_1 \leq -\bar{u}_2 < 0 < \bar{u}_1 \leq t_2$	0
$-\bar{u}_2 \leq t_1 < 0 \leq t_2 < \bar{u}_1$	0
$t_1 \leq -\bar{u}_2 < 0 \leq t_2 < \bar{u}_1$	0
$-\bar{u}_2 \leq t_1 \leq t_2 \leq 0 < \bar{u}_1$	t_2
$t_1 \leq -\bar{u}_2 \leq t_2 \leq 0 < \bar{u}_1$	t_2
$t_1 \leq t_2 \leq -\bar{u}_2 < 0 < \bar{u}_1$	$-\bar{u}_2$

Hence, summarizing the PNEs in Table C.1 immediately yields the results. □

C.9 Proof of Theorem 4.3

Proof. Since both $\tilde{z}_1(a;\varepsilon)$ and $\tilde{z}_2(a;\varepsilon)$ are convex functions in 'a', $\tilde{z}_1(a;\varepsilon)+\tilde{z}_2(a;\varepsilon)$ is also a convex function in 'a'. Hence, the first order condition will yields that the minimum of $\tilde{z}_1(a;\varepsilon)+\tilde{z}_2(a;\varepsilon)$ is achieved at

$$t_c = \frac{\Phi+\sqrt{\Psi}}{\gamma_1\epsilon_1\beta_2-\epsilon_2\gamma_2\beta_1}, \tag{C.32}$$

where

$$\begin{aligned}
\Phi &= \gamma_1\epsilon_1\beta_2(m_2-\bar{u}_2)-\gamma_1\epsilon_2 q_{21}+\gamma_2\epsilon_2\beta_1(m_1-\bar{u}_1)-\gamma_2\epsilon_1 q_{12}\\
\Psi &= \Psi_1\Psi_2+{\gamma_1}^2{\epsilon_1}^2{\beta_2}^2 q_{12}q_{21}-2\,\gamma_1\epsilon_1\beta_2\gamma_2\epsilon_2\beta_1 q_{12}q_{21}+{\gamma_2}^2{\epsilon_2}^2{\beta_1}^2 q_{12}q_{21}\\
\Psi_1 &= -\gamma_1\epsilon_1\beta_2\,(m_1-\bar{u}_1+m_2-\bar{u}_2)+\gamma_1\epsilon_2 q_{21}+\gamma_2\epsilon_1 q_{12}\\
\Psi_2 &= -\gamma_2\epsilon_2\beta_1\,(m_1-\bar{u}_1+m_2-\bar{u}_2)+\gamma_1\epsilon_2 q_{21}+\gamma_2\epsilon_1 q_{12}.
\end{aligned}$$

The convexity of $\tilde{z}_1(a;\varepsilon)+\tilde{z}_2(a;\varepsilon)$ immediately implies that the CPOA is

$$a^\star = \begin{cases} -\bar{u}_2 & : \quad t_c < -\bar{u}_2 \\ t_c & : \quad -\bar{u}_2 \le t_c \le \bar{u}_1 \\ \bar{u}_1 & : \quad t_c > \bar{u}_1, \end{cases}$$

which can be interpreted as the implementation stated in the theorem.

The last thing we need to prove is that

$$t_1 \le t_c \le t_2.$$

In fact, suppose $t_c < t_1$. Then by the strict convexity of $\tilde{z}_i(a;\varepsilon)$ and $t_1 \le t_2$ (Lemma 4.3), we have $\tilde{z}_1'(t_c;\varepsilon) < 0$ and $\tilde{z}_2'(t_c;\varepsilon) < 0$, where the derivative is taken over 'a'. Then, $\tilde{z}_1'(t_c;\varepsilon)+\tilde{z}_2'(t_c;\varepsilon) < 0$, contradicting to $\tilde{z}_1'(t_c;\varepsilon)+\tilde{z}_2'(t_c;\varepsilon) = 0$. Therefore, $t_c < t_1$ does not hold. Similarly, we can disprove $t_c > t_2$. □

Bibliography

Allen, S. (1958), "Redistribution of Total Stock Over Several User Locations," *Naval Research Logistics Quarterly*, 5, 337–345.

Anderson, R. and May, R. (1991), *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, USA.

Anonymous (2005), "Preparing for an influenza pandemic," *The Economist*, Sep. 24, 95-96.

Armstrong, M. (1996), "Multiproduct nonlinear pricing," *Econometrica*, 64, 51–75.

Ball, F. (1991), "Dynamic Population Epidemic Models," *Math. Biosci.*, 107, 299–324.

Bauch, C. and Earn, D. (2004), "Vaccination and the Theory of Games," *Proc. Natl. Acad. Sci.*, 101, 13391–13394.

Bauch, C., Galvani, A., and Earn, D. (2003), "Group Interest versus Self-interest in Smallpox Vaccination Policy," *Proc. Natl. Acad. Sci.*, 100, 10564–10567.

Belloni, A., Lopomo, G., and Wang, S. (2010), "Multidimensional Mechanism Design: Finite-Dimensional Approximations and Efficient Computation," *Operations Research*, 58, 1079–1089.

Belloni, A., Lopomo, G., and Wang, S. (2011), "Does Private Information about Inventories Matter?" working paper.

Border, K. (1991), "Implementation of reduced form auctions: a geometric approach," *Econometrica*, 59, 1175–1187.

Brandeau, M. (2004), *Allocating Resources to Control Infectious Diseases*, chap. 17, pp. 443–464, Kluwer Academic Publishers.

Brandeau, M., Zaric, G., and Richter, A. (2003), "Resource Allocation for Control of Infectious Diseases in Multiple Independent Populations: beyond Cost-effectiveness Analysis," *J. Health Econ.*, 22, 575–598.

Brusco, S., Lopomo, G., and Marx, L. (2011), "The Economics of Contingent Re-auctions," *American Economic Journal: Microeconomics*, 3, 165–193.

Cachon, G. (2001), "Stock Wars: Inventory Competition in a Two-Echelon Supply Chain with Multiple Retailers," *Management Science*, 49, 658–674.

Cachon, G. (2003), *Supply Chain Coordination with Contracts*, chap. in Handbooks in Operations Research and Management Science: Supply Chain Management, North-Holland.

Cachon, G. (2004), "The Allocation of Inventory Risk in a Supply Chain: Push, Pull, and Advance-Purchase Discount Contracts," *Management Science*, 50, 222–238.

Cachon, G. and Fisher, M. (2000), "Supply Chain Inventory Management and the Value of Shared Information," *Management Science*, 46, 1032–1048.

Cachon, G. and Lariviere, M. A. (1999), "Capacity Choice and Allocation: Strategic Behavior and Supply Chain Performance," *Management Science*, 45, 1091–1108.

Cachon, G. and Lariviere, M. A. (2001), "Contracting to Assure Supply: How to Share Demand Forecasts in a Supply Chain," *Management Science*, 47, 629–646.

Cachon, G. and Lariviere, M. A. (2005), "Supply Chain coordination with revenue-Sharing Contract: Strengths and Limitations," *Management Science*, 51, 30–44.

Cachon, G. and Zhang, F. (2006), "Procuring Fast Delivery: Sole Sourcing with Information Asymmetry," *Management Science*, 52, 881–896.

Capasso, V. (1993), *Mathematical Structures of Epidemic Systems*, Springer-Verlag, Berlin, Heidelberg, New York.

Chen, F. (2007), "Auctioning Supply Contracts," *Management Science*, 53, 1562–1576.

Chen, F., Federgruen, A., and Zheng, Y.-S. (2001), "Coordination Mechanism for a Distribution System with One Supplier and Multiple Retailers," *Management Science*, 47, 693–708.

Corbett, C. (2001), "Stochastic Inventory Systems In A Supply Chain With Asymmetric Information: Cycle Stocks, Safety Stocks, and Consignment Stock," *Operations Research*, 49, 487–500.

Corbett, C. and de Groote, X. (2000), "A Supplier's Optimal Quantity Discount Policy Under Asymmetric Information," *Management Science*, 46, 444–450.

Daley, D. and Gani, J. (1994), *A Deterministic General Epidemic Model in a Stratified Population*, chap. 8, pp. 117–132, Probability, Statistics and Optimization: A Tribute to Peter Whittle, John Wiley & Sons ltd., Chichester.

Dasgupta, P., Hammond, P., and Maskin, E. (1979), “The Implementation of Social Choice Rules: some Results on Incentive Compatibility,” *Review of Economic Studies*, 46, 185–216.

Deshpande, V. and Schwarz, L. (2005), “Optimal Capacity Choice and Allocation in Decentralized Supply Chains,” working paper.

Diekmann, O. and Heesterbeek, J. (2000), *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, John Wiley & Sons, Chichester.

Fenner, F., Henderson, D., Arita, I., Jezek, Z., and Ladnyi, I. (1988), “Smallpox and its eradication,” *World Health Organization, Geniva.*

Gantmacher, F. (2005), *Applications of the Theory of Matrices*, Dover Publications, Mineola, N.Y.

Grötschel, M., Lovász, L., and Schrijver, A. (1981), *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, New York, NY.

Grötschel, M., Lovász, L., and Schrijver, A. (1988), “The ellipsoid method and its consequences in combinatorial optimization,” *Combinatorica*, 1, 169–197.

Hill, A. and Longini, I. (2003), “The critical vaccination fraction for heterogeneous epidemic models,” *Math. Biosci.*, 181, 85–106.

Iyengar, G. and Kumar, A. (2008), “Optimal Procurement Mechanisms for Divisible Goods with Capacitated Suppliers,” *Rev. Econ. Design*, 12, 129–154.

Kaplan, E., Craft, D., and Wein, L. (2003), “Analyzing bioterror response logistics: the case of smallpox,” *Math. Biosci.*, 185, 33–72.

Klein, E., Galvani, A., Malani, A., and Laxminarayan, R. (2007), “Impact of Incentives for Surveillance and Reporting on the Spread of an Influenza Epidemic,” DIMACS Workshop on ”Game Theoretic Approaches to Epidemiology and Ecology”.

Lau, A. H. L., Lau, H.-S., and Wang, J.-C. (2008), “How a dominant retailer might design a purchase contract for a newsvendor-type product with price-sensitive demand,” *Eur. J. Oper. Res.*, 190, 443–458.

Lee, H. and Whang, S. (1999), “Decentralized Multi-Echelon Supply Chains: Incentive and Information,” *Management Science*, 45, 633–640.

Lee, H. L. and Whang, S. (2000), “Information Sharing in a Supply Chain,” *International Journal of Technology Management*, 20, 373–387.

Mabrey, V. and Janik, K. (2009), "Zhu Zhu Pets: Hamsters to Save Christmas?" http://abcnews.go.com/Nightline/HolidayShopping/zhu-zhu-pets-holiday-seasons-toy/story?id=9137390.

Manelli, A. and Vincent, D. (2007), "Multidimensional Mechanism Design: Revenue Maximization and the Multiple-Good Monopoly," *Journal of Economic Theory*, 137, 153–185.

Maskin, E. and Riley, J. (1984), "Monopoly with Incomplete Information," *Rand Journal of Economics*, 15, 171–196.

Maskin, E. and Riley, J. (1986), "Optimal Auctions with Risk Averse Buyers," *E-conometrica*, 52, 1473–1518.

Maskin, E. and Tirole, J. (1990), "The principal-agent relationship with an informed principal, the case of private values," *Econometrica*, 58, 379–409.

May, R. and Anderson, R. (1984), "Spatial heterogeneity and the design of immunization programs." *Math. Biosci.*, 72, 83–111.

Milgrom, P. and Segal, I. (2002), "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

Monto, A. (2005), "The Threat of an Avian Influenza Pandemic," *New Engl. J. Med.*, 352, 323.

Mussa, M. and Rosen, S. (1978), "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.

Myerson, R. (1979), "Incentive-compatibility and the Bargaining Problem," *Econometrica*, 47, 61–73.

Myerson, R. (1981), "Optimal Auction Design," *Mathematics of Operation Research*, 6, 58–73.

Myerson, R. (1983), "Mechanism Design By an Informed Principle," *Econometrica*, 51, 1767–1797.

Mylovanov, T. and Tröger, T. (2008), "Optimal auction design and irrelevancy of private information," .

Padberg, M. W. and Rao, M. R. (1980), "The Russian method and integer programming," working paper.

Porteus, E. and Whang, S. (1991), "On Manufacturing/Marketing Incentives," *Management Science*, 37, 1166–1181.

Rochet, J. and Choné, P. (1998), "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66, 783–826.

Rockafellar, R. T. (1970), *Convex Analysis*, Princeton University Press, Princeton, NJ.

Rosenthal, J. (2000), *A First Look at Rigorous Probability Theory*, World Scientific.

Seneta, E. (1981), *Non-negative Matrices and Markov Chains*, Springer-Verlag, New York, Heidelberg, Berlin, 2nd edn.

Smith, C. (1964), "Factors in the transmission of virus infections from animals to man." *Sci. Basis Med. Annu. Rev.*, 125, 50.

Smolinski, M., Hamburg, M., and Lederberg, J. (2003), "Microbial threats to health: emergence, detection, and response," *Washington: Institute of Medicine*, p. 32.

Sun, P., Yang, L., and de Véricourt, F. (2009), "Selfish Drug Allocation for Containing an International Influenza Pandemic at the Onset," *Operations Research*, 57, 1320–1332.

Wang, S., de Véricourt, F., and Sun, P. (2009), "Decentralized Resource Allocation to Control an Epidemic: A Game Theretic Approach," *Mathematical Biosciences*, 222, 1–12.

Wernau, J. (2009), "Zhu Zhu Pets: Toy successes hard to predict, but robotic hamsters got a craze rolling," http://articles.chicagotribune.com/2009-12-07/news/0912070126_1_zhu-zhu-pets-toy-trend-hot-toy.

Wilson, R. (1993), *Nonlinear Pricing*, Oxford University Press.

Zhang, F. (2010), "Procurement Mechanism Design in a Two-Echelon Inventory System with Price-Sensitive Demand," *Manufacturing & Service Operations Management*, 12, 608–626.

Zhang, H., Nagarajan, M., and Sošić, G. (2009), "Dynamic Supplier Contracts under Asymmetric Inventory Information," forthcoming in Operations Research.

Zipkin, P. (1980), "Simple Ranking Methods For Allocation of One Resource," *Management Science*, 26, 34–43.

Zipkin, P. (2000), *Foundations of Inventory Management*, McGraw-Hill.

Biography

Shouqiang (Qiang) Wang was born on October 10, 1982 in Guiyang, the capital city of Guizhou province, P. R. China. In 2001, he was enrolled to the pre-med program of Peking Union Medical College and studied in the biology department at Peking University, Beijing, China. One year later, he transferred to School of Mathematical Sciences at Peking University and graduated from there in 2006 with a Bachelor of Science in Mathematics and a Bachelor of Art in Economics. In his junior year, he studied as a visiting student at the mathematics department of Uppsala University, Sweden. Right after his undergraduate study, he continued his graduate study in decision science group at the Fuqua School of Business, Duke University, Durham, USA. In 2011, he received his Master of science in Statistical Science from Department of Statistical Science at Duke as well. He is scheduled to complete his PhD in Business Administration in May of 2011 and will join the management department of Clemson University as an assistant professor, pending successful defense of this dissertation.