Capacity Management: Intra-Firm and Inter-Firm Perspectives

by

Anyan Qi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in The University of Michigan
2014

Doctoral Committee:

Associate Professor Hyun-Soo Ahn, Co-Chair
Associate Professor Amitabh Sinha, Co-Chair
Professor Xiuli Chao
Assistant Professor Xun Wu

To my mother and grandmother

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors and dissertation committee co-chairs, Professors Hyun-Soo Ahn and Amitabh Sinha. When I look back the numerous days and nights when we worked together, I feel extremely lucky to be able to work with Hyun-Soo and Amitabh. During my years in Michigan, they have provided me with sufficient academic guidance, have contributed significant amount of time, and have been tolerant about the mistakes I have made along my PhD journey. To me, they are not only my academic advisors, but also my life advisors, who help me grow up as a qualified researcher. Without them, the dissertation, which is build upon three co-authored research papers with my advisors, cannot be completed. I would also like to thank my dissertation committee members, Professors Xiuli Chao and Brian Wu. Both Xiuli and Brian have raised critical questions which helped me to significantly improve the quality of the dissertation.

Beyond my dissertation committee members, I would like to thank professors in Department of Technology and Operations for their generous input in my PhD study. Specifically, I would like to thank Professors Ravi Anupindi and Damian Beil for generously supporting me in my job hunting process, sharing knowledge with me during their research seminars, and giving me the opportunity to learn in their classroom. I also appreciate Professor Bill Lovejoy for his research seminars as well as the creative idea to provide me a valuable opportunity to work in the industry. I significantly benefited from these experiences as well as his sincere advice. I also learned many cutting-edge topics from research seminars provided by Professors Wally

Hopp, Roman Kapuscinski, Steve Leider, Peter Lenk, and Jun Li, which directly or indirectly inspired the dissertation. I am also appreciative about Professors Owen Wu and Steve Leider's tremendous help for my teaching.

The completion of the dissertation also significantly benefited from the help from my colleagues. The senior PhD students in my department have provided me with sufficient help in my study and life since my first day in the program. They are Min Choi, Suleyman Demirel, Sripad Devalkar, Bin Hu, Shanshan Hu, Zhixi Wan, Dadi Wang, Lai Wei, Yan Yin, and Man Yu. I sincerely appreciate their generous help. I have also benefited from academic discussions with PhD students in my year or younger, including Bam Amornpetchkul, Ruthy Beer, Yao Cui, and Santhosh Suresh. I also thank Hakjin Chung and Liang Ding for their input to the successful completion of our hospital project.

Finally, I would like to thank Mr. Brian Jones, Ms. Roberta Perry, Ms. Kelsey Zill at the PhD Program Office at Ross School of Business for their help. I would like to thank my mother, Ling Sun, and grandmother, Yuehua Sun, for their unconditional love to support my PhD study. I am also appreciative about my cousin Ji Li for her encouragement and support, as well as the joyful time I spent with my friends during my spare time.

TABLE OF CONTENTS

LIST OF FIGURES

<u>Figure</u>

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Capacity Management: Intra-Firm and Inter-Firm Perspectives

by

Anyan Qi

Co-Chairs: Hyun-Soo Ahn and Amitabh Sinha

Capacity management is challenging. Many decisions regarding capacity are made before full information is known, often requiring large and irrevocable expenditures. Moreover, the consequences of wrong capacity decisions critically affect the firm's bottom line. In recent years, the capacity decision has become of particular interest, reflecting two principal trends. First, advances in information technology that provide huge amounts of data about operations and demand offer firms potential to utilize this big data in making capacity decisions. Second, although supply chains today are highly decentralized with complex topologies, many buying firms and suppliers aim to maintain tight relationships with initiatives such as supplier development, among which capacity investment is an important strategic decision.

Corresponding to the two streams, we analyze a firm's capacity management decision, *how much capacity a firm should have and why*, at both intra-firm and inter-firm levels. At the intra-firm level, we investigate how a firm should learn demand information and leverage the information in capacity decisions. In Chapter II, we formulate a firm's capacity adjustment plan when the demand distribution is unknown as a s-tochastic dynamic program, and derive the optimal policy and date-driven heuristics.

At the inter-firm level, using game theory, we examine how a firm should manage its capacity at a shared supplier given two contractual constraints: *exclusive*, where other firms cannot access the leftover, and *first-priority*, where they can. In Chapter III where firms compete and the capacity cost consists of a fixed and a variable portion, we find that firms tend to invest more aggressively under the exclusive contract. Therefore, sometimes the firm may benefit from letting a competitor free-ride on the invested capacity. In Chapter IV where firms may or may not compete and the capacity cost has a variable portion, we characterize two equilibria: *a prisoner's dilemma*, where both firms choose the exclusive capacity which is not Pareto-optimal, and *a free-rider equilibrium*, where one firm chooses the first-priority capacity and allows the other with exclusive capacity to free ride. Both equilibria can be sustained when the firms serve independent markets, but not when they compete in a Cournot market.

CHAPTER I

Introduction

Capacity management, featuring high yet irrevocable expenditure, is a key operational decision that firms need to make. This decision is particularly acute in the modern business world with a highly uncertain environment and rather complex supply chain relationship. On one hand, an incorrect capacity decision at a firm's own site may lead to severe mismatch between supply and demand. On the other hand, a capacity decision at the supplier's site may play a strategic role in a complex supply chain. Therefore, it is challenging to decide how much capacity a buying firm should have. In the dissertation, we attempt to analyze a firm's capacity management decisions at both intra-firm and inter-firm levels.

In Chapter II, we analyze how a firm should leverage early demand data in its capacity decision at the intra-firm level. In a complex supply chain environment, the demand distribution may not be available as typically assumed in the literature and firms need to infer the true demand based on sales history. For example, Mahindra and Mahindra, the largest utility vehicle producer in India, launched a new car XUV 500 in September 2011 with an initial capacity of approximately 2,000 units per month. However, they severely underestimated the demand, and confirmed the order of over 15,000 vehicles in just two rounds of bookings. Finally, they doubled their capacity in May 2012, which was already 9 months after the launch of the product.

Clearly, the firm faces a trade-off between exploration and exploitation: While it can delay the decision to adjust capacity to explore the demand, it has to suffer from the potential lost demand and more expensive outsourcing cost. An interesting and important question therefore arises: When, and by how much, should a firm adjust its capacity?

To investigate this question, we formulate a firm's capacity adjustment plan within an environment of unknown demand distribution as a stochastic dynamic programming. When the firm has only a single adjustment opportunity, corresponding to cases with long capacity adjustment leadtime and high adjustment cost, we characterize the firm's optimal policy: while the target capacity level always increases in the likelihood of demand being high, the decision to adjust capacity is not necessarily monotone with respect to the likelihood. As the optimal policy is difficult to compute, we also derive a simple data-driven heuristic which only depends on the demand observations, and show the heuristic is asymptotically optimal with fast convergence rate. When the firm has multiple adjustment opportunities, corresponding to cases with short capacity adjustment leadtime and low adjustment cost, we show the optimal policy as a control band policy characterized by two switching curves, and derive a rather simple heuristic to compute the capacity decision. Finally, we use numerical study to analyze the performance of the heuristics.

In Chapter III and IV, we explore how firms should invest in or reserve the supplier's capacity when the supplier is shared with other firms, which becomes common in several industries such as electronics, cosmetics, and fashion. For instance, Apple's iPhone uses components made by more than a dozen suppliers and is assembled by a third-party manufacturer, while these suppliers and manufacturer also serve other products such as Samsung's Galaxy. In these relationships, how the leftover capacity can be used critically affects how much capacity firms should reserve or build as, if unspecified, it may be used to fill its competitor's orders. In fear of such free-riding, some firms put contractual constriants when reserving capacity. Neutrogena, a Johnson and Johnson company, invested in the capacity of their supplier who also serves Neutrogena's competitors such as L'Oreal and Estee Lauder. To avoid its supplier using the capacity to fill other orders, it claimed exclusive rights by placing counters on the invested machines. In another case, DowAgrosciences pays to reserve capacity at their key ingredients suppliers who also supply to its competitor, DuPont Agriculture. However, they only demand the first priority in utilizing the capacity, allowing other firms to access the leftover. Depending on different market environment, supply conditions, and capacity cost structures, firms and suppliers may have different preferences over the capacity types. Therefore, it is natural to ask: (Chapter III) When the capacity is random and the demand is deterministic, how should the competing buying firms invest in the shared supplier, and how should different stakeholders choose the capacity type? (Chapter IV) When the capacity is deterministic and the demand is random, how should buying firms endogenously choose these capacity constraints? How would firms' investment decisions be affected by whether the firms are competing with each other or not?

In Chapter III, we build game theoretic models, where two competing buying firms share a common supplier, and explore how firms invest in the supplier's capacity given the two contractual restraints: exclusive and first-priority, when the demand is deterministic, the capacity is random, and the capacity investment cost consists of a fixed and a variable portion. Our results suggest that firms tend to invest more aggressively under the exclusive contract, and therefore, sometimes the firm may benefit from letting a competitor free-ride on the invested capacity because allowing free-riding can reduce the competition intensity. Essentially, this chapter provides managerial insights that firms considering investing in suppliers who also supply their competitors must consider the consequences of their investment via the lens of a multi-player game, rather than myopically focusing on increases access to capacity.

In Chapter IV, we also build game-theoretic models where firms share a common supplier and examine how the relationship between the two firms affects the capacity investment decisions, when the demand is random, the capacity is deterministic, and the capacity investment cost is linear in the invested capacity size. In some cases, the firms are not directly competing in the market, for example, when they serve two geographically separated markets. In other cases, firms may directly compete against each other such as in a Cournot market. In these two market structures, demands for both firms are independent and positively-correlated respectively. For a given market structure, firms first choose the capacity types: to share (first-priority) or not to share (exclusivity). Then they build capacity at the supplier contingent upon the capacity types, and finally serve the market. Our analysis shows that even if choosing the first-priority is Pareto-optimal for both firms, buying firms tend to choose the exclusive capacity, driven by the incentive of capacity cost and the possibility to drive the other firm to build excess capacity. This provides a parsimonious explanation about the widely observed exclusive claim attached to the capacity investment to the supplier, as well as the managerial insights that doing so may trap the buying firms in a prisoner's dilemma. Interestingly, we also observe that a free-rider equilibrium can be sustained, where one firm chooses the first-priority capacity, builds a larger capacity, and allows the other firm with the exclusive capacity to free-ride on its invested capacity. This, in general, is driven by the capacity cost. These free-rider and prisoner's dilemma equilibria are observed in the independent market, but not in the Cournot market. This is because in the Cournot market, where the demands are positively correlated, firms are less likely to access the other firm's leftover capacity. Therefore, both firms simply choose the cheaper capacity and do not need to consider much about the benefit to pool the demand uncertainty under the first-priority capacity.

CHAPTER II

Capacity Investment with Demand Learning

2.1 Introduction

In most cases building capacity requires significant time and resource commitment, thus many firms need to make capacity decisions when there exist significant demand uncertainties. While early capacity installation enables a firm to seize a time-to-market opportunity, installing capacity with little market information may result in a significant mismatch when the capacity level is significantly different from the realized demand. Even if the firm realizes such a mismatch, changing the capacity level is often difficult and costly to the firm in both time and money. Increasing capacity level through adding new machines and/or hiring new workers is expensive and often irrevocable. Downsizing the capacity level, which typically requires layoffs and equipment divestment or salvage, can also be costly. In addition to the financial cost, changing the capacity level often requires a considerable amount of time. New machines or workers may take several weeks or months to be ready for production. More importantly, if capacity installation is delayed or insufficient, the firm will miss out market opportunities and significantly hit its bottom line. As choosing a “perfect” initial capacity level well before a selling season is a near-impossible task, many firms adjust its initial capacity level after observing some demand information in the early stage of a planning horizon. For this strategy to be successful, the firm should be able to evaluate the benefit and cost between two options– waiting it out (gathering more information) and committing to an action (adjusting the capacity level)– a classic trade-off between *exploration* and *exploitation*.

The set of problems that we consider is well illustrated in the following two examples. A major ODM (original design manufacturer) that the authors have intimate knowledge about serves a number of major cosmetics companies. Many of the products that the ODM produces are seasonal and sensitive to fashion trends, thus they have a short selling season of about three to four months. As there exist significant uncertainties about demand volume and type (e.g., which one of twenty different red shades will be popular?), it is impossible to stock the finished products in advance and therefore the ODM produces in a make-to-order environment. For some products, e.g., make-up compacts and eyeshadow pallets, the firm's capacity is bounded by the number of molds and fixtures designed for the specific products. When the demand of a particular product significantly surges beyond their existing capacity, the firm needs to either produce them with existing equipment in overtime or expand its capacity by procuring additional molds or fixtures. In this setting, not satisfying the order is not an option as the firm may lose a client. As it takes significant time to get new molds (typically two weeks to one month compared to the three to four months of a selling season), if the firm would like to increase its capacity level, it must do so very early in the selling season; otherwise, it will be too late to use the adjusted capacity in production.

In the automobile industry, Mahindra & Mahindra (M&M), the largest utility vehicle producer in India, launched the XUV 500 model in September 2011 with an initial capacity of about 2,000 vehicles per month. However, the company confirmed orders of over 15,000 vehicles in just two rounds of bookings. In fact, in the second round, over 25,000 booking applications were received, among which only 7,200 winners were chosen with a lottery. Finally, in June 2012, M&M announced a ramp-up of their capacity to about 4,000 vehicles per month (Thakkar, 2012). As the demand continues to grow, in January 2013, M&M announced another ramp-up of the capacity to 4,500 vehicles per month (Philip, 2013). In this setting, the product life cycle is long relative to the leadtime to adjust the capacity. Therefore, M&M has multiple opportunities to adjust capacity, as illustrated in the two adjustments.

Motivated by these observations, we examine a make-to-order firm's capacity decision using demand observation: ***when****, and by* ***how much****, should a firm adjust its capacity?* To investigate this question, we consider a firm selling a single product for a finite planning horizon when the firm has only partial information about random demand. In each period, the firm observes the realized demand and collects more information. Based on the information, the firm actively updates its knowledge about the demand, and uses this updated knowledge in the capacity adjustment decision. We consider two different stylized settings explicitly. In the first setting, the firm has only a single chance to adjust capacity (*a single-adjustment scenario*). This scenario is appropriate in settings where the leadtime for capacity investment (disinvestment) is long relative to a short planning horizon and/or the cost associated with capacity adjustment is significant as illustrated in our cosmetics example. In the second setting, the firm has multiple opportunities to change its capacity (*a multiple-adjustment scenario*). This scenario is appropriate when the leadtime for capacity adjustment is short and/or it is easy (or relatively inexpensive) to adjust the capacity level as illustrated in the M&M example. We specifically choose these settings as we will show that, the number of opportunities that the firm has to adjust its capacity critically affects the structure of the optimal policy and asymptotically optimal heuristics.

In both single-adjustment and multiple-adjustment scenarios, we first formulate the problem as a stochastic dynamic program and characterize the structure of the optimal policy. Then, for each scenario, we propose a data-driven heuristic policy that is not only implementable but also *asymptotically optimal* with an analytic performance bound. In the single adjustment scenario, we show that the optimal policy is counter-intuitive. In particular, we show whether to adjust capacity level or not in a given period is not monotone in the firm's posterior belief about demand. In particular, under the optimal policy, the firm may increase the capacity level when the likelihood of high demand is moderate, but switch to stay put and collect more demand observations when the likelihood of high demand becomes even higher. Thus, the firm's belief about a demand type does not monotonically affect the optimal capacity adjustment decision. In addition to the non-monotonicity, the stochastic control problem has a very large state-space as the firm's belief about demand type is our state variable. Consequently, solving and implementing the optimal policy quickly becomes computationally intractable even when there are only several possible demand types. To overcome this, we propose a two-step data-driven heuristic, which only depends on the firm's observed demand data. We prove that this heuristic is *asymptotically optimal* in the case where the true demand follows a stochastic process with stationary and independent increment in time, Specifically, we show that the regret (the percentage profit loss relative to an upper bound when the firm has complete demand information) converges to 0 rapidly as the problem scale increases. We then consider the multiple-adjustment scenario and show that the optimal policy is a control band policy, where in each period the firm will adjust the capacity up to a threshold if the capacity level is significantly low relative to the inferred demand, adjust the capacity down to another threshold if the capacity level is significantly high, and stay put in between. For this setting, we propose a different data-driven heuristic in which the firm adjusts capacity in exponentially increasing intervals and show that this policy is indeed asymptotically optimal under the regret criterion.

We illustrate the performance of our heuristics using a numerical study where some of the key parameters and data are derived from actual production and sales data of an automobile instead of using a randomized test bed in order to highlight the fact that our heuristic only requires real-time demand data and a few parameters that can be either inferred or collected by the firm. The numerical study demonstrates the value of using demand learning in capacity decision, and show that our heuristics is very robust with respect to problem parameters and assumptions.

The rest of the chapter is organized as follows. The related literature is reviewed in Section 2.2. The optimal policy of the stochastic dynamic program for the single-adjustment scenario is presented and discussed in Section 4.2. In Section 2.4, we propose a two-step heuristic and proves its asymptotic optimality. In Section 2.5 we consider the multiple-adjustment scenario. Similarly to the single-adjustment scenario, we first characterize the optimal policy of a corresponding stochastic dynamic program. We then propose a data-drive heuristic policy under which the firm adjusts its capacity in exponentially increasing intervals and show that this policy is asymptotically optimal. We present the set-up and results of our numerical study in Section 2.6 and conclude the chapter in Section 2.7.

2.2 Literature Review

There is an extensive body of literature in the general area of capacity management. Manne (1967), Freidenfelds (1981) and Luss (1982) provide surveys on the earlier literature. In the early work, the main focus is to expand capacity to meet growing demand with no uncertainties. Therefore, the firm is able to make optimal capacity expansion plans to balance economy-of-scale savings and the cost associated with a mismatch between demand and supply. For problems with uncertain demand, Davis et al. (1987) uses the piecewise-deterministic Markov process to model an optimal capacity expansion problem with leadtime. Dixit and Pindyck (1994) provide a survey about the real options approach to analyze investment without detailed operational implications. When the dynamic capacity adjustment is costly and partially irreversible, Eberly and Van Mieghem (1997) present the optimal capacity policy as a control limit policy, labeled as the ISD (invest-stay put-divest) policy. Van Mieghem (2003) provides a comprehensive review about recent developments.

Among more recent literature on capacity management, a number of papers assume the firm has complete information about the parameterized demand distribution. Among them, Chao et al. (2009) characterize a firm's optimal capacity policy when the existing capacity is subject to deterioration and random supply constraints. Besanko et al. (2010) study an oligopoly in which firms make lumpy capacity investment and disinvestment, and show that while firms build excess capacity for a preemption race in the short run, capacity coordination can be achieved in the long run. Wang et al. (2013) show the optimal capacity policy for two competing technologies is a control limit policy. On contrary to these works, our work emphasizes the firm's active role of learning about demand and using it for capacity decisions.

A number of papers consider demand learning in operation context. Boyacı and Özer (2010) consider a firm acquiring information via pricing and advance selling, and characterize the firm's optimal policy to stop collecting information and building capacity as a control band policy. Kwon and Lippman (2011) analyze a firm's optimal strategy to invest in project-specific assets with a real option approach, where the firm's profit follows a Brownian motion, and characterize the optimal policy as a control band policy. Kaminsky and Yuen (2011) show a pharmaceutical firm's investment strategy to acquire clinical trial information and build capacity as a threshold policy. In contrast to these papers, we characterize the firm's optimal policy to adjust capacity (increasing or decreasing) in two different settings – single and multiple adjustment cases. While the setting is similar, we show that the optimal policy and methodology that enables us to characterize the optimal policy can be quite different. In addition to optimal policy, this chapter proposes a simple heuristic that is data-driven and asymptotically optimal for each of the two settings. For each heuristic, we provide a theoretical bound and derive the convergence rate. These heuristics overcome challenges of determining an optimal policy under incomplete information Lovejoy (1993).

Our analysis of the optimal policy is closely related to literature on partially observed Markov decision processes (POMDPs), with a particular emphasis on demand learning with Bayesian updating. That is, decision makers know the family of distributions, and update their knowledge about key parameters characterizing the distribution with new observations. Monahan (1982) and Lovejoy (1991) provide surveys about early works in POMDP. Demand learning in a Bayesian fashion has been applied in inventory management (e.g., Scarf 1959, Azoury 1985, Eppen and Iyer 1997, Lariviere and Porteus 1999, Burnetas and Gilbert 2001, Chen and Plambeck 2008). Recently, Aviv and Pazgal (2005) analyze a firm's pricing decision using the POMDP framework. In this chapter, we analyze a different operational decision, capacity, which is costly to adjust and the adjustment process is often associated with a non-trivial leadtime.

Methodologically, our heuristics are closely related with the recent research on data-driven optimization. Most papers have focused on inventory (Huh et al. 2011, Besbes and Muharremoglu 2013) and pricing (Burnetas and Smith 2000). Some papers also use regret to quantify the heuristics. For example, Huh and Rusmevichientong (2009) analyze a firm's inventory decision with censored demand and no knowledge about demand distribution. They show that using policies derived from online convex optimization, the regret asymptotically converges to 0. Besbes and Zeevi (2009) propose a dynamic pricing algorithm when the demand function is not known, and show that the regret asymptotically converges to 0. To the best of our knowledge, we are one of the first to apply data-driven optimization in the capacity management setting. In contrast with inventory and pricing decisions, a firm usually has limited opportunities to adjust its capacity, and the adjustment process is often costly and lengthy, which makes the problem somewhat challenging.

2.3 Capacity Investment with a Single Adjustment Opportunity

We first consider the case where the firm has a single opportunity to adjust (add or remove) capacity during a planning horizon. This model is appropriate in an environment where the leadtime for changing capacity level is considerably long (relative to the planning horizon) and/or the cost of adjusting capacity is high. In each period, the firm decides whether to change its capacity with existing information about demand or decide to delay the decision and observe the demand for one more period. We assume that the firm has incomplete information about the demand: while the firm knows the demand pattern or distribution family about the demand, some key parameters characterizing the demand are unknown. Thus, the key decisions of the firm in each period are if the firm should change the capacity or not, and if so, by how much.

We consider a firm serving a single product for a finite horizon of J periods, with period 1 and J as the starting and ending periods respectively. We assume that each period is of length τ units of time, which will be useful to derive the heuristic in Section 2.4. There are $I \in \mathbb{N}$ potential demand types: θ_i for $i \in \{1, 2, ..., I\}$ and $\theta_{i_1} < \theta_{i_2}$ if $i_1 < i_2$. The demand type parameter, θ_i, determines the demand distribution. Thus, for given demand type i, the demand in period j, D_j, is represented by a random variable $D_j|\theta_i = \lambda_j(\theta_i) + \xi_j|\theta_i$ where $\lambda_j(\theta_i)$ is the mean demand of $D_j|\theta_i$, and $\xi_j|\theta_i$ is a random term with mean 0. We assume the random term ξ_j is independent across periods. A number of demand processes can be expressed in this way and our results on the optimal policy apply to a large class of random variables and demand processes (see remark on demand process in Section 2.3.2).

We assume that demand in each period is stochastically ordered in the demand type parameter: $D_j|\theta_{i_1} \preceq_{st} D_j|\theta_{i_2}$ for $i_1 \leq i_2$. Thus, demand stochastically increases in the demand type index, i. We use $F_j(\cdot|\theta_i)$ and $f_j(\cdot|\theta_i)$ to denote the cumulative distribution and the density function (probability mass function in the case of discrete demand) of $D_j|\theta_i$. Finally, for ease of exposition, we write $\lambda_j(\theta_i)$ as $\lambda_{j,i}$, and assume $\lambda_{j,I} < \infty$ for all j for analytical tractability.

Because the true demand type is unknown, the firm observes sales (demand) and uses the observations to update its belief about the true demand type. The firm's information about the demand evolves as follows. Let the vector $\boldsymbol{\pi_1}$ be the firm's prior distribution of the demand type at the beginning of period 1: $\boldsymbol{\pi_1} = (\pi_{1,1}, \ldots, \pi_{1,I})$ where $\pi_{1,i} = \Pr(\Theta = \theta_i)$. At the beginning of period j ($j > 1$), the firm's information about the demand type is represented by an information vector $\boldsymbol{\pi_j} \triangleq (\pi_{j,1}, \pi_{j,2}, ..., \pi_{j,I})$. The $\pi_{j,i}$ is defined as the posterior distribution of the demand being type i given the past demand history, i.e., $\pi_{j,i} \triangleq \Pr(\Theta = \theta_i|\boldsymbol{d_{j-1}})$ where $\boldsymbol{d_{j-1}} \triangleq (d_1, d_2, ..., d_{j-1})$, and d_k indicates the realized demand in period k. After the firm observes d_j at the end of period j, the information vector is updated following Bayes' rule:

$$\pi_{j+1,i} = \frac{\pi_{j,i} f_j(d_j|\theta_i)}{\sum_{k=1}^{I} \left[\pi_{j,k} f_j(d_j|\theta_k)\right]}. \tag{2.1}$$

Before the realization of D_j, the information vector is a vector of random variables (denoted by $\boldsymbol{\Pi_{j+1}}$), which we prove below satisfies the martingale property. (All proofs are provided in the Appendix.)

Lemma II.1 (Martingale property of the posterior distribution)**.**

$$E\left[\boldsymbol{\Pi_{j_2}}|\boldsymbol{\Pi_{j_1}}\right] = \boldsymbol{\Pi_{j_1}},\ \textit{for}\ j_1 \leq j_2. \tag{2.2}$$

This lemma implies that given the current distribution about the demand types, the conditional posterior distributions in the future periods are the same as the current one in expectation. This result allows us to derive the firm's expected profit in future periods given the current distribution.

In each period, the firm observes the realized demand, d_j, and fulfills the demand using the firm's existing capacity in that period. For each unit it satisfies with existing capacity, the firm accrues a profit of p, which represents the revenue minus the variable production cost (excluding any capacity cost). If demand exceeds the firm's capacity, it is satisfied by a more expensive outside option such as outsourcing, overtime production, or using other production facilities or resources owned by the firm. Let c_1 be the per-unit outside option cost (we call this *outside option* or *outsourcing* cost). Note that c_1 represents the cost premium of producing one unit using the firm's outside option. In addition to production costs, the firm also incurs an overhead cost to maintain the existing capacity, denoted by c_0 per unit capacity and unit time. As this cost represents the firm's cost to own and maintain the capacity, it is incurred whether the capacity is used or not in that period. To avoid trivial cases, we assume $p \geq c_1 > c_0$, i.e., the unit profit is higher than the unit cost associated with the outside option, otherwise the firm will not outsource any demand; the unit outsourcing cost is higher than the cost to maintain one unit of the firm's own capacity for one period, otherwise the firm will not have incentive to build any capacity. A similar cost structure was applied in Chao et al. (2009).

When the firm's capacity level is μ and the firm's belief about the demand type is $\boldsymbol{\pi_j}$, the firm's expected operating profit $h_j(\boldsymbol{\pi_j}, \mu)$ in period j (note that each period is τ units of time) is:

$$\begin{aligned} h_j(\boldsymbol{\pi_j}, \mu) &\triangleq E_{\Theta}\left\{ E_{D_j|\Theta}\left[pD_j - c_1\left(D_j - \mu\tau\right)^+ - c_0\mu\tau|\Theta\right] \middle| \, \boldsymbol{\pi_j} \right\} \\ &= \sum_{k=1}^{I} \pi_{j,k} E\left[pD_j - c_1\left(D_j - \mu\tau\right)^+ - c_0\mu\tau|\Theta = \theta_k\right], \text{ where } x^+ \triangleq \max\{x, 0\}. \end{aligned} \tag{2.3}$$

We note that in our base model, there is no inventory carryover and demand is not censored as the extra demand beyond the firm's own capacity can be satisfied by an outside option. We provide a discussion and extension of these features in Section 2.3.2.

We next describe the firm's capacity decision. At the beginning of the planning horizon, the firm has an initial capacity level μ_0. This is for generality. Of course, the firm may start with no existing capacity: $\mu_0 = 0$, or the firm may use the prior distribution to choose a capacity level or use the existing (legacy) capacity: $\mu_0 > 0$. In each period, the firm decides whether it should (1) continue to observe the demand and keep the initial capacity level, or (2) stop observing the demand and change the capacity. As changing capacity often requires considerable amount of time, we assume there is a leadtime of l periods.

To be more specific, suppose that the firm has a capacity level of μ in period j. If the firm already adjusted its capacity in previous periods (and perhaps is waiting to be installed) or has decided to wait, then no adjustment will be made and the firm will fulfill the demand with existing capacity μ and an outside option as described above[1]. On the other hand, if the firm decides to change the capacity level from μ to μ' in period j, the firm's existing capacity will be changed to μ' after l periods (in period $j + l$). We assume that both increasing and decreasing the capacity level are costly to the firm. Let c_a be the cost of adding one unit of capacity and γ_a be the cost of decreasing one unit of capacity. Thus, the cost associated with changing the capacity level from μ to μ', denoted by $\hat{C}(\mu, \mu')$, is

$$\hat{C}(\mu, \mu') \triangleq c_a(\mu' - \mu)^+ + \gamma_a(\mu - \mu')^+ \tag{2.4}$$

Notice that if the firm does not change the capacity, $\hat{C}(\mu, \mu) = 0$. We assume $c_a \geq 0$ and $c_a + \gamma_a \geq 0$, indicating that it is costly to reverse the installed capacity[2]. Note

[1]Since each period is τ units of time, the maximum demand the firm can satisfy with its own capacity in this case is $\mu\tau$.

[2]A similar assumption was made in Eberly and Van Mieghem (1997).

that $\gamma_a < 0$ implies that the firm may salvage a portion of its capacity cost, and $\gamma_a \geq 0$ implies that downsizing the capacity is costly to the firm (e.g., the firm needs to pay the layoff costs). To avoid trivial cases, we also assume $c_1(J-l)\tau \geq c_a + c_0(J-l)\tau$ and $c_0(J-l)\tau \geq \gamma_a$. The first assumption implies that it is less costly to increase a unit of capacity and maintain it than outsourcing this unit to the more expensive outside option for the whole time after the adjustment. The second assumption implies that it is cheaper to shrink one unit of capacity than holding it for the whole time after the adjustment.

We allow the set of capacity levels (denoted by $\mathcal{K}$) to be discrete or continuous. When capacity level is primarily determined by the number of key machines or production lines, it may be appropriate that the capacity level must be chosen from a discrete set, i.e., $\mathcal{K} = \{\delta_k, k = 1, 2, ..., |\mathcal{K}|, \delta_k \text{ increasing in } k\}$ where $|\mathcal{K}|$ is the cardinality of the set $\mathcal{K}$. Otherwise, capacity levels can be continuous (e.g., the capacity is measured by the available labor hours), i.e., $\mathcal{K} = \mathbb{R}^+$.

To model the firm's capacity decision, we first introduce the state vector $\boldsymbol{\omega_j} = (\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, v_{j-1})$. Here, $\boldsymbol{\pi_j}$ is the firm's belief about demand type given the demands up to period $j-1$, and $\hat{\mu}_{j-1}$ is defined as the induced capacity position at the end of period $j-1$ (since the capacity leadtime is l periods, $\hat{\mu}_{j-1}$ is the capacity level at the end of period $j+l-1$. In general, for period k, we have $\hat{\mu}_k = \mu_{k+l}$ and $\mu_k = \hat{\mu}_{k-l}$). Lastly, v_{j-1} is defined as an indicator to denote whether capacity has been changed on or prior to period $j-1$. Formally, if capacity adjustment is made in period j, we define

$$v_k = \begin{cases} 0 & \text{if } k < j \\ 1 & \text{if } k \geq j \end{cases} \tag{2.5}$$

We next describe the transition of the state vector. We first observe that the transition of $\boldsymbol{\pi_j}$ (specified in equation (2.1)) follows Lemma 1. To describe how capacity position changes, we first introduce u_j to represent the firm's decision to adjust capacity in period j:

$$u_j = \begin{cases} 0 & \text{if the firm decides to stay put and continue to observe the demand} \\ 1 & \text{if the firm decides to adjust capacity in period } j \end{cases} \tag{2.6}$$

As the firm has only a single opportunity to adjust the capacity, the feasible action space to adjust capacity in period j for given v_{j-1}, $\mathcal{A}(v_{j-1})$, is contingent upon whether the firm has adjusted the capacity or not, i.e.,

$$\mathcal{A}(v_{j-1}) = \begin{cases} \{0, 1\} & \text{if } v_{j-1} = 0; \\ \{0\} & \text{if } v_{j-1} = 1. \end{cases} \tag{2.7}$$

If $u_j = 1$, the firm adjusts the capacity level from the initial level μ_0 to maximize the expected profit from period $j + l$ till the end of the planning horizon based on the information vector $\boldsymbol{\pi_j}$.

$$\begin{aligned} \hat{\mu}_j^a(\boldsymbol{\pi_j}) &\triangleq \arg\max_{\mu \in \mathcal{U}} E\left[\sum_{k=j+l}^{J} h_k(\mathbf{\Pi_k}, \mu) - \hat{C}(\mu_0, \mu) \middle| \boldsymbol{\pi_j}\right] \\ &= \arg\max_{\mu \in \mathcal{U}} \left\{\sum_{k=j+l}^{J} h_k(\boldsymbol{\pi_j}, \mu) - \hat{C}(\mu_0, \mu)\right\}. \end{aligned} \tag{2.8}$$

The equality follows Lemma II.1 and the fact that $h_k(\mathbf{\Pi_k}, \mu)$ is linear in $\mathbf{\Pi_k}$. When the maximizer is not unique, as a tie-breaking rule, the firm chooses the smallest capacity level, i.e., $\hat{\mu}_j^a(\boldsymbol{\pi_j}) = \min_i\{\hat{\mu}_i\}$. Then, the firm's (induced) capacity position transits as follows.

$$\hat{\mu}_j(\boldsymbol{\omega_j}, u_j) = \begin{cases} \hat{\mu}_j^a(\boldsymbol{\pi_j}) & \text{if } u_j = 1; \\ \hat{\mu}_{j-1} & \text{if } u_j = 0. \end{cases} \tag{2.9}$$

We observe that using $\hat{\mu}_j^a(\boldsymbol{\pi_j})$ defined in equation (2.8) in the dynamic program turns the firm's decision problem into an optimal stopping problem, i.e., when to pull the trigger and adjust the capacity to the level specified by $\hat{\mu}_j^a(\boldsymbol{\pi_j})$. For ease of exposition, we suppress the dependence of $\hat{\mu}_j^a(\boldsymbol{\pi_j})$ on $\boldsymbol{\pi_j}$ when there is no confusion.

Having characterized the state transition, we next define the objective function. Given the indicator v_{j-1}, and the starting capacity position $\hat{\mu}_{j-1}$, if the firm adjusts its capacity position to $\hat{\mu}_j^a$ in period j (i.e., $u_j = 1$), with a capacity leadtime of l periods, it accrues profit in period $j+l$ with capacity $\mu_{j+l} = \hat{\mu}_j^a$, but pays a capacity adjustment cost in period j. Otherwise the firm's capacity level in period $j+l$ will be $\hat{\mu}_{j-1}$. Formally, we have

$$\begin{aligned} H_j(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, v_{j-1}, u_j) &\triangleq E[h_{j+l}(\boldsymbol{\Pi_{j+l}}, \hat{\mu}_j(\boldsymbol{\omega_j}, u_j)) - \hat{C}\left(\hat{\mu}_{j-1}, \hat{\mu}_j(\boldsymbol{\omega_j}, u_j)\right) |\boldsymbol{\pi_j}] \\ &= h_{j+l}(\boldsymbol{\pi_{j+l}}, \hat{\mu}_j(\boldsymbol{\omega_j}, u_j)) - \hat{C}\left(\hat{\mu}_{j-1}, \hat{\mu}_j(\boldsymbol{\omega_j}, u_j)\right) \\ &= \begin{cases} h_{j+l}(\boldsymbol{\pi_j}, \hat{\mu}_j^a) - \hat{C}\left(\mu_0, \hat{\mu}_j^a\right) & \text{if } u_j = 1 \\ h_{j+l}(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}) & \text{if } u_j = 0 \end{cases} \end{aligned} \tag{2.10}$$

The first equality follows Lemma II.1 and the fact that $h_{j+l}(\boldsymbol{\Pi_{j+l}}, \hat{\mu}_j)$ is linear in $\boldsymbol{\Pi_{j+l}}$. For ease of exposition, we suppress the dependency of $\hat{\mu}_j(\boldsymbol{\omega_j}, u_j)$ on $\boldsymbol{\omega_j}$ and u_j when there is no confusion.

To represent the firm's capacity decision as a dynamic program, we define a *policy* as a sequence of functions mapping the information states to the action space $\mathcal{A}(v_{j-1})$ for all $j \leq J-l$, i.e., $\{u_j(\boldsymbol{\omega_j}), j = 1, 2, ..., J-l\}$. We notice that with a leadtime of l, the firm should not adjust its capacity after period $J-l$. Let $\mathcal{G}$ denote the set of all the admissible policies, and the firm's objective is to find a policy $g^* \in \mathcal{G}$ to maximize the expected total profit,

$$\max_{g \in \mathcal{G}} \sum_{k=1}^{l} E\left[h_k(\boldsymbol{\Pi_k}, \mu_0) \middle| \boldsymbol{\pi_1}\right] + \sum_{k=1}^{J-l} E^g\left[H_k(\boldsymbol{\Pi_k}, \hat{\mu}_{k-1}, v_{k-1}, u_k) \middle| \boldsymbol{\pi_1}\right] \tag{2.11}$$

where the expectation is taken over D_j for all j at time zero. Due to the l-period leadtime, the expected profit of the first l periods, $\sum_{k=1}^{l} E\left[h_k(\mathbf{\Pi_k}, \mu_0)|\,\boldsymbol{\pi_1}\right]$, is independent of the firm's capacity adjustment policy. Therefore, it is sufficient to maximize

$$\max_{g\in\mathcal{G}} \sum_{k=1}^{J-l} E^{g}\left[H_k(\mathbf{\Pi_k}, \hat{\mu}_{k-1}, v_{k-1}, u_k)|\boldsymbol{\pi_1}\right] \tag{2.12}$$

Define a partial policy $g_j \triangleq \{u_k(\boldsymbol{\pi_k}, \hat{\mu}_{k-1}, v_{k-1}), k = j, ..., J-l\}$ and the set of all the admissible partial policies by $\mathcal{G}_j$. Then at the beginning of period j, given the initial states $\boldsymbol{\pi_j}$, $\hat{\mu}_{j-1}$ and v_{j-1}, the firm's optimal value-to-go function is

$$V_j(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, v_{j-1}) = \max_{g_j\in\mathcal{G}_j} \sum_{k=j}^{J-l} E^{g_j}\left[H_k(\mathbf{\Pi}_k, \hat{\mu}_{k-1}, v_{k-1}, u_k)|\,\boldsymbol{\pi_j}\right] \tag{2.13}$$

Then, the optimal value-to-go functions satisfy the following recursive optimality equations for all $j \in \{1, 2, ..., J-l\}$.

$$\begin{aligned} &V_j(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, v_{j-1}) = \max_{u_j\in\mathcal{A}(v_{j-1})} \left\{H_j(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, v_{j-1}, u_j) + E\left[V_{j+1}(\mathbf{\Pi_{j+1}}, \hat{\mu}_j, v_j)|\,\boldsymbol{\pi_j}\right]\right\} \\ &V_k(\boldsymbol{\pi_k}, \hat{\mu}_k, v_k) = 0, \text{ for } k > J-l \end{aligned} \tag{2.14}$$

To simplify the optimality equations above, we observe the following: for $j = 1, 2, ..., J-l$, if the firm has not adjusted the capacity before period j, i.e., $v_{j-1} = 0$, we have $\hat{\mu}_{j-1} = \mu_0$. In this case, if the firm decides to adjust its capacity in period j, i.e., $u_j = 1$, then for $k = j+1, ..., J-l$, we have $\mathcal{A}(v_{k-1}) = \{0\}$ and $u_k = 0$, and therefore the firm's value-to-go function is as follows:

$$\begin{aligned}
L_j^a\left(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, v_{j-1}\right) &= L_j^a\left(\boldsymbol{\pi_j}, \mu_0, 0\right) \\
&\triangleq h_{j+l}\left(\boldsymbol{\pi_j}, \hat{\mu}_j^a\right) - \hat{C}\left(\mu_0, \hat{\mu}_j^a\right) + E\left[V_{j+1}\left(\boldsymbol{\Pi_{j+1}}, \hat{\mu}_j^a, 1\right) \middle| \boldsymbol{\pi_j}\right] \\
&= h_{j+l}\left(\boldsymbol{\pi_j}, \hat{\mu}_j^a\right) - \hat{C}\left(\mu_0, \hat{\mu}_j^a\right) + \sum_{k=j+1}^{J-l} H_k\left(\boldsymbol{\pi}_j, \hat{\mu}_j^a, 1, 0\right) \\
&= \sum_{k=j+l}^{J} h_k\left(\boldsymbol{\pi_j}, \hat{\mu}_j^a\right) - \hat{C}\left(\mu_0, \hat{\mu}_j^a\right) \qquad (2.15)
\end{aligned}$$

We note that the firm's induced capacity position $\hat{\mu}_j^a$ maximizes the value-to-go function (see equation (2.8)), and the firm needs to pay a one-time capacity adjustment cost of $\hat{C}\left(\mu_0, \hat{\mu}_j^a\right)$. After the adjustment, the firm does not have another opportunity to change the capacity (recall that $\mathcal{A}(1) = \{0\}$). Therefore, the firm's expected operating profit in period k is simply $H_k(\boldsymbol{\pi}_j, \hat{\mu}_j^a, 1, 0)$, which in turn equals $h_k(\boldsymbol{\pi}_j, \hat{\mu}_j^a)$ from equation (2.10).

If the firm has not adjusted the capacity ($v_{j-1} = 0$), and decides to delay decision one more period ($u_j = 0$), then we use the superscript s for "stay put", and have the value-to-go function as

$$L_j^s(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, v_{j-1}) = L_j^s(\boldsymbol{\pi_j}, \mu_0, 0) \triangleq h_{j+l}(\boldsymbol{\pi_j}, \mu_0) + E\left[V_{j+1}(\boldsymbol{\Pi_{j+1}}, \mu_0, 0) \middle| \boldsymbol{\pi_j}\right] \qquad (2.16)$$

By delaying the adjustment, the firm earns a profit based on the starting capacity level in this period. However, it maintains the option to change the capacity in the future, as reflected by the term $E\left[V_{j+1}(\boldsymbol{\Pi_{j+1}}, \mu_0, 0) \middle| \boldsymbol{\pi_j}\right]$.

On the other hand, if the firm already adjusted the capacity before, i.e., $v_{j-1} = 1$, then for $k = j, ..., J - l$, we have $\mathcal{A}(v_{k-1}) = \{0\}$ and $u_k = 0$, and we have

$$\begin{aligned} L_j^s(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, v_{j-1}) &= L_j^s(\boldsymbol{\pi_j}, \mu_0, 1) \\ &\triangleq \sum_{k=j}^{J-l} H_k(\boldsymbol{\pi}_j, \hat{\mu}_{j-1}, 1, 0) = \sum_{k=j+l}^{J} h_k(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}) \end{aligned} \tag{2.17}$$

To sum up, we have the following value-to-go functions contingent upon whether the capacity has been adjusted or not.

$$V_j(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, 0) = V_j(\boldsymbol{\pi_j}, \mu_0, 0) = \max\left\{L_j^a(\boldsymbol{\pi_j}, \mu_0, 0), L_j^s(\boldsymbol{\pi_j}, \mu_0, 0)\right\} \tag{2.18}$$

$$V_j(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, 1) = L_j^s(\boldsymbol{\pi_j}, \mu_0, 1) \tag{2.19}$$

When the maximum in equation (2.18) is attained by $L_j^a(\boldsymbol{\pi_j}, \mu_0, 0)$, it is optimal to adjust the capacity. Otherwise, the firm should delay the adjustment and continue to observe the demand. For ease of exposition, we suppress the dependence on μ_0 and v_{j-1}, and write $V_j(\boldsymbol{\pi_j}, \mu_0, 0)$, $L_j^a(\boldsymbol{\pi_j}, \mu_0, 0)$ and $L_j^s(\boldsymbol{\pi_j}, \mu_0, 0)$ as $V_j(\boldsymbol{\pi_j})$, $L_j^a(\boldsymbol{\pi_j})$ and $L_j^s(\boldsymbol{\pi_j})$ respectively. Therefore, to characterize the firm's optimal policy to stop observing the demand and adjust the capacity, we only need to compare $L_j^a(\boldsymbol{\pi_j})$ and $L_j^s(\boldsymbol{\pi_j})$. Note that, in the single adjustment case, the problem of choosing "when to adjust" and "by how much" is recast as an optimal stopping time problem.

2.3.1 Optimal Policy

We now characterize the firm's optimal capacity policy, starting with the case when possible capacity levels are discrete ($\mathcal{K} = \{\delta_k, k = 1, 2, ..., |\mathcal{K}|\}$). We first define a convex partition of the space of feasible information vectors $\boldsymbol{\pi}$: $\mathcal{P} = \{\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_I) : \sum_{i=1}^{I} \pi_i = 1, \pi_i \geq 0\}$.

Definition II.2. $\mathbb{P} = \{\mathcal{P}_k, \mathcal{P}_k \subset \mathcal{P}\}$ is a convex partition of $\mathcal{P}$, if the following conditions are satisfied:

(i) $\emptyset \notin \mathcal{P}_k$;

(ii) $\bigcup_k \mathcal{P}_k = \mathcal{P}$;

(iii) if $k \neq r$, then $\mathcal{P}_k \bigcap \mathcal{P}_r = \emptyset$;

(iv) for any $\alpha \in (0,1)$, if $\boldsymbol{\pi_1} \in \mathcal{P}_k$ and $\boldsymbol{\pi_2} \in \mathcal{P}_k$, then $\alpha\boldsymbol{\pi_1} + (1-\alpha)\boldsymbol{\pi_2} \in \mathcal{P}_k$.

In other words, $\mathbb{P}$ is a collection of subsets of information vectors where each subset is non-empty and convex, and the union of these subsets is $\mathcal{P}$.

We next characterize the firm's optimal policy to adjust the capacity. We use $\boldsymbol{\pi} \preceq \boldsymbol{\pi'}$ to denote that the posterior distribution $\boldsymbol{\pi}$ is smaller than $\boldsymbol{\pi'}$ in the first order stochastic sense, i.e., $\sum_{k=1}^{i} \pi_k \geq \sum_{k=1}^{i} \pi'_k$ for all $i = 1, 2, ..., I$. The following proposition characterizes the optimal policy.

Proposition II.3 (Optimal capacity policy: Discrete capacity case)**.** *Let $\hat{\mu}_j^*(\boldsymbol{\pi})$ be the optimal capacity position in period j given information vector $\boldsymbol{\pi}$. For $j = 1, 2, ..., J-l$:*
(i) $L_j^a(\boldsymbol{\pi})$ and $L_j^s(\boldsymbol{\pi})$ are convex in $\boldsymbol{\pi}$. Therefore, $V_j(\boldsymbol{\pi})$ is convex in $\boldsymbol{\pi}$.
(ii) Let $\mathcal{P}_{jk} = \{\boldsymbol{\pi} : \hat{\mu}_j^a(\boldsymbol{\pi}) = \delta_k\}$. Then, $\mathbb{P}_j = \{\mathcal{P}_{jk} : \mathcal{P}_{jk} \neq \emptyset, k = 1, ..., |\mathcal{K}|\}$ is a convex partition of $\mathcal{P}$.
(iii) In each $\mathcal{P}_{jk} \in \mathbb{P}_j$, there exists at most one convex set $\mathcal{S}_{jk} \subseteq \mathcal{P}_{jk}$ such that if $\boldsymbol{\pi} \in \mathcal{S}_{jk}$, it is optimal to adjust the capacity position to δ_k, $\hat{\mu}_j^*(\boldsymbol{\pi}) = \hat{\mu}_j^a(\boldsymbol{\pi}) = \delta_k$. If $\boldsymbol{\pi} \notin \bigcup_k \mathcal{S}_{jk}$, then it is optimal to wait: $\hat{\mu}_j^*(\boldsymbol{\pi}) = \hat{\mu}_{j-1}(\boldsymbol{\pi})$.
(iv) Let $S_j = \bigcup_k S_{jk}$. If $\boldsymbol{\pi}, \boldsymbol{\pi'} \in S_j$ and $\boldsymbol{\pi} \preceq \boldsymbol{\pi'}$, then $\hat{\mu}_j^*(\boldsymbol{\pi}) \leq \hat{\mu}_j^*(\boldsymbol{\pi'})$.

Part (iii) of Proposition II.3 implies that the firm's decision to adjust capacity level in the current period is not monotone in its belief about the demand type, $\boldsymbol{\pi}$. That is, it is possible that the firm may increase the capacity when the likelihood of high demand is low, but wait to observe more demand when the likelihood of high demand becomes even higher, i.e., $\hat{\mu}_j^*(\boldsymbol{\pi}) > \hat{\mu}_j^*(\boldsymbol{\pi'})$ for $\boldsymbol{\pi} \prec \boldsymbol{\pi'}$. Thus, as $\boldsymbol{\pi}$ stochastically increases, the optimal policy can switch multiple times between waiting and adjusting. Within each $\mathcal{P}_{jk} \in \mathbb{P}_j$, it is optimal to adjust capacity only when $\boldsymbol{\pi}$

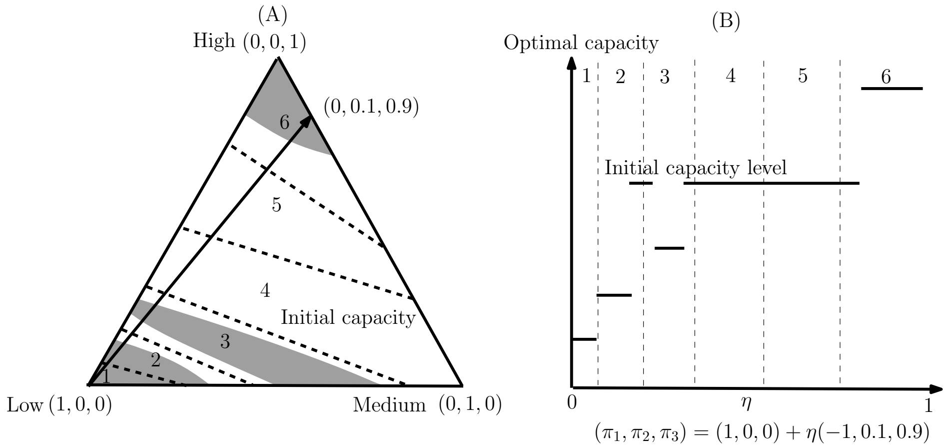

Figure 2.1: An illustrative example of optimal policy with three demand types and discrete capacity levels. Region $k = 1, 2, ..., 6$ corresponds to $\mathcal{P}_{jk}$. The grey area (if any) within each region corresponds to $\mathcal{S}_{jk}$. The initial capacity level corresponds to the optimal target capacity in region 4. The information vector is (π_1, π_2, π_3).

falls in a convex subset $\mathcal{S}_{jk}$. If $\boldsymbol{\pi} \in \mathcal{P}_{jk} \backslash \mathcal{S}_{jk}$, it is optimal to wait. Recall that the problem defined in equation (2.18) is indeed an optimal stopping problem. Thus, one would expect that the optimal policy would be characterized by a monotone threshold (switching curve) in information vector as $\boldsymbol{\pi}$ stochastically increases. Proposition II.3 shows that it is not the case. Although this is quite counter-intuitive at first, this phenomenon indeed reflects the primary trade-off that the firm juggles—exploration versus exploitation. On one hand, the firm would like to exploit benefits from the current information by adjusting the capacity now. On the other hand, if a few more observations of the demand (and, resultantly, updated belief) may shift the firm's target capacity level considerably, it might be beneficial to wait. Part (iv) shows that in regions where it is optimal to change the capacity level, the target capacity level increases in the information state, $\boldsymbol{\pi}$. In other words, given that the firm changes the capacity in the same period, the optimal capacity level is monotonically increasing in the information vector.

Figure 2.1(A) illustrates how the optimal policy changes in $\boldsymbol{\pi}$. In this case, the space of feasible information vectors $\mathcal{P}$ is partitioned into 6 convex subsets ($\mathcal{P}_{jk}$), and each subset corresponds to a different level of $\hat{\mu}^a = \delta_k$ (i.e., the induced capacity level given the firm decides to adjust capacity in that period). The shaded areas correspond to the regions in which it is optimal to adjust the capacity ($\mathcal{S}_{jk}$). Notice that the firm may choose to adjust capacity for the whole region ($\mathcal{S}_{j1} = \mathcal{P}_{j1}$), or choose to wait for the whole region ($\mathcal{S}_{j5} = \emptyset$).

Observe that $\hat{\mu}^a$ increases in the information vector, i.e., $\hat{\mu}^a(\boldsymbol{\pi}) \leq \hat{\mu}^a(\boldsymbol{\pi'})$ when $\boldsymbol{\pi} \preceq \boldsymbol{\pi'}$. However, the decision to adjust the capacity is not monotone in $\boldsymbol{\pi}$. In this case, given that the initial capacity corresponds to the optimal target level in region 4, as the information vector increases, the optimal decision on when to change the capacity is not monotone. For example, consider Figure 2.1(B). This figure shows how optimal policy changes when the information state changes from $(1, 0, 0)$ to $(0, 0.1, 0.9)$ in the direction of $(-1, 0.1, 0.9)$: thus, the information vector is stochastically ordered along the line. The firm first chooses to adjust down (regions $1-2$), then stay put (regions $2-3$), then adjust down again (region 3), then stay put again (regions $3-6$), and finally adjust up (region 6). While one may think that this discontinuity is driven by the fact that the feasible capacity level must be chosen from a discrete set, we show that the same result holds even when the capacity level is a continuous variable, as shown in the next proposition.

Proposition II.4 (Optimal capacity policy: Continuous capacity case)**.** *Let $\hat{\mu}^*_j(\boldsymbol{\pi})$ be the optimal capacity position in period j given information vector $\boldsymbol{\pi}$. For $j = 1, ..., J - l$,*

(i) $L^a_j(\boldsymbol{\pi})$ and $L^s_j(\boldsymbol{\pi})$ are convex in $\boldsymbol{\pi}$. Therefore, $V_j(\boldsymbol{\pi_j})$ is convex in $\boldsymbol{\pi}$.

(ii) Let $S_j = \{\boldsymbol{\pi} : L^a_j(\boldsymbol{\pi}) > L^s_j(\boldsymbol{\pi})\}$. For $\boldsymbol{\pi}$ and $\boldsymbol{\pi'} \in S_j$, if $\boldsymbol{\pi} \preceq \boldsymbol{\pi'}$, then $\hat{\mu}^*_j(\boldsymbol{\pi}) \leq \hat{\mu}^*_j(\boldsymbol{\pi'})$.

Continuous capacity level is the limiting case of discrete capacity level as the number of potential capacity levels increases and the difference between two adjacent

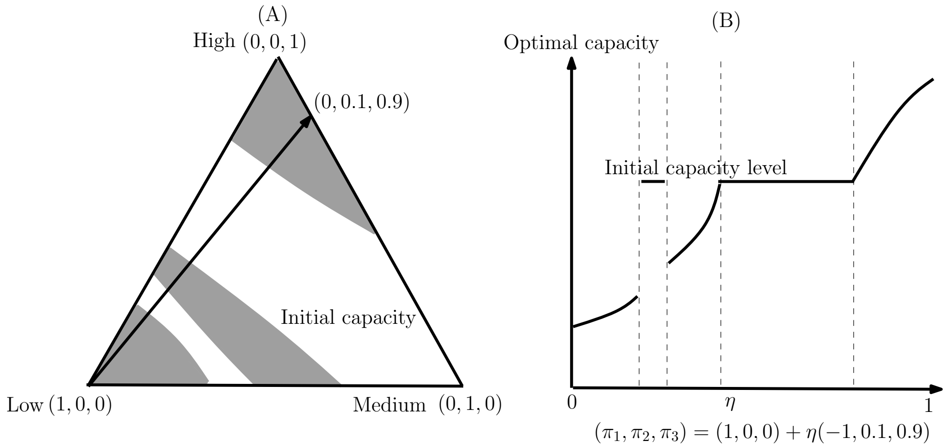

Figure 2.2: An illustrative example of optimal policy with three demand types and continuous capacity levels. The grey areas are the capacity adjustment regions S_j. The initial capacity level corresponds to the optimal target capacity in the middle region. The information vector is (π_1, π_2, π_3).

levels is infinitesimally small. Thus as in the discrete type case if the firm decides to adjust the capacity, then the target capacity position increases in the information vector. However, the firm's decision about whether adjust the capacity in this period does not change monotonically with respect to the increased likelihood, as this when-to-stop decision is determined by comparing two convex functions, $L_j^a(\boldsymbol{\pi})$ and $L_j^s(\boldsymbol{\pi})$, in the optimal stopping problem. The optimal policy is illustrated in Figure 2.2 with a three-demand-type example.

In this case, we observe that the value-to-go function is not necessarily a concave function in the initial capacity. Therefore, the optimal policy cannot be simply characterized by the control limit policy shown in Eberly and Van Mieghem (1997).

In Section 2.5, we will consider the multiple adjustment case and highlight the difference. The next section will derive a heuristic policy and analyze its performance.

2.3.2 Remarks

We briefly discuss some of the modeling features and assumptions, the rationale behind them, and the consequences of removing or relaxing them.

Demand process. The optimal policy characterized in Section 2.3.1 can be applied to a large class of random variables and demand processes. In our base model, we have finite demand types and each type is characterized by a demand type parameter. However, our model can be extended to accommodate more general features. First, demand type i can be characterized by a vector of parameters $\boldsymbol{\theta_i}$. We only require the demand stochastically increases in the demand type index, i.e., $D_j|\boldsymbol{\theta_{i_1}} \preceq_{st} D_j|\boldsymbol{\theta_{i_2}}$ for $i_1 \leq i_2$. Thus as long as the demand type forms an ordered set, our results apply. Second, if there are uncountably infinite demand types, i.e., the prior and posterior distributions are characterized by a continuous distribution function, Proposition II.3 and II.4 still hold. That is, assuming the firm decides to adjust the capacity, the target capacity increases as the likelihood of demand being high increases. As in the base model, the decision to adjust the capacity is not monotone in the likelihood. Finally, the optimal policy still holds when the demand is non-stationary; for example, $D_j|\theta_i$ may represent a non-stationary Poisson process with the mean demand $\lambda_j(\theta_i)$ following a Bass diffusion curve where the market size is θ_i and the coefficient of innovation and coefficient of imitation are fixed across all the demand types. In this case, the random term $\xi_j|\theta_i$ represents a "shifted Poisson" distribution, which has mean 0 and variance $\lambda_j(\theta_i)$.

Censored demand. As the demand beyond capacity is satisfied by an outside option (e.g., outsourcing, overtime, or temporarily using the capacity designated for a different product), demand is fully observed and not censored. However, our model can be extended to accommodate censored demand. In the case of unobservable lost sales, the posterior distribution can be updated as follows:

$$\pi_{j+1,i} = \begin{cases} \frac{\pi_{j,i} f_j(d_j|\theta_i)}{\sum_{k=1}^{I}\left[\pi_{j,k} f_j(d_j|\theta_k)\right]} & \text{if } d_j < \mu \\ \frac{\pi_{j,i} Pr(D_j \geq \mu|\Theta=\theta_i)}{\sum_{k=1}^{I}\left[\pi_{j,k} Pr(D_j \geq \mu|\Theta=\theta_k)\right]} & \text{if } d_j \geq \mu \end{cases} \tag{2.20}$$

Correspondingly, the firm's expected profit in one period is changed as follows.

$$h_j^c(\boldsymbol{\pi_j}, \mu) = \sum_{k=1}^{I} \pi_{j,k} E\left[p \min\{D_j, \mu\tau\} - c_0\mu\tau | \Theta = \theta_k\right] \tag{2.21}$$

Then, following a similar process of defining equation (2.14), we can define the value-to-go function $V_j^c(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, v_{j-1})$. Following a similar proof as that of Proposition II.3, we can show that an optimal policy with similar structure holds.

Inventory. In our model, we assume there is no inventory carried over between two consecutive periods as the demand beyond the capacity is satisfied by an outside option at a higher cost. This higher cost to satisfy demand beyond capacity partially captures the effect of inventory in the backlog case. For example, toy manufacturers in China were constrained by capacity due to low labor retention, and therefore had to use more expensive expediting methods to ship toys from China to the U.S. (Mattioli and Burkitt, 2013). On the other hand, inventory is known as a substitute for capacity in firm operations. Therefore, assuming there is no inventory allows us to isolate the substitution effect and focus on capacity management.

Discount factor. We implicitly assume the discount factor is 1 in this work. This follows the fact that the finite life cycle of the product is relatively short, and therefore we can neglect the time value of wealth. We note that the analytic results, i.e., Proposition II.3 and II.4, will also hold if the discount factor is less than 1. Nevertheless assuming the discount factor to be 1 simplifies the notation to evaluate the performance of the asymptotically optimal heuristic in Section 2.4, and this is standard in asymptotic analysis, for example, see Besbes and Zeevi (2009).

2.4 Near-Optimal Heuristic and Performance Evaluation

As Section 4.2 shows, the optimal policy is very complicated and difficult to implement even for problems with finite demand types and capacity levels. One of the reasons is that the state space—which includes the information vector $\boldsymbol{\pi}$—is uncountably infinite, and therefore computing the exact optimal policy is computationally intractable for large problems. For smaller problems, a fine mesh approximation with linear interpolation can approximate the value-to-go function and hence the optimal policy (as we do for one part of our validation case study in Section 6), but in general, the curse of dimensionality makes it impossible to find the optimal policy.

Therefore, we propose a simple two-step heuristic. The firm observes demand for a specific amount of time (τ_n units of time whose value depends on the problem size) and then adjusts the capacity based on the observed demand. We then show that, in an asymptotic regime, this heuristic is near-optimal when the underlying demand process is a stationary process with unknown mean under the *regret* criterion, which quantifies the gap between an upper-bound (based on information relaxation and deterministic approximation) and the value-to-go function derived from the two-step heuristic.

We scale up both demand and capacity by a coefficient n to define the asymptotic regime. For this, consider the firm's problem with a planning horizon $[0, T]$ within which the firm reviews its decision periodically. Let τ_n be the time between two consecutive decision opportunities so that the corresponding decision problem is a discrete-time dynamic program with $J_n = T/\tau_n$ periods. Likewise, let l_n be the capacity lead time (described in the number of periods): $l_n = l_t/\tau_n$ where $l_t \in [0, T]$. Without loss of generality, we assume $J_n = T/\tau_n$ and $l_n = l_t/\tau_n$ are integers.

We assume that the firm's demand follows a stationary random process with an unknown average demand rate. Let $\{N(t), t \geq 0\}$ denote a standard random process with stationary and independent increment, which satisfies $N(0) = 0$, has mean $E[N(t)] = t$, and variance $Var[N(t)] = \sigma^2 t$ for $t \geq 0$. When the demand type is i, we define $\{N(n\lambda_i t), t \geq 0\}$, $i \in \{1, 2, ..., I\}$ as the demand process and the demand type parameter $\theta_{i,n} = n\lambda_i$. That is, given demand type i, the firm's demand in period j

Table 2.1: The two-step heuristic

	Given the period length of τ_n,
1.	The firm serves the demand in period 1 with initial capacity $n\mu_0$. Let $n\hat{\lambda}_{\tau_n}$ be the observed demand rate in period 1.
2.	The firm adjusts its capacity position to $n\hat{\lambda}_{\tau_n}$.
3.	The firm serves demand from period 2 to $l_n + 1$ with the initial capacity $n\mu_0$, and from $l_n + 2$ to J_n with capacity $n\hat{\lambda}_{\tau_n}$.

is $D_j|\theta_{i,n} = N(n\lambda_i j\tau_n) - N(n\lambda_i(j-1)\tau_n)$, and therefore the demand in each period is a sequence of i.i.d. random variables with mean $n\lambda_i\tau_n$ and variance $\sigma^2 n\lambda_i\tau_n$. We assume that the firm's initial capacity is scaled up as $n\mu_0$. All other aspects of the model (e.g., costs, revenue, etc.) are the same as the original model considered in Section 4.2.

To show the asymptotic optimality, we impose the following assumptions on τ_n and λ_i.

Assumption II.5. $\lim_{n\to\infty} \tau_n = 0$*;* $\lim_{n\to\infty} n\tau_n = \infty$*;* $\lambda_i \in [0,\ M]$.

The assumption stipulates that, as the problem scale increases, the length of the observation period (τ_n) asymptotically converges to 0 at a relatively slow speed, and the demand rate for any type i is bounded from above. With this set-up, we now introduce and analyze the two-step heuristic, denoted by *(ts)*.

The two-step heuristic. In the heuristic, the firm observes demand for one period comprising τ_n units of time, and then uses the observed demand rate to adjust the capacity for the rest of the time horizon, as specified in Table 2.1. We will show that this simple policy is asymptotically optimal with an appropriately chosen τ_n (which is a function of the scale parameter n).

Under the two-step heuristic, the firm always adjusts the capacity to the observed demand rate in the first period. Therefore, we define

$$\hat{\lambda}_{i,\tau_n} \triangleq \frac{N(n\lambda_i\tau_n)}{n\tau_n}, \tag{2.22}$$

Then the firm's expected value-to-go function under the heuristic is as follows.

$$V_{0,n}^{ts}(\boldsymbol{\pi_1}) = \sum_{i=1}^{I} \pi_{1,i} E \left\{ \begin{array}{l} pD_{l_n+1} - c_1\left(D_{l_n+1} - n\mu_0\tau_n\right)^+ - c_0 n\mu_0\tau_n \\ -\hat{C}(n\mu_0, n\hat{\lambda}_{i,\tau_n}) \\ + \sum\limits_{j=l_n+2}^{J_n} \left[pD_j - c_1\left(D_j - n\hat{\lambda}_{i,\tau_n}\tau_n\right)^+ - c_0 n\hat{\lambda}_{i,\tau_n}\tau_n\right] \end{array} \middle| \, \theta_{i,n} \right\} \tag{2.23}$$

For ease of exposition, we suppress the dependency of $V_{0,n}^{ts}(\boldsymbol{\pi_1})$ on $\boldsymbol{\pi_1}$ when there is no confusion.

As the two-step heuristic is a feasible policy for the corresponding optimal stopping problem, it follows that the value-to-go function under the two-step heuristic, $V_{0,n}^{ts}$, is a lower bound of the value-to-go function under the optimal policy, denoted by $V_{0,n}^{*}$. However, because of the complexity of the optimal policy and the curse of dimensionality, the exact value function under the optimal policy, denoted by $V_{0,n}^{*}$ is difficult to compute. Hence, we will introduce an upper-bound of $V_{0,n}^{*}$ to evaluate the performance of the heuristic.

Upper bound. We derive an upper bound of $V_{0,n}^{*}$ based on information structure relaxation. Consider a hypothetical model, where the information of demand type is revealed to the firm in the first period. In this case, the firm has *full information (fi)* about the demand type, and is able to decide the optimal capacity position contingent upon the demand type. Consequently, we obtain the firm's value-to-go function as follows:

$$V_{0,n}^{fi}(\boldsymbol{\pi_1}) = \max_{\mu_1,\ldots,\mu_I} \sum_{i=1}^{I} \pi_{1,i} E \left\{ \begin{array}{l} \sum\limits_{j=l_n+1}^{J_n} \left[pD_j - c_1\left(D_j - n\mu_i\tau_n\right)^+ - c_0 n\mu_i\tau_n\right] \\ -\hat{C}(n\mu_0, n\mu_i) \end{array} \middle| \, \theta_{i,n} \right\} \tag{2.24}$$

We observe that the value-to-go function above is concave in the demand. Therefore, by Jensen's inequality, we have an upper bound of $V_{0,n}^{fi}$ from a *deterministic (d)* problem as follows:

$$\begin{aligned} V_{0,n}^{d}(\boldsymbol{\pi_1}) &= \max_{\mu_1,\ldots,\mu_I} \sum_{i=1}^{I} \pi_{1,i} \left\{ \left[pn\lambda_i - c_1 n(\lambda_i - \mu_i)^+ - c_0 n\mu_i\right](J_n - l_n)\tau_n - \hat{C}(n\mu_0, n\mu_i) \right\} \\ &= \sum_{i=1}^{I} \pi_{1,i} \left\{ (p - c_0)n\lambda_i(J_n - l_n)\tau_n - \hat{C}(n\mu_0, n\lambda_i) \right\} \end{aligned} \tag{2.25}$$

In the deterministic problem described in equation (2.25), the optimal target capacity for demand type i is $\mu_i^* = \lambda_i$. It is not a surprise that the firm's optimal action is to adjust the capacity to the mean instead of a newsvendor type fractile, because the decision problem is deterministic, and there is no uncertainty in the demand. Finally, it follows that $V_{0,n}^* \leq V_{0,n}^{fi} \leq V_{0,n}^d$.

Performance evaluation. To evaluate the performance of the policy in the asymptotic regime, we analyze the metric of *regret*, which measures the gap between the value-to-go function under the heuristic and the deterministic upper bound. Formally, the regret of the two-step heuristic is defined as $R_n^{ts} = 1 - V_{0,n}^{ts}/V_{0,n}^d$. In the following, we say a heuristic is asymptotically optimal if the regret converges to 0 as the scale factor n increases to infinity. For two sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$. Then we characterize the asymptotic regret as follows.

Proposition II.6 (Asymptotic regret: Two-step heuristic)**.** *If* $\tau_n \asymp n^{-\frac{1}{3}}$ *for all* n*,* *the two-step heuristic is asymptotically optimal and* $R_n^{ts} = O\left(n^{-\frac{1}{3}}\right)$.

We first observe that the firm sets $\tau_n \asymp n^{-\frac{1}{3}}$ corresponding to a problem scale of n. This reflects the exploration-exploitation tradeoff the firm faces. For a given problem scale n, the firm has incentive to set a long observation period to *explore* the demand so that it can obtain more demand information. However, the longer the observation period is, the less time is left for the firm to *exploit* the benefit of its knowledge about demand by adjusting the capacity. Therefore, the firm has to choose an appropriate period length to balance this tradeoff. As the problem scale increases, more demand information is available within a unit of time. Therefore, the firm is able to reduce the observation period and starts to exploit its knowledge earlier.

2.5 Capacity Investment with Multiple Adjustment Opportunities

We now move on to the case where the firm can adjust capacity multiple times. At the beginning of each period, the firm first decides whether it will adjust its capacity or not, and if so, by how much. Then the demand is realized and satisfied using the firm's capacity and (if short) an outside option. At the end of the period, the firm updates the posterior distribution of demand types. We present the case where the capacity level set is continuous, i.e., $\mathcal{K} = \mathbb{R}^+$, and there are still I demand types, but the analysis for the discrete capacity level is similar. We use a superscript m to indicate the multiple adjustment model.

Except for multiple adjustment opportunities, the problem setting is identical to the one considered in Section 4.2. Let $\boldsymbol{\pi_1}$ be the prior distribution of demand type, and $\boldsymbol{\pi_j}$ be the posterior distribution at the beginning of period j, as shown in equation (2.1). Each period is still of length τ. As in equation (2.3), let $h_j(\boldsymbol{\pi_j}, \mu)$ be the expected profit in period j (not including the capacity adjustment cost) when the information vector is $\boldsymbol{\pi_j}$ and the current capacity level is μ. As before, l is the leadtime for capacity adjustment: hence there is a l-period lag between capacity position and actual capacity: $\hat{\mu}_j = \mu_{j+l}$ or $\mu_j = \hat{\mu}_{j-l}$. We assume that the adjustment decision cannot be canceled or reversed. That is, if the firm later adjusts capacity down, the firm still incurs the cost for that, and the adjustment also takes effects l periods later. As it is feasible for the firm to make the adjustment decision in every period, the firm does not need to keep track of whether the capacity has been adjusted or not. Therefore, the auxiliary information state v_j is no longer needed. Thus, $(\boldsymbol{\pi_j}, \hat{\mu}_{j-1})$ is the state vector.

Given state $(\boldsymbol{\pi_j}, \hat{\mu}_{j-1})$, define $H_j^m(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, \hat{\mu}_j)$ to be the expected operating profit in period $j+l$ (that is, when the capacity is in effect) minus the capacity adjustment cost that the firm incurs.

$$\begin{aligned} H_j^m(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, \hat{\mu}_j) &\triangleq E[h_{j+l}(\boldsymbol{\Pi_{j+l}}, \hat{\mu}_j) - \hat{C}(\hat{\mu}_{j-1}, \hat{\mu}_j)\,|\boldsymbol{\pi_j}] \\ &= h_{j+l}(\boldsymbol{\pi_j}, \hat{\mu}_j) - \hat{C}(\hat{\mu}_{j-1}, \hat{\mu}_j) \end{aligned} \tag{2.26}$$

The equality follows Lemma II.1 and the fact that $h_{j+l}(\boldsymbol{\Pi_{j+l}}, \hat{\mu}_j)$ is linear in $\boldsymbol{\Pi_{j+l}}$. We also define a policy g as $\{\hat{\mu}_j(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}), j = 1, 2, ..., J-l\}$ and $\mathcal{G}^m$ as the set of all the admissible policies. Likewise, let $g_j \triangleq \{\hat{\mu}_k(\boldsymbol{\pi_k}, \hat{\mu}_{k-1}), k = j, j+1, ..., J-l\}$ and $\mathcal{G}_j^m$ denote a partial policy and the set of all the admissible partial policies, respectively. Then, the firm's problem is to determine a policy $g^* \in \mathcal{G}^m$ to maximize the total expected profit,

$$\max_{g\in\mathcal{G}^m} \sum_{k=1}^{l} E\left[h_k(\boldsymbol{\Pi_k}, \mu_0)|\boldsymbol{\pi_1}\right] + \sum_{k=1}^{J-l} E^g\left[H_k^m(\boldsymbol{\Pi}_k, \hat{\mu}_{k-1}, \hat{\mu}_k)|\boldsymbol{\pi_1}\right] \tag{2.27}$$

where the expectation is taken over D_j for all j at time zero. As the profit from the first l periods is not affected by the firm's capacity decision, the decision problem is to find a policy that maximizes the following function:

$$V_1^m(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}) = \max_{g\in\mathcal{G}^m} \sum_{k=1}^{J-l} E^g\left[H_k^m(\mathbf{\Pi}_k, \hat{\mu}_{k-1}, \hat{\mu}_k)|\boldsymbol{\pi_1}\right] \tag{2.28}$$

Then, the optimal value-to-go function is recursively defined as follows: for all $j \in \{1, 2, ..., J-l\}$,

$$\begin{aligned} V_j^m(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}) &= \max_{\hat{\mu}\in\mathbb{R}^+} E\left[H_j^m(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}, \hat{\mu}) + V_{j+1}^m(\mathbf{\Pi_{j+1}}, \hat{\mu})|\boldsymbol{\pi_j}\right] \\ &= \max_{\hat{\mu}\in\mathbb{R}^+} \left\{h_{j+l}\left(\boldsymbol{\pi_j}, \hat{\mu}\right) - \hat{C}\left(\hat{\mu}_{j-1}, \hat{\mu}\right) + E\left[V_{j+1}^m(\mathbf{\Pi_{j+1}}, \hat{\mu})|\boldsymbol{\pi_j}\right]\right\}; \\ V_k^m(\boldsymbol{\pi_k}, \hat{\mu}_{k-1}) &= 0 \text{ for } k > J-l. \end{aligned} \tag{2.29}$$

We next show that the optimal policy is a control band policy, similar to Eberly and Van Mieghem (1997).

Proposition II.7 (Optimal policy for multiple adjustment opportunities)**.** *Suppose the firm has information vector* $\boldsymbol{\pi}$ *and capacity position* $\hat{\mu}_{j-1}$ *at the beginning of period* j*. Then, the optimal capacity position, denoted by* $\hat{\mu}^*(\boldsymbol{\pi})$*, is characterized by two thresholds* $\underline{\mu}_j(\boldsymbol{\pi})$ *and* $\bar{\mu}_j(\boldsymbol{\pi})$*, such that:*
(i) If $\hat{\mu}_{j-1} < \underline{\mu}_j(\boldsymbol{\pi})$*, it is optimal for the firm to adjust the capacity position up to* $\hat{\mu}^*(\boldsymbol{\pi}) = \underline{\mu}_j(\boldsymbol{\pi})$*.*
(ii) If $\underline{\mu}_j(\boldsymbol{\pi}) \leq \hat{\mu}_{j-1} \leq \bar{\mu}_j(\boldsymbol{\pi})$*, it is optimal for the firm to stay put, i.e.,* $\hat{\mu}^*(\boldsymbol{\pi}) = \hat{\mu}_{j-1}$*.*
(iii) If $\hat{\mu}_{j-1} > \bar{\mu}_j(\boldsymbol{\pi})$*, it is optimal for the firm to adjust the capacity position down to* $\hat{\mu}^*(\boldsymbol{\pi}) = \bar{\mu}_j(\boldsymbol{\pi})$*.*

Intuitively, as it is costly to adjust capacity, even if the firm is entitled with the flexibility to make adjustment in every period, it may not do so. The firm will adjust the capacity only if the current capacity position is significantly lower or higher than the expected level. In contrast to the single adjustment case, where the value-to-go function is not necessarily concave in the initial capacity, in the multiple adjustment case, the value-to-go function is concave in the starting capacity. This concavity enables a simple characterization of the optimal policy by two state-dependent capacity adjustment thresholds. However, the optimal policy is still computationally complex with the information vector $\boldsymbol{\pi}$. This result expands the control limit policy in Eberly and Van Mieghem (1997) by explicitly incorporating the demand learning process and the leadtime to build capacity. The optimal policy allows us to derive a near-optimal heuristic below, which is computationally simple, yet performs asymptotically optimally. Before that, we first discuss the monotonicity of switching curves with respect to the information state.

We next show how the two thresholds $\underline{\mu}_j(\boldsymbol{\pi})$ and $\bar{\mu}_j(\boldsymbol{\pi})$ change in the information state $\boldsymbol{\pi}$. When there are two demand types, the information vector $\boldsymbol{\pi} = (\pi_1, \pi_2)$ can be written as $(1-\pi_2, \pi_2)$ (i.e., losing one degree of freedom), enabling us to reduce the information state to π_2, the probability of demand being the high type. We show in the following lemma that both thresholds increase in the probability of high demand type when the demand distributions satisfy the monotone likelihood ratio property. For ease of exposition, we write π_2 as π.

Lemma II.8 (Monotonicity of switching curves: Two demand type case)**.** *If the likelihood ratio $f_j(d|\theta_2)/f_j(d|\theta_1)$ increases in d, then $\underline{\mu}_j(\pi)$ and $\bar{\mu}_j(\pi)$ increase as the probability of the high demand type, $\pi = P(\Theta = \theta_2)$, increases.*

Lemma II.8 implies that as the high demand type becomes more likely, both invest-up-to and divest-down-to thresholds increase in the two-type case. In other words, the switching curves and the resultant capacity levels that the firm sets under the optimal policy are both monotone in π. This is a sharp contrast to the result of the single adjustment case, in which the optimal policy (and the resultant capacity level) is not monotone in the information state. This highlights the key difference of the decision problems that the firm faces in single and multiple adjustments. In the single adjustment, the firm needs to decide two things: *when to adjust* and *how* *much*. As a result, the firm may decide to wait even in the state when π_2 is high while it is optimal to increase capacity when π_2 is lower. In the multiple adjustment case, however, the firm only needs to worry about how much capacity to adjust in each period. Although there is a stay-put interval, this is cost-driven. If the current capacity is near the target capacity (i.e., the optimal capacity when the capacity cost is ignored in that period), the capacity cost is higher than the profit difference. On the other hand, note that, in the single adjustment case, the non-monotonicity is primarily opportunity driven (if we change the capacity in this period, we cannot change the capacity again).

When there are more than two demand types, one might conjecture that the same result could be directly extended. However, proving the result becomes formidable for two reasons. First, when there are more than two types, the information state $\boldsymbol{\pi}$ is no longer a completely ordered set. Therefore, even when we start with two ordered information vectors, $\boldsymbol{\pi} \preceq \boldsymbol{\pi'}$, future states may not necessarily preserve the same ordering in all sample paths even if we assume monotone likelihood ratios. Second, the value-to-go function corresponding to adjusting the capacity is no longer linear in the information vector, which makes it difficult to apply similar techniques to prove the monotonicity in the target capacity level using Lemma II.1.

2.5.1 Near-Optimal Heuristic and Performance Evaluation

We next derive a simple near-optimal heuristic similar to the one in Section 2.4. The setting is entirely identical other than the fact that the firm is able to adjust its capacity multiple times during the decision horizon $[0, T]$ (equivalently period 1 to J_n in discrete time). We also show that in an asymptotic regime (same as the one defined in Section 2.4), this multi-step heuristic (ms) is asymptotically optimal, and provide a performance upper bound for the heuristic under the regret criterion. To show the asymptotic optimality, we also impose Assumption II.5 on τ_n and λ_i.

Table 2.2: The multi-step heuristic

Given the period length τ_n and the number of adjustment opportunities K_n,

1. The firm serves the demand in period 1 with the initial capacity $n\mu_0$.
2. For $\kappa = 1 : K_n$
   a. The firm adjusts the capacity position at the start of period 2^κ to the observed average demand during the first $2^\kappa - 1$ periods denoted by $n\bar{\lambda}_\kappa$. The capacity level will be updated accordingly l_n periods later.
   b. The firm serves the demand from period 2^κ to $2^{\kappa+1} - 1$ using the (updated) capacity.

   End
3. The firm serves the demand in the remaining periods using the updated capacity.

The multi-step heuristic. In this heuristic, the firm adjusts its capacity only in a subset of the J_n periods, instead of doing it in every period. Specifically, the κ^{th} adjustment of the capacity position occurs at the beginning of period 2^κ, and the actual change of capacity levels occurs at the start of period $2^\kappa + l_n$, for $\kappa = 1, 2, ..., K_n$, where K_n is the largest integer such that $l_n + \sum_{\kappa=1}^{K_n+1} 2^{\kappa-1} \leq J_n$, i.e., $K_n \triangleq \lfloor \log_2(J_n - l_n + 1) \rfloor - 1$. That is, the time between the $\kappa - 1^{th}$ and κ^{th} adjustments is $2^{\kappa-1}\tau_n$ ($2^{\kappa-1}$ periods). The intuition for choosing the exponentially increasing periods between two consecutive adjustment decisions is that as more demand information is collected, adding new observations is less likely to change the information state in a significant way. The details of the heuristic are illustrated in Table 2.2.

In this heuristic, the firm always adjusts the capacity position to the observed demand rate. To evaluate the value-to-go function under this heuristic, we denote the observed demand rate contingent upon the demand type i by $n\bar{\lambda}_{i,\kappa}$ for $\kappa \geq 1$.

Then, we first define $\bar{\lambda}_{i,\kappa}$ recursively below.

$$\bar{\lambda}_{i,1} \triangleq \frac{D_1|\theta_{i,n}}{n\tau_n}$$
$$\bar{\lambda}_{i,\kappa} \triangleq \frac{\bar{\lambda}_{i,\kappa-1}n(2^{\kappa-1}-1)\tau_n + \sum_{j=2^{\kappa-1}}^{2^\kappa-1} D_j|\theta_{i,n}}{n(2^\kappa-1)\tau_n}, \ \kappa = 2,3,...,K_n \tag{2.30}$$

For notational simplicity, we also define $\bar{\lambda}_{i,0} \triangleq \mu_0$. Then we have the firm's expected value-to-go function under this heuristic as follows.

$$V_{0,n}^{ms}(\boldsymbol{\pi_1}) = \sum_{i=1}^{I} \pi_{1,i} E \left\{ \begin{array}{l} \sum_{\kappa=0}^{K_n-1} \sum_{j=l_n+2^\kappa}^{l_n+2^{\kappa+1}-1} \left[pD_j - c_1\left(D_j - n\bar{\lambda}_{i,\kappa}\tau_n\right)^+ - c_0 n\bar{\lambda}_{i,\kappa}\tau_n \right] \\ + \sum_{j=l_n+2^{K_n}}^{J_n} \left[pD_j - c_1\left(D_j - n\bar{\lambda}_{i,K_n}\tau_n\right)^+ - c_0 n\bar{\lambda}_{i,K_n}\tau_n \right] \\ - \sum_{\kappa=1}^{K_n} \hat{C}\left(n\bar{\lambda}_{i,\kappa-1}, n\bar{\lambda}_{i,\kappa}\right) \end{array} \middle| \theta_{i,n} \right\} \tag{2.31}$$

As this heuristic is a feasible policy for the corresponding optimal capacity adjustment problem, we have that $V_{0,n}^{ms} \le V_{0,n}^{m*}$, where $V_{0,n}^{m*}$ denotes the value-to-go function under the optimal policy. As the optimal policy is not computationally tractable, we need to derive an upper bound of the value-to-go function under the optimal policy in order to evaluate the performance of the heuristic.

Upper bound. We first observe that the $V_{0,n}^{d}$ (see equation (2.25) in Section 2.4) is still an upper bound of $V_{0,n}^{m*}$. This is because in the deterministic stationary demand setting, once the firm obtains full information about the demand type, even if the firm is able to adjust capacity any time, it is still optimal to adjust it only once at the beginning of the time horizon as the adjustment is costly. That is, we still have the optimal target capacity $\mu_i^* = \lambda_i$, and $V_{0,n}^d$ as follows.

$$V_{0,n}^d = \sum_{i=1}^{I} \pi_{1,i} \Big\{ (p-c_0)n\lambda_i(J_n - l_n)\tau_n - c_a n(\lambda_i - \mu_0)^+ - \gamma_a n(\mu_0 - \lambda_i)^+ \Big\} \tag{2.32}$$

Performance evaluation. To analyze the performance of the heuristic, we evaluate the asymptotic behavior of the regret of the multi-step heuristic, defined as $R_n^{ms} = 1 - V_{0,n}^{ms}/V_{0,n}^{d}$. We derive the following characterization of the asymptotic regret.

Proposition II.9 (Asymptotic regret: Multi-step heuristic). *If $\tau_n \asymp n^{-\frac{1}{3}}$ for all n, the multi-step heuristic is asymptotically optimal and $R_n^{ms} = O\left(n^{-\frac{1}{3}}\right)$.*

The intuition of the proof is that as the firm observes more demand information and adjusts capacity to match the observed average demand rate, we are able to bound outsourcing costs and capacity adjustment costs by the bound shown in Proposition 1 in Gallego (1992), which derived a one-sided deviation bound for the class of distributions with finite mean and variance. As noted above, we choose the exponentially increasing time between two consecutive decisions because the adjustment is costly, and with more information learned, it is less necessary for the firm to learn about demand frequently. Finally, the time interval τ_n is set to minimize the derived upper bound.

Recall that when the firm has only one chance to adjust its capacity, the upper bound of the regret is also $O\left(n^{-1/3}\right)$ (see Proposition II.6). Here, although the upper bound of the regret is still of the same order, the capacity adjustment cost makes a difference. With multiple adjustment opportunities, the firm is able to correct errors that it might have made in a one shot decision, and therefore, the regret should be smaller. However, when capacity adjustment is very costly, with multiple capacity adjustments specified in the heuristic, the firm needs to pay a higher total capacity adjustment cost as it chases the mean demand. Therefore, the benefit from the learning-while-doing may be diluted. In Section 2.6.2, we compare the two heuristics numerically.

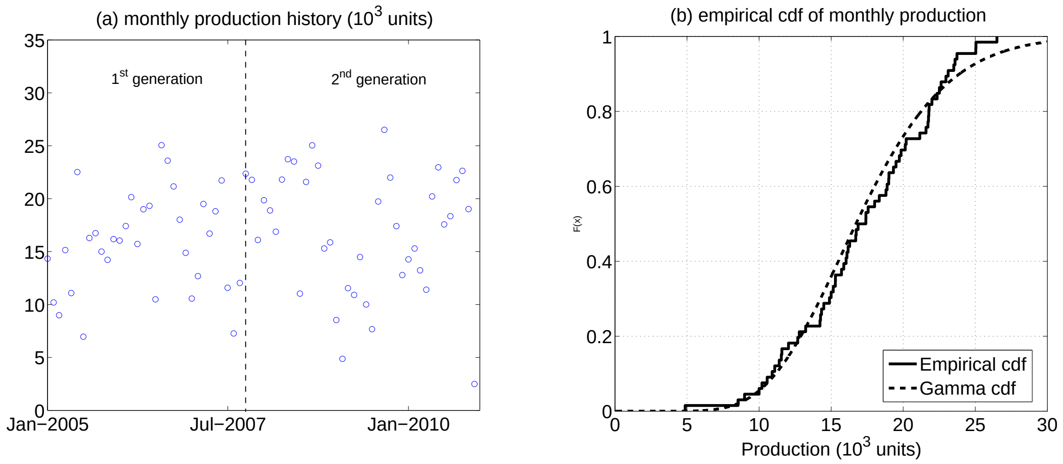

Figure 2.3: Monthly production of Ford Focus in North America market (a), and the empirical cdf and gamma distribution with sample mean and variance (b), from January 2005 to December 2010

2.6 Numerical Study: Ford Focus, the Third Generation

To demonstrate the performance and robustness of our heuristic, we develop a numerical study where the model premises (such as demand pattern, problem scale, cost, and profit) are drawn from practice. Specifically, our example utilizes production and financial data related to the Ford passenger sedan, the Focus. Given the data from the first two generations of the Focus in the North American market, the numerical study illustrates how one could use our heuristics in a setting where the managers of Ford Focus need to decide to adjust the capacity for its third generation. Although we need to make some simplifying assumptions because of the lack of precise accounting data, we show that the results and conclusions are quite robust to model parameters and our assumptions.

2.6.1 Data and Parameter Estimation

In this section, we briefly describe how we collect data and recover the demand and cost parameters, with details deferred to the appendix.

Demand. Our focus is on how the assembly factory should adjust its capacity based on orders received from the dealership. Therefore, the demand is reflected by the number of Ford Focus sedans produced at the assembly factory. This system can be approximately considered as a make-to-order system as Ford receives orders from dealers before the orders are factored into its production plan, and the inventory is held at the dealership level.

To analyze the demand pattern, we first collect monthly production data of Ford Focus in the North America market from January 2005 to December 2010 from the database of *Automotive News Data Center*[3]. There are two (redesigned) generations of Focus during this period: the first from January 2005 to September 2007, and the second from October 2007 to December 2010. Although there is seasonality within each year, which is affected by factors such as mid-year discount when manufacturers switch production to the next year model and end-of-year sales to boost sales figure, we observe the demand pattern is plausibly *stationary* within each generation: see Figure 2.3(a). As the demand pattern of the first two generations is similar, we group the production data from January 2005 to December 2010 and observe that the monthly demand approximately follows a gamma distribution with a mean of 17.21 thousand units per month and a standard deviation of 5.04; see Figure 2.3(b).[4] When we construct the empirical cumulative distribution function, we excluded the data point in July or August if the production in that month is approximately half of a regular month to account for the regular summer shutdown, and the production quantity for December 2010 which is significantly below the average production level as it is the transition time from the second generation production to the third generation. We denote the two key parameters of the gamma distribution by a and b, i.e.,

[3]Automotive News Data Center: http://www.autonews.com/section/datacenter.

[4]We observe that during the automotive industry crisis (2008-2010), the demand pattern of Focus did not change. This may be because the Focus is a fuel-efficient model, and therefore the substantial increase in the prices of automotive fuels did not cause a significant drop in sales, unlike the sport utility vehicles and pickup trucks, whose demands declined in the same period.

the probability density function is characterized as

$$f(x|a,b) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}}$$

where $\Gamma(\cdot)$ represents the gamma function.

In fact, we test the cleaned production data from January 2005 to December 2010 with a gamma distribution where the estimated parameters are $a = 11.67$ and $b = 1.47$ using a one-sample Kolmogorov-Smirnov test, which yields a p-value of 0.74, supporting our choice for the demand distribution. Therefore, we model the monthly demand (with the unit of a thousand cars) for Focus using a stationary gamma distribution.

In our numerical example, we postulate that the decision maker (Ford) has three possible scenarios (demand types) for the third generation Focus. In the medium scenario, the demand will remain at the same level as the first two generations: monthly demand will follow the above gamma distribution. In the other two scenarios, the demand for the third generation (released in May 2011) are either lower or higher than the first two generations as the customers may not like the product, or the economic environment improves. Thus the average monthly demand will be either dropped by or raised by 5 thousand units (which is about one standard deviation). That is, the two key parameters are $a = 8.28$ and $b = 1.47$ for the low demand case, and $a = 15.06$ and $b = 1.47$ for the high demand case. We assume that the parameter b, which stands for the ratio between variance and mean, stays stationary, i.e., a higher demand is associated with a higher variance. We will later show that our result is quite robust with respect to the misspecification of the average demand parameters.

It is important to note that our heuristic (which is data driven) does not rely on knowledge about the prior distribution or the exact demand distribution for each type.

These information is necessary only to evaluate the performance of the heuristic (i.e., computing the regret). Also, it should be noted that our heuristic applies to more general settings (e.g., there are more than three market scenarios, or the unknowns are a vector of parameters rather than a single parameter).

Finally, as the second generation Focus was on sale for three years, we assume the decision horizon T for the third generation is also 3 years, starting from January 2011. Following the convention of the asymptotic analysis, we also assume when $n = 1$, the average medium type demand in the three year horizon is 1 unit and $\tau_1 = 36$ months. Therefore, the problem scale in the base case is $n_1 = 17.21 \times 10^3 \times 36 = 619,610$, and we assume the firm reviews demand and makes capacity adjustment decisions in a monthly scale at the current demand level, i.e., $\tau_{n_1} = 1$ (recall that $\tau_{n_1} \asymp n_1^{-1/3}$.) We will illustrate the impact of market size on the performance of the heuristic in the numerical study. We also assume there is no leadtime, i.e., $l = 0$, and we will study the impact of leadtime later.

Initial capacity. Our target is to analyze Ford's capacity adjustment decision for the third generation. Therefore, besides the demand information, we also need information about the capacity. Since Ford does not publish their exact capacity, we use the maximum production quantity from January 2010 to December 2010 as the starting capacity, i.e., 22.97 thousand cars per month.[5]

Cost/profit parameters. We use aggregated cost parameters at the firm level to approximate the ones at the product level. Specifically, we recover the gross capacity of Ford using Ford's public financial reports and data, and then identify the unit profit and capacity related costs at the firm level. Although we acknowledge that these are rough estimates, the performance of our heuristic is quite robust to the

[5]According to Ford Motor Company (2012), the vehicle assembly capacity is categorized as *installed capacity* and *manned capacity*. Installed capacity refers to "the physical capability of a plant and equipment to assemble vehicles if fully manned". Manned capacity refers to "the degree to which the installed capacity has been staffed". In this numerical example, we use capacity to refer to the *installed* capacity that is specific to Ford Focus.

Table 2.3: Production and capacity related profit/cost parameters

Estimated cost parameters for Focus	Value
Capacity adjustment (upgrading) cost c_a	\$4,487 *month/unit*
Capacity adjustment (downgrading) cost γ_a	\$448.7 *month/unit*
Capacity overhead cost c_0	\$181.1 *per unit*
Capacity outsourcing cost c_1	\$362.2 *per unit*
Unit profit p (excluding capacity related cost)	\$1,270 *per unit*
Average retail price	\$22,154 *per unit*

cost parameters. Note that the cost parameters would be significantly more accurate if one extracts the cost information from an ERP or internal accounting system. We summarize the cost parameters derived from Ford's Annual Report in 2012 (Ford Motor Company, 2012) in Table 2.3, and relegate the details of estimations to Appendix A.

2.6.2 Numerical Analysis

In the numerical study, we evaluate the performance of our heuristics using the regret with respect to its deterministic upper bound. Specifically, for given scale parameter n, define $R_n^{ts} = 1 - V_{0,n}^{ts}/V_{0,n}^{d}$ and $R_n^{ms} = 1 - V_{0,n}^{ms}/V_{0,n}^{d}$ to be regrets associated with the two-step heuristic and multi-step heuristic, respectively. The decision horizon, demand distributions, initial capacity, and profit and cost parameters are the ones specified in Section 2.6.1. In what follows, we first present the impact of various parameters and demand assumptions (market size, leadtime, misspecified demand, and cost and profit parameters) on the performance of our two-step heuristic. We then compare the performance of the two-step heuristic with the multi-step heuristic. Finally, we show the performance of the two-step heuristic with respect to the optimal policy.

To evaluate the value-to-go function under the two-step heuristic, $V_{0,n}^{ts}$, for a given prior vector $\boldsymbol{\pi_1}$, we apply a simulation approach with 10^6 experiments. In each round, a demand distribution (a demand type) is first generated according to the prior, then a

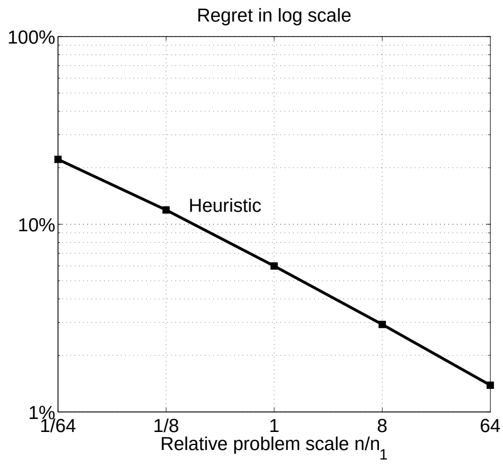

Figure 2.4: Regret with respect to market size when the prior $\boldsymbol{\pi_1} = (0.2, 0.4, 0.4)$

sample path of demand in each period is generated according to the distribution. For each sample path generated, the firm follows the two-step heuristic, and the resultant profit is calculated. We use the average of the 10^6 observations to approximate $V_{0,n}^{ts}$. The deterministic upper bound, $V_{0,n}^{d}$, is computed following equation (2.25).

Market size. We first analyze the impact of market size, which is determined by the scale factor n. From Proposition II.6, when the scale factor of the decision problem is n, setting the length of the learning period as $\tau_n \asymp n^{-1/3}$ results in asymptotic convergence at most on the order of $n^{-1/3}$. In Figure 2.4, we show that as $log(n)$ increases linearly, the log of the regret decreases linearly. In the base case ($n_1 = 619, 610$), we assume the firm reviews the demand information monthly, and adjusts capacity based on the observation in the first month, i.e., $\tau_{n_1} = 1$ month. To analyze the impact of the market size, we let n be $8^k n_1, k = -2, -1, ..., 2$, corresponding to a τ_n of $2^{-k}, k = -2, -1, ..., 2$ month respectively. That is, as the magnitude of demand increases, the firm can adjust capacity within a smaller window of demand data. For instance, when $k = -2$, τ_n is 4 months, and when $k = 2$, τ_n is 1/4 months (about 1 week). In Figure 2.4, we observe that the log of regret decreases at the slope of -0.33, corresponding to the $n^{-1/3}$ convergence rate. This implies that the absolute difference between the upper bound and the heuristic is sub-linear in n. The cases

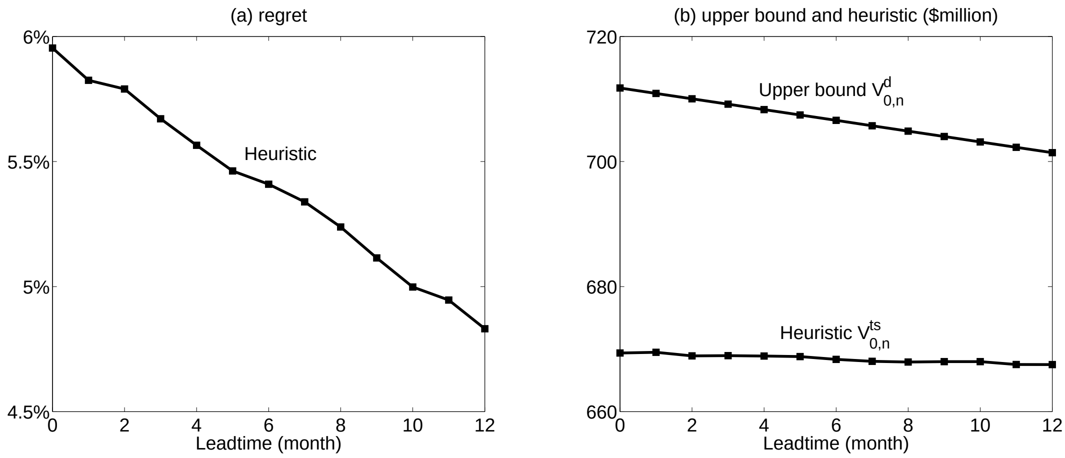

Figure 2.5: Regret and value-to-go function with respect to leadtime when the prior $\boldsymbol{\pi_1} = (0.2, 0.4, 0.4)$

are similar when the priors are different, so for the interest of space the details are not shown here.

Leadtime. In the base case, we normalize the leadtime as 0. One may think that this might favor the two-step heuristic, but the result is the opposite. The performance of the two-step heuristic improves as the leadtime becomes longer. To show this, we change the leadtime l from 0 to 12 months when the review period τ_{n_1} is 1 month and compute the total revenue of the planning horizon (i.e., the value function plus the revenue of the first l periods (before any adjustment is made). In Figure 2.5(a), we observe that the regret decreases as the leadtime increases: the relative profit loss due to the lack of information decreases in leadtime. Although this is counter-intuitive at first glance, we observe from Figure 2.5(b) that, with a longer leadtime, the benefit of full information decreases, thus the performance of the deterministic upper bound deteriorates substantially, resulting in the decrease in the regret.

Misspecified demand. The base case has assumed three demand types: low, medium and high. In the low demand scenario, we assume the average demand decreases to 12.21 thousand units per month. In the medium demand scenario, we

Table 2.4: Regret for the misspecified demand and cost parameters (The number of observations for each case is 147.)

Parameters	Regret			
	mean	stdev	min	max
low demand mean θ_l	6.02%	0.31%	5.78%	7.71%
high demand mean θ_h	5.81%	0.51%	4.32%	6.77%
outsourcing and overhead cost $\beta = (c_1 - c_0)/c_0$	6.08%	0.60%	4.99%	7.71%
downsizing and expansion cost $\gamma = \gamma_a/c_a$	6.01%	0.30%	5.45%	6.92%

assume that the demand remains at the same level as the demand for the first and second generation. In the high demand scenario, we assume the average demand has increased to 19.71 thousand units. However, these assumptions may not be accurate. Therefore, we now analyze the case when the firm has incorrect information about the demand type and resulting distribution. When calculating the deterministic upper bound, the firm still has complete information about the demand.

Note that our two-step heuristic does not depend on the firm's knowledge about the high type demand: The demand information is needed for evaluation and comparison only. In the analysis, we vary the average demand of low demand type from -20% to 40% in the increment of 10%, and similar for the high demand type. For each set of demand parameters, we also vary the prior as $(0.2i, 0.2j, 1 - 0.2i - 0.2j)$ where $i = 0, 1, ..., 5$ and $j = 0, 1, ..., 5 - i$. We summarize the test statistics in the first two rows of Table 2.4. We observe that the average regret with respect to the relaxed upper bound is only about 6% with a range less than 2.45%, which indicates the performance of the regret is quite robust with respect to the misspecified demand parameters.

Cost parameters. We also analyze the impact of the cost parameter changes on the two-step heuristic. In particular, we examine this by varying the relative difference between the outsourcing cost and capacity overhead cost: $\beta = (c_1 - c_0)/c_0$ fixing c_0, and the ratio of the downsizing cost to the expansion cost: $\gamma = \gamma_a/c_a$ fixing c_a. In our base case, we have $\beta = (362.2 - 181.1)/181.1 = 1$ and $\gamma = 448.7/4,487 = 0.1$ (see

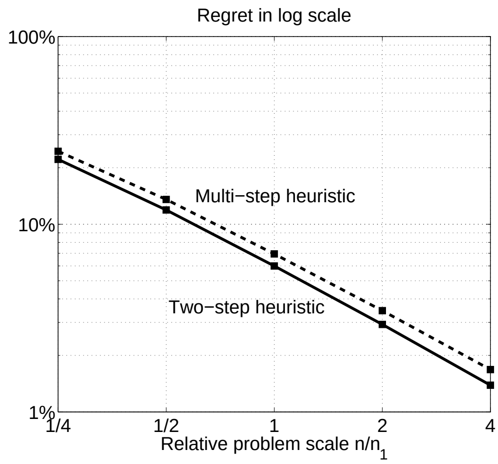

Figure 2.6: Two-step heuristic vs. multi-step heuristic when the prior $\boldsymbol{\pi_1} = (0.2, 0.4, 0.4)$

Table 2.3). Similarly to the misspecified demand scenario, we also vary the prior as $(0.2i, 0.2j, 1-0.2i-0.2j)$ where $i = 0, 1, ..., 5$ and $j = 0, 1, ..., 5-i$. In a quite broad range of β (from 0.7 to 1.3) and γ (from -0.3 to 0.3), the regret does not change in any significant manner (see the third and fourth row of Table 2.4). These results show that the heuristic is quite robust with respect to the cost parameters, as the increase in the regret is smaller than 2.72% when the cost parameters and the prior vary.

Single vs. multiple adjustments. We now compare the two-step heuristic with the multi-step heuristic. The capacity adjustment cost is specified in Section 2.6.1. In Figure 2.6, we observe that as the market size increases, the regrets of both policies decrease. In this case, as the firm needs to pay a much higher adjustment cost under the multi-step heuristic, which dominates the benefit from extra opportunities to adjust capacity, we observe that the regret under the multi-step heuristic is higher than the one under the two-step heuristic. However, when the capacity adjustment cost is small, as one may expect, the regret under the multi-step heuristic is lower than the one under the two-step heuristic, which reflects the benefit of learning-while-doing.

Heuristic vs. optimal policy. To simplify the computation for the optimal

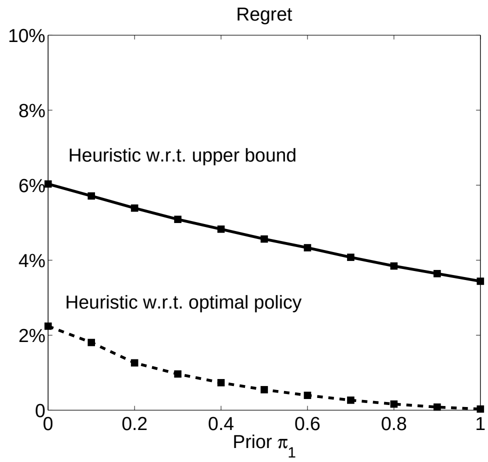

Figure 2.7: Regret of the two-step heuristic with respect to upper bound and optimal policy. There are only two demand types here and π_1 indicates the prior of the demand being high type

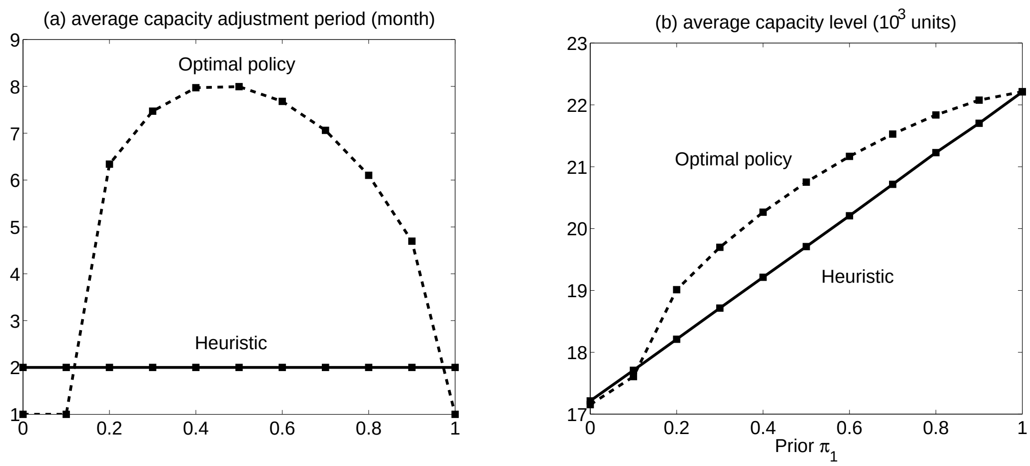

Figure 2.8: Firm's capacity decision under the two-step heuristic and optimal policy

policy, we consider only two demand types in this part: *medium* and *high*. As there are only two demand types, we use π_j, the posterior distribution of high demand, to denote the information vector. Figure 2.7 shows the regret of the two-step heuristic. Compared to the deterministic upper-bound (which assumes the knowledge of full information and no randomness), the regret of our data-driven heuristic is no more than 6.03%. We use the deterministic upper bound to define the regret, because a large state-space makes it intractable to compute the optimal policy and resultant value function. In the two demand-type case, however, we can numerically approximate the value function of the optimal policy, $V^*_{0,n}$, with linear interpolation (i.e., evaluating the value at a set of fine fixed grid points and then approximating values for the rest of the states using linear interpolation). As Figure 2.7 shows, the regret (compared to the optimal policy) is less than 2.24%.

From the timing perspective (Figure 2.8(a)), the firm always adjusts its capacity in the second month under the two-step heuristic. On the other hand, under the optimal policy the firm adjusts capacity early (in the first period) when the prior is close to the extremes (π_1 close to 0 or 1), and delays the decision when there is no dominant demand type in the prior. In addition, Figure 2.8(b) shows that, on average, the capacity levels under the optimal policy and the two-step heuristic are fairly close when the firm adjusts the capacity at the beginning of the decision period, because the newsvendor fractile is 0.5 as determined by the capacity outsourcing cost and overhead cost and therefore the optimal capacity level is close to the average demand. When the firm is less certain about the demand type and prefers to delay the capacity adjustment to the future, consistent with the conventional wisdom, the firm invests more in capacity compared to the average capacity level built under the two-step heuristic.

2.7 Conclusion

We analyze a firm's capacity investment decision for a product with a finite life cycle, and investigate ***when***, and by ***how much***, the firm should adjust its capacity. When the firm can adjust the capacity once in a planning horizon, we show that in each period, as the likelihood of demand being high increases, interestingly, the firm may alternate its decision to pull the trigger (adjust capacity) or delay the adjustment multiple times. This contrasts most of the results in a stopping problem where the optimal decision to stop tends to be monotone. On the other hand, if the firm decides to adjust the capacity, the target capacity level increases in the likelihood. While the structure of optimal policy is quite interesting from the analytic point of view, it cannot be easily computed or implemented. Instead, we show that there is a very simple but provably well-performing data-driven heuristic when demand follows a stochastic process with stationary and independent increment. In this heuristic, the firm observes demand during an exploration period, and then adjusts capacity to match the observed demand rate. By choosing an appropriate exploration period length, the firm is able to balance the exploration and exploitation tradeoff, and the regret of the heuristic asymptotically converges to 0.

When the firm has multiple opportunities to adjust capacity, we show the firm's optimal policy is a control band policy, characterized by two state-dependent thresholds. Under this policy, in each period, the firm stays put to observe the demand when the capacity is between the two thresholds, and adjusts its capacity to the lower threshold only when the capacity is below it, and vice versa. We also characterize a simple but asymptotically optimal heuristic, in which the firm predetermines a set of time points at which the firm will adjust its capacity to match the observed demand rate. The time between two consecutive decisions increases exponentially, reflecting the fact that the adjustment is costly, and it is less necessary for the firm to adjust capacity frequently with more demand information collected. The multiple adjustments enable the firm to correct errors in early decisions. However, when the capacity adjustment cost is high, the multiple adjustments also yield a higher adjustment cost, which dilutes the benefit of the learning-while-doing. The optimal policy and heuristics are illustrated using production and sales data of the Ford Focus.

CHAPTER III

Investing in a Shared Supplier in a Competitive Market: The Stochastic Capacity Case

3.1 Introduction

In many supply chains, multiple firms source from the same set of suppliers. Such supply chain structures benefit from achieving economies of scale, and obtaining reliable and high-quality supply, but there are also risks such as the firms being exposed to shortage of supplies or greater vulnerability to supply disruptions. In order to mitigate the risks, many firms invest in shared suppliers, even if they compete against each other.

For example, firms may invest in the supplier to expand the supplier's capacity, avoid the supplier's bankruptcy, or improve the quality of products. Neutrogena directly invested in its South Korean supplier, Cosmax, which also served many other cosmetic companies. Intel invested in ASMI by purchasing 4% of its total common shares to foster material and equipment development (LaPedus, 2009), even though ASMI is also a supplier to AMD, Intel's main rival. Of course, firms invest in suppliers not just for capacity expansion; for instance, GM provided $210 million to AAM in 2009 to help keep it out of bankruptcy (Haywood, 2009), and Walmart sent teams of experts to help Chinese suppliers improve sustainability efforts while these suppliers also supplies to other stores in US (Aston, 2009). These investments can broadly be considered under the umbrella of "supplier development" (Handfield et al., 2000).

When a firm invests in a supplier which also serves its competitors, there is a natural competitive threat that arises: the competitors may be able to take advantage of the original firm's investment in the supplier, which would intensify the market competition for the end product and therefore reduce the investing firm's profits. To avoid this, the investing firm may impose contractual constraints on the supplier that dictate exactly how the increased capabilities of the supplier can be used. In this chapter, we specifically focus on the contractual relationships governing firms' investments in their supplier's capacity and their consequences.

It is worth considering Foxconn's recent investment in Sharp to illustrate the framing of our model. In early 2012, it was reported that Foxconn (also known as Hon Hai Precision Industry Co.) invested $1.6 billion in Sharp: very specifically, the investment included a 46.5% stake in a single LCD factory in Sakai, Japan, and an agreement to buy 50% of the LCD panels produced in that factory (Dignan, 2012). In this case, Foxconn claimed exclusive use of the 50% capacity, while Sharp is free to use the remaining 50% to supply other buying firms, including Sharp's own products in the smartphone/tablet/TV market which remains highly competitive.

Another motivation to invest in a shared supplier is to prevent other firms from receiving preferential treatment when fulfilling a contract. When a significant portion of the supplier's capacity could be first tapped by the investing firm, the non-investing firm's ability to accrue profits from meeting the demand will be reduced, relatively. Such concern is manifested when competing buyers share a supplier. This can partly explain Samsung's involvement in Sharp shortly after Foxconn's investment. Samsung also invested $110 million in Sharp, in order to "prevent its competitor, particularly Hon Hai and Apple, from gaining too much control over Sharp", and "secure a steady supply of LCD panels from diversified sources" (Osawa and Lee, 2013).

In this chapter, we consider a network of two firms competing in the market and sharing a common supplier, and examine the question of how the different contractual forms affect the firms' investment in the supplier's capacity, and the consequences of such investments for the buying firms and the supplier. While it is straightforward to evaluate if such investment is beneficial when the supply chain consists of one supplier and one buyer, it is not clear whether such an investment is beneficial to a firm when a supplier serves multiple firms. Investing in the supplier's capacity can provide a buyer increased access to the supplier's capacity, but such investment is costly and also may benefit the non-investing firm if it also has more access to the capacity (a spillover effect).

To prevent unintended spillover, buyers' investment often comes with conditions such as exclusive use or prioritized access of the production resources. While restricting the autonomy of the supplier's operations, a buying firm's investment can ensure it has enough supply to satisfy the demand. Therefore, for both supplier and buyer, the economics of supplier investment becomes more complex with presence of competing buying firms. In this chapter, we develop a model that captures the investment decisions and market competition, and study the consequences of the contracts that accompany the investments in terms of profits to the buyers and the supplier.

Our contributions. We examine two common forms of contracts used in practice: *Exclusive* (the investing firm gets exclusive access to a portion of the supplier's capacity that cannot be used by the non-investing firm), and *First-Priority* (the non-investing firm can access the unused portion of the investing firm's capacity, if any). We completely characterize equilibrium outcomes in terms of the number of investing firms and capacity investment levels. In equilibrium, the number of investing firms decreases as the fixed capacity investment cost increases, and within a regime where the number of investing firms remains the same, the capacity level decreases in the variable capacity cost.

Specifically, we identify when and to what extent the spillover effect occurs. The spillover effect occurs when the capacity type is first-priority, and the fixed capacity investment cost is intermediate. The extent to which the spillover effect occurs is critically determined by the variable capacity cost, as it determines how much capacity the investing firm will invest to build and consequently how much capacity the competing firm will be able to tap into.

We next examine the impact of the spillover effect on the supply chain performance by comparing the exclusive and first-priority capacity. We observe that up to a certain extent, the spillover curbs competition between the firms and therefore discourages them from investing in the supplier. As a result, the equilibrium first-priority capacity is lower than the equilibrium exclusive capacity, and so is the number of investing firms. On the other hand, the spillover effect mitigates the risk of both firms being trapped in a prisoner's dilemma, resulting in a better outcome for both firms.

Given the observation about the equilibrium capacity levels, we further explore the buying firms' and the supplier's preference about the capacity types. We find the buying firms' preference is determined by two effects: the *leading* effect as the firm is the only investor and has advantages in accessing capacity, and the *spillover* effect where the non-investing firm can tap into the investing firm's capacity. We show that the investing firm does not always prefer the exclusive capacity. The investing firm typically prefers the exclusive capacity. However, when the exclusive capacity triggers the other firm also to invest, the firm loses its benefit from the leading effect and therefore may prefer the first-priority capacity. The non-investing firm, however, always prefers the first-priority capacity in the hope of getting benefit from the spillover effect. We also find that the supplier's preference is driven by the tradeoff between the *over-investment* in the exclusive capacity, and the *flexibility* in utilizing the first-priority capacity. We observe that the supplier finds the exclusive capacity more attractive when strictly more firms invest with the exclusive capacity due to the significant over-investment in capacity. When only one firm invests under both exclusive and first-priority contracts, the supplier's preference is determined by which of the two effects is stronger and depends on the context.

Finally, we compare the supply chain performance relative to a first-best benchmark where the downstream firms are combined as a monopoly and are vertically integrated with the supplier, that is, one firm owns both levels of the supply chain. We analyze the two sources of inefficiency: competition, and misaligned incentive with non-zero wholesale price (double marginalization). We identify the inefficiency caused by competition in the supply chain, and find that the spillover effect can partially mitigate the over-investment in exclusive capacity and therefore improve the efficiency of the supply chain. We also identify the impact of the wholesale price, which also decreases the equilibrium capacity level. Therefore, we find that both the wholesale price and the spillover effect can reduce the over-investment associated with the exclusive capacity.

3.1.1 Literature review

Our work falls within the literature of supplier development, which refers to the set of activities undertaken by a buyer to identify, measure, and improve supplier performance (Krause et al., 1998). These activities have been identified and studied along the dimensions of the level of efforts committed by the buying firms (Krause, 1997), whether supplier development is a reactive or strategic process (Krause et al., 1998), and how supplier development influences accumulation and allocation of social capital (Krause et al., 2007). Handfield et al. (2000) provided a taxonomy of supplier development, and identified various pitfalls that may occur.

Many papers have studied different ways to improve supply chain efficiency by firms' investment in suppliers such as cost reduction (Iyer et al., 2005), quality improvement (Zhu et al., 2007), capacity investment (Li and Debo, 2009; Li, 2013), reliability improvement (Wang et al., 2010), and financial subsidy (Babich, 2010). In all these papers, there is a *single* buyer with one or more suppliers; none of the above papers consider the case of competing firms investing in a shared supplier.

Two papers studied multiple firms investing in a shared supplier to improve the reliability. Wadecki et al. (2011) proposed a model in which firms may share an unreliable supplier, whose probability of disruption can be reduced through firms' subsidy, and showed that lower subsidies are likely to be offered when firms compete. Wang et al. (2014) considered the effect of knowledge spillover when one or both firms invest in a shared supplier, and showed spillover often improves the firms' profits. In both papers, random yield is considered and capacity competition is absent. In contrast, the supplier's limited capacity, and the amount of capacity that can be accessed by each firm under different contract structures, are key features of our model. Thus, we model capacity as a finite random variable and examine how the firms' investment decisions are influenced by how much capacity the firm and its competitor can access.

Our work is also related with the capacity management literature, about which Van Mieghem (2003) provided a review. Among the papers that address capacity issues in an outsourcing setting, Plambeck and Taylor (2005) analyzed the impact of contract manufacturing on firms' innovation and the supplier's capacity investment, and showed that firms might be better off by trading capacity among themselves rather than outsourcing to a supplier. Ülkü et al. (2005) analyzed firms' time of entry in an uncertain market in an outsourcing setting, and showed that firms may subsidize the supplier's capacity investment to accelerate the entry process. Ülkü et al. (2007) analyzed a model in which firms and a supplier differ in forecast accuracy. They investigate premium-based schemes to induce the best party, firms or the supplier, to bear the risk of building the capacity, and concluded that the schemes work well. Li et al. (2011) analyzed three capacity reservation options, in which the firms may or may not access the other firm's reserved capacity and may or may not need to pay a fee to access. They investigated which option is preferred from the supplier's perspective. Compared with these papers, our work differs in two aspects. First, competition is a central theme in our work. In fact, firms' demands are independent in all these papers expect for Li et al. (2011) where the demand for firms might be correlated. In contrast, the fact that firms compete in the end market is a key feature of our model. Second, these papers feature deterministic capacity, while in our model the capacity is stochastic. The stochastic capacity arising from uncertain yields or technology motivates firms to invest in the supplier.

3.2 Model

We consider a supply chain with a single supplier (denoted by s), and two competing firms, labeled 1 and 2. The supplier will produce a new product (or critical component), and sell it at a wholesale price c to both firms. Thus, in the absence of capacity investment, both firms are ex ante symmetric. The two firms compete *á la* Cournot in the downstream market with the following inverse demand function: $P(q_1, q_2) = a - b(q_1 + q_2)$, where a represents the total market size, and b the price sensitivity to the total quantity.

To fulfill the buying firms' orders, the supplier needs to build capacity. Prima facie, the supplier incurs the investment cost for the capacity. In this stage, buying firms have opportunities to invest in the supplier's capacity. In return for the investment, the firms are endowed with a portion of capacity for use, for which the firms can impose restrictions that limit how the supplier uses. Details of these restrictions will determine the amount of the capacity that the firms can use at their discretion, and resultantly, the output produced for the two firms.

To capture this, we consider a two-stage model. In the first stage, the firms decide whether to invest in the supplier, and if so, by how much. In the second stage, based on the outcome of the investment game, the firms set their order quantities and compete in the end market. We consider a case where the firm i's capacity investment cost includes both a *fixed* cost, w_0, and a *variable* cost, wk_i, where w is the unit variable cost and $k_i \in \mathbb{R}^+$ measures the size of the capacity investment level that a supplier will have if there is no yield or efficiency loss. For example, the fixed capacity investment cost may represent the fixed cost associated with commissioning and starting up a new facility with the firms' investment in the supplier, and the variable cost may represent the cost to purchase tooling or hire new workers, which is proportional to the size of the invested capacity. This structure of capacity investment cost is also observed in the literature, for example, in Van Mieghem (2003) and Ye and Duenyas (2007).

We assume that at the time of investment, firms do not know exactly the yield of capacity, thus the capacity level after the investment has uncertainty. This reflects the fact that many factors besides capital investment (e.g., physical capacity, available technology, process yield, and staffing plan) influence the actual capacity at the time of production. This is consistent with the stream of literature where the supply is unreliable, see e.g., Wadecki et al. (2011) and Wang et al. (2014). This setting is valid in several contexts such as high tech capacity installation, agriculture, vaccine production, etc. For instance, in AMOLED manufacturing, the production yield is largely unknown at the time of capacity investment. Even when they are close to reach the threshold of mass production capability, the acheived yield is still below 70% [1]. In agriculture for example, while firms may invest to increase the supplier's capacity, the realized capacity is random until the harvest of the produce. There exists another stream of literature where the capacity is deterministic and the demand is stochastic. For example, one may refer to Plambeck and Taylor (2005) and Li et al. (2011).

[1]http://www.oled-info.com/reports-korea-suggest-both-lgd-and-sdc-increased-oled-tv-production-yields

Specifically, the supplier's total capacity is $K_s = (k_0+k_1+k_2)\boldsymbol{\xi}$, where k_0 represents the supplier's base capacity, k_i $(i = 1, 2)$ is the capacity level invested by firm i, and $\boldsymbol{\xi}$ is a random variable with support $[0, 1]$ reflecting the yield between the realized capacity and the theoretical maximum capacity projected at the time of investment. It follows that the supplier's total capacity increases in the first order stochastic sense as the buying firms' investment increases. The fact that all the three parts of the supplier's capacity, $k_0\boldsymbol{\xi}$, $k_1\boldsymbol{\xi}$, and $k_2\boldsymbol{\xi}$ are positively correlated, reflects that these capacities are built for the same product at the supplier's site, which are subject to the same risk of random yield or supplier disruption.

In return for the investment, the firms impose contractual restrictions that limit how the installed capacity should be used. While there are many different forms of restriction used in practice, one widely-used form is an "exclusive" contract: the investing firms demand the exclusive use of the firm-invested capacity, and disallow the supplier from accessing the invested capacity to serve other firms even if the order quantities leave some of that capacity unused. Another widely used form is that the investing firms demand to fulfill their orders first ("first priority"), but the supplier is free to use any leftover. One of the central questions we ask is how these restrictions impact firms' investment and quantity decisions.

If neither firm invests in the supplier, then both firms are identical from the perspective of the supplier. In this case, no such priority will be given to either firm. If the total quantity ordered by both firms is less than the supplier's capacity, allocation is trivial. When the total quantity is greater than the supplier's capacity, we follow Cachon and Lariviere (1999) and use the *uniform allocation* rule (also see Sprumont (1991)). Under the uniform allocation rule, if one firm orders more than the other firm, it will receive the minimum of its own order quantity and the capacity left from serving the other firm. Unlike other allocation rules that can induce order inflation (see Cachon and Lariviere (1999) for details), the uniform allocation rule is known to be truth-inducing, and non-manipulable; it results in both firms ordering their preferred quantities within the limits of an upper bound that enforces that the total quantity does not exceed the capacity. We assume that the uniform allocation rule is common knowledge in our game, so both downstream firms are incentivized to order their optimal quantities under all cases of investment. Since we assume that the two buying firms are ex ante identical, if neither firm invests, the supplier is effectively allocating one-half of the capacity to each firm in equilibrium; if the firms' order quantities exceed capacity, they will each get one-half following the uniform allocation rule. Therefore, in the case of neither firm investing, we use ξ to represent the realization of $\boldsymbol{\xi}$, and then firm i's capacity reservation level is $\frac{k_0}{2}\xi$.

The exclusive contract is now defined as firm 1 reserves a capacity level $(\frac{k_0}{2}+k_1)\xi$ exclusively. If firm 1 does not fully use the allotted capacity, the remaining capacity will be wasted. Consequently, firm 2 is allocated a capacity of $(\frac{k_0}{2}+k_2)\xi$, which is the maximum quantity that firm 2 can order. On the other hand, in the first priority contract, firm 1 is allowed to order first, but the other firm can place an order up to the remaining capacity: $(\frac{k_0}{2}+k_2)\xi$ plus leftover capacity from firm 1, if any. When only one firm or both firms invest, without loss of generality, we assume $k_1 \geq k_2$, i.e., the capacity investment size by firm 1 is greater than or equal to the investment size by firm 2.

The remaining sequence of events is as follows. After the capacity has been realized, firms place their orders (with quantities specified by q_1 and q_2) subject to the above capacity constraints and compete in the downstream market. Therefore, firm i's second-stage profit is given as $q_i P(q_1, q_2) - cq_i$. Likewise, the supplier's second-stage revenue is simply $c(q_1 + q_2)$.

The total profit of each firm is the expected second-stage profit, minus investment cost, if any. We assume that the buyer's investment is entirely used for building capacity, so the supplier's overall profit is simply the expected value of $c(q_1 + q_2)$.

Note that relaxing this assumption does not change our results. For ease of exposition, we assume that neither the firms nor the supplier incurs any other costs, although incurring additional unit production costs at both firms will not change the analytical findings. We assume that the buyers and supplier are profit-maximizing, risk-neutral agents and all game parameters are common knowledge.

Remark on sequence of events. In our base model, we frame the sequence of events as firms first invest to improve the stochastic capacity at the supplier, then place an order *after* the capacity is realized. However, an alternative framework to interpret the sequence of events is that firms first invest in and order from the supplier. Then, after the capacity uncertainty is resolved, firms may choose not to use up all the ordered components and only utilize a portion of them to produce the final products at a production cost of c per unit. The second interpretation reflects the semiconductor or hightech industry practice where firms place orders before the capacity yield is realized or the production runs are completed. For ease of exposition, we follow the first interpretation in the description of our model.

3.3 Exclusive capacity contract

We first analyze the case where the capacity contract is *exclusive*, i.e., the invested capacity cannot be accessed by the other firm. To determine an equilibrium, we solve the game by backward induction. Therefore, we first present the analysis for the second-stage ordering subgame for given realized capacity $k_s = (k_0 + k_1 + k_2)\xi$ and first-stage investment decisions, then solve the first-stage investment game.

3.3.1 Second-stage quantity game

As we have introduced in Section 4.2, let (k_1, k_2) denote the capacity investment sizes of the two firms. Given the realized supplier capacity $k_s = (k_0 + k_1 + k_2)\xi$ and investment decisions in the first stage, firms decide their order quantities (q_1, q_2)

subject to realized capacity and restrictions placed in return for investments. Hence, for given (q_1, q_2), the market clearing price, $P(q_1, q_2)$, is given by $P(q_1, q_2) = a - b(q_1 + q_2)$. The equilibrium production quantities are determined by solving the following:

$$\pi_1^e(k_1, k_2, \xi) = \max_{q_1 \le (\frac{k_0}{2} + k_1)\xi} q_1 P(q_1, q_2) - cq_1; \pi_2^e(k_1, k_2, \xi) = \max_{q_2 \le (\frac{k_0}{2} + k_2)\xi} q_2 P(q_1, q_2) - cq_2.$$

The equilibrium order quantities and resultant profits are presented in Lemma III.1, in which we assume $k_1 \geq k_2$ without loss of generality. For ease of exposition, throughout the chapter, we use $m_0(k_1, k_2)$ to represent the unit margin if neither firm is constrained by the invested capacity (k_1, k_2) when ordering optimally. Likewise, under a subgame perfect equilibrium $m_1(k_1, k_2)$ represents the unit margin if only one firm is constrained by the invested capacity (k_1, k_2). As we assume $k_1 \geq k_2$ and firms compete in the Cournot market, if only one firm is constrained, that firm can only be firm 2. Similarly, we use $m_2(k_1, k_2)$ to represent the unit margin if both firms are constrained.

These are the following:

$$\begin{aligned}
m_0(k_1, k_2, \xi) &= \frac{a - c}{3} \\
m_1(k_1, k_2, \xi) &= \frac{a - b(\frac{k_0}{2} + k_2)\xi - c}{2} \\
m_2(k_1, k_2, \xi) &= a - b(k_0 + k_1 + k_2)\xi - c
\end{aligned}$$

For simplicity, we suppress the dependency of $m_i(k_1, k_2, \xi)$ on (k_1, k_2, ξ) when there is no confusion. All proofs are relegated to the appendix.

Lemma III.1 (Firms' equilibrium order quantity and ex post profit)**.**
Let the capacity investment sizes be (k_1, k_2). The resulting subgame yields the following:

realized yield ξ	*order quantity* (q_1^*, q_2^*)	*ex post profit* $(\pi_1^e,\ \pi_2^e)$
$0 \le \xi \le \frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}$	$\big((\frac{k_0}{2}+k_1)\xi, (\frac{k_0}{2}+k_2)\xi\big)$	$\big(m_2(\frac{k_0}{2}+k_1)\xi, m_2(\frac{k_0}{2}+k_2)\xi\big)$
$\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)} < \xi \le \frac{a-c}{3b(\frac{k_0}{2}+k_2)}$	$\left(\frac{a-c-b(\frac{k_0}{2}+k_2)\xi}{2b}, (\frac{k_0}{2}+k_2)\xi\right)$	$\left(\frac{m_1[a-c-b(\frac{k_0}{2}+k_2)\xi]}{2b}, m_1(\frac{k_0}{2}+k_2)\xi\right)$
$\frac{a-c}{3b(\frac{k_0}{2}+k_2)} \le \xi \le 1$	$\big(\frac{a-c}{3b},\ \frac{a-c}{3b}\big)$	$\left(\frac{m_0(a-c)}{3b},\ \frac{m_0(a-c)}{3b}\right)$

We note that the lemma illustrates the case where $\frac{a-c}{3b(\frac{k_0}{2}+k_2)} \le 1$. For the cases where $\frac{a-c}{3b(\frac{k_0}{2}+k_2)} > 1$, the analysis is exactly the same except that one or two regions in the table are empty. We discuss these cases in Appendix B.

3.3.2 First-stage investment game

Building on Lemma III.1, we now analyze the firms' investment decisions in the first stage. In this stage, the firm needs to make capacity investment decisions: whether to invest in the supplier, and if so, by how much.

Before proceeding to the equilibrium analysis, we first define firms' expected profit, given firms' capacity investment size (k_1, k_2). For notational simplicity, we define an indicator $\mathbf{1}_C$: $\mathbf{1}_C = 1$ if condition C is met and 0 otherwise. Let $V_i^e(k_1, k_2)$ be the firm i's expected profit when the capacity investment sizes of the two firms are (k_1, k_2) and the two firms will follow a subgame perfect ordering strategy in the second stage. Using these terms and applying the results from Lemma III.1, we write $V_i^e(k_1, k_2)$ as follows.

$$
\begin{aligned}
V_1^e(k_1,k_2) &= E\left[\pi_1^e(k_1,k_2,\boldsymbol{\xi})\right] - w_0\mathbf{1}_{\{k_1>0\}} - wk_1 \\
&= E\left\{\mathbf{1}_{\left\{\boldsymbol{\xi}\le\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}\right\}}\left[m_2\left(\frac{k_0}{2}+k_1\right)\boldsymbol{\xi}\right] + \mathbf{1}_{\left\{\boldsymbol{\xi}>\frac{a-c}{3b(\frac{k_0}{2}+k_2)}\right\}}\left[\frac{m_0(a-c)}{3b}\right]\right. \\
&\qquad \left. + \mathbf{1}_{\left\{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}<\boldsymbol{\xi}\le\frac{a-c}{3b(\frac{k_0}{2}+k_2)}\right\}}\left[\frac{m_1[a-c-b(\frac{k_0}{2}+k_2)\boldsymbol{\xi}]}{2b}\right]\right\} \\
&\quad - w_0\mathbf{1}_{\{k_1>0\}} - wk_1 \\
&= \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}} m_2\left(\frac{k_0}{2}+k_1\right)\xi f(\xi)d\xi + \int_{\frac{a-c}{3b(\frac{k_0}{2}+k_2)}}^{1} \frac{m_0(a-c)}{3b} f(\xi)d\xi \\
&\quad + \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}}^{\frac{a-c}{3b(\frac{k_0}{2}+k_2)}} \frac{m_1[a-c-b(\frac{k_0}{2}+k_2)\xi]}{2b} f(\xi)d\xi - w_0\mathbf{1}_{\{k_1>0\}} - wk_1
\end{aligned} \tag{3.1}
$$

$$
\begin{aligned}
V_2^e(k_1,k_2) &= \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}} m_2\left(\frac{k_0}{2}+k_2\right)\xi f(\xi)d\xi + \int_{\frac{a-c}{3b(\frac{k_0}{2}+k_2)}}^{1} \frac{m_0(a-c)}{3b} f(\xi)d\xi \\
&\quad + \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}}^{\frac{a-c}{3b(\frac{k_0}{2}+k_2)}} m_1\left(\frac{k_0}{2}+k_2\right)\xi f(\xi)d\xi - w_0\mathbf{1}_{\{k_2>0\}} - wk_2
\end{aligned} \tag{3.2}
$$

For both firms, the (expected) profit is the sum of three terms, minus investment costs. The first term represents the profit when the realized yield is small and the resultant capacity is binding. In this case both firms order and use up all the available quantity. The second term represents the firm's profit when the realized capacity is sufficiently large and neither firm is bounded by its capacity constraint. The third term represents the case where firm 2 is capacity-constrained while firm 1 which invested more capacity is not. With the induced profit functions, we find the equilibrium by analyzing the corresponding game between the two firms.

Note that there are three possible equilibrium regimes: *neither firm invests, one firm invests*, or *both firms invest.* In each regime, the investing firm(s) will also decide how much to invest in the supplier's capacity. If neither firm investing and both firms investing are both equilibria, we use the equilibrium that the firms remain at the status quo and do not invest in the supplier. This is because the status quo of the game is that neither firm invests in the supplier, and it is natural to choose this equilibrium as a focal equilibrium. We next evaluate the corresponding profits and identify the conditions under which each specific equilibrium arises. We characterize how the equilibrium evolves from one to another with respect to the fixed investment cost, w_0, and the variable investment cost, w, as follows.

Proposition III.2 (Firm's equilibrium capacity investment: Exclusive capacity). *There exist two equilibrium switching curves, $\underline{w}_0^e(w)$ and $\overline{w}_0^e(w)$, such that $\underline{w}_0^e(w) \leq \overline{w}_0^e(w)$, and*

i) When the fixed cost w_0 is small ($w_0 \leq \underline{w}_0^e(w)$), both firms invest k^e in the supplier. The equilibrium capacity, k^e, is such that

$$k^e \triangleq \left\{ k : \int_0^{\frac{a-c}{3b(\frac{k_0}{2}+k)}} \left[a - c - 3b\left(\frac{k_0}{2} + k\right)\xi \right] \xi f(\xi) d\xi - w = 0 \right\}, \quad (3.3)$$

which decreases in w. Furthermore, there exists a function $w_0^e(w)$ such that the equilibrium leads to a prisoner's dilemma in the region for all $w \in [w_0^e(w), \underline{w}_0^e(w)]$. Under the prisoner's dilemma we have $V_i^e(k^e, k^e) \leq V_i^e(0,0)$ for $i = 1, 2$.

ii) When the fixed cost w_0 is intermediate ($\underline{w}_0^e(w) < w_0 \leq \overline{w}_0^e(w)$), only one firm (labeled as firm 1) invests in the supplier, where the equilibrium

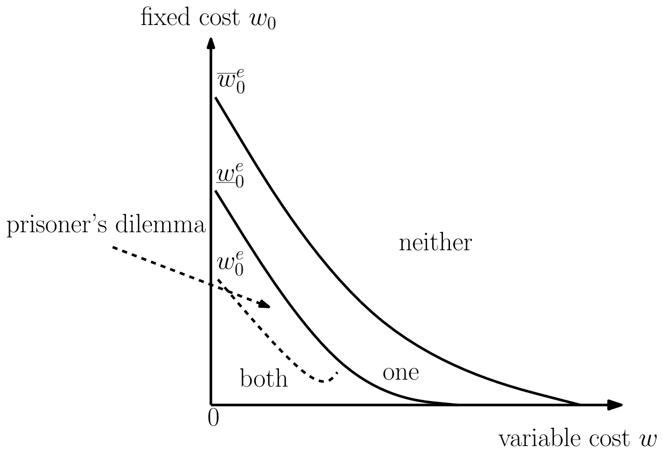

Figure 3.1: Equilibrium investment outcomes with exclusive capacity. "neither": neither firm invests. "one": one firm invests. "both": both firms invest.

capacity, k_1^e, is

$$k_1^e \triangleq \left\{ k_1 : \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1)}} \left[a-c-b\left(\frac{3k_0}{2}+2k_1\right)\xi \right] \xi f(\xi) d\xi - w = 0 \right\}, \tag{3.4}$$

which is decreasing in w.

iii) When the fixed cost w_0 is high ($w_0 > \overline{w}_0^e(w)$), neither firm invests in the supplier.

Furthermore, $\underline{w}_0^e(w)$ and $\overline{w}_0^e(w)$ decrease in the variable cost w.

The equilibrium can be seen in Figure 3.1. For a given variable capacity cost w, we observe that the number of investing firms decreases as the fixed capacity cost w_0 increases. When the fixed capacity investment cost is fairly low, i.e., $w_0 \leq \underline{w}_0^e(w)$, the entry barrier to invest in the supplier is low, and therefore, both firms invest in the supplier's capacity. The level of invested capacity, however, is determined by the variable cost of the capacity investment w, as shown in equation (3.3). What is surprising is, when the fixed capacity investment cost is close to the threshold $\underline{w}_0^e(w)$, as both firms find it dominant to invest in the supplier, both firms can be trapped in a prisoner's dilemma, where both firms earn a lower profit than what they would have earned if neither firm had invested in the supplier's capacity. While over-investment in capacity also occurs when $w_0 \leq w_0^e$, in this case, the fixed investment cost is not big enough to eat up the firms' profit too much, thus preventing the firms from being trapped in a prisoner's dilemma. The prisoner's dilemma equilibrium arises as a combination of both over-investment in capacity and the non-negligible fixed investment cost.

When the fixed capacity cost is in the intermediate range, i.e., $\underline{w}_0^e(w) < w_0 \leq \overline{w}_0^e(w)$, an asymmetric equilibrium where only one firm chooses to invest in the supplier is sustained. In this regime, the investing firm is able to gain enough profit being the capacity leader in the market while the non-investing firm finds it not necessary to invest in the supplier because the gain from investing in the supplier is limited and not enough to cover the associated investment costs.

When the fixed capacity cost is very high, i.e., $\overline{w}_0^e(w) < w_0$, neither firm invests in the supplier's capacity, and each firm will rely on the supplier's base capacity. The transition of equilibrium from two firm investing to one firm investing to neither firm investing is illustrated in Figure 3.1.

For a given fixed capacity cost w_0, we also observe that the number of investing firms decreases as the variable capacity cost w increases. To understand this, we also need to understand the impact of increased variable cost w on the buying firm's profit. When both firms invest in the supplier, an increase in w has both a direct and an indirect impact on the buying firms' profits. The direct impact of a higher variable cost is that the firms will have less incentive to invest in the supplier's capacity (k^e decreases in w,) because it is more expensive to do so, and as a result, the direct effect negatively affects the buying firm's capacity and profit. The indirect impact of a higher variable cost is that as the capacity of both firms decreases, it increases the market price and therefore the indirect effect positively affects the buying firm's profit. Therefore, whether an increase in the variable cost will lead to increase or decrease in the buying firm's profit will depend on the relative magnitude of the two effects, and this is why we may observe the non-monotonicity in the lower border for the prisoner's dilemma region, as indicated by the dashed line in Figure 3.1.

When only one firm invests in the supplier, a higher variable cost discourages the investing firm from investing in the supplier's capacity (k_1^e decreases in w,) and the investing firm's profit decreases. Therefore, if the investing firm's profit is less than the profit with zero firm investing at a given variable cost w, then this will still be the case as the variable cost becomes even higher. Thus we show that the switching curve $\overline{w}_0^e(w)$ decreases in w. For the non-investing firm, however, the decrease in the invested capacity lowers the quantity available to the market and raises the market price. In addition, the non-investing firm cannot access the investing firm's leftover capacity with the exclusive capacity anyway. Therefore, the non-investing firm's profit increases with respect to w. We observe that as w increases, this increase in the non-investing firm's profit when only one firm invests in the supplier outweighs the possibility to increase the firms' profit when both firms invest in the supplier. Thus we observe the monotonicity in the equilibrium switching curve $\underline{w}_0^e(w)$.

3.4 First-priority capacity contract

Another form of restriction is to claim *first-priority* rather than *exclusivity*. Under this contract, the firm's invested capacity will be used first for the investing firm, and any leftover capacity invested but not used by the investing firm can be used to fulfill other orders.

The sequence of events is the same as the exclusive contract in Section 4.2. In the first stage, firms decide whether to invest or not, and if so, by how much. In the second stage, firms engage in the quantity competition. However, the firm with less invested capacity (say firm 2) can use any leftover capacity of firm 1, i.e., firm 2 can now order up to $k_s - q_1$ instead of $(\frac{k_0}{2} + k_2)\xi$. Then, the decision problems that the two firms face are as follows:

$$\pi_1^f(k_1, k_2, \xi) = \max_{q_1 \le (\frac{k_0}{2}+k_1)\xi} q_1 P(q_1, q_2) - cq_1; \pi_2^f(k_1, k_2, \xi) = \max_{q_2 \le k_s - q_1} q_2 P(q_1, q_2) - cq_2.$$

We present the equilibrium order quantity and resultant profit in Lemma III.3.

Lemma III.3 (Firms' equilibrium order quantity and ex post profit)**.**
Let the capacity investment sizes be (k_1, k_2)*, and the resulting subgame yields the following:*

realized yield ξ	*order quantity* (q_1^*, q_2^*)	*ex post profit* $\left(\pi_1^f,\ \pi_2^f\right)$
$0 \le \xi \le \frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}$	$\left((\frac{k_0}{2}+k_1)\xi, (\frac{k_0}{2}+k_2)\xi\right)$	$\left(m_2(\frac{k_0}{2}+k_1)\xi, m_2(\frac{k_0}{2}+k_2)\xi\right)$
$\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)} < \xi \le \frac{2(a-c)}{3b(k_0+k_1+k_2)}$	$\left(\frac{a-c}{b} - k_s, 2k_s - \frac{a-c}{b}\right)$	$\left(m_2(\frac{a-c}{b} - k_s), m_2(2k_s - \frac{a-c}{b})\right)$
$\frac{2(a-c)}{3b(k_0+k_1+k_2)} \le \xi \le 1$	$\left(\frac{a-c}{3b},\ \frac{a-c}{3b}\right)$	$\left(\frac{m_0(a-c)}{3b},\ \frac{m_0(a-c)}{3b}\right)$

We observe that under any realization of ξ, there are only two outcomes: either both firms are constrained by the capacity, or neither firm is constrained. In other words, a situation where only one firm is constrained does not arise. This is not surprising when ξ is very low (so the capacity is really tight) and ξ is very high (so the capacity is sufficiently large,) but even when the realized capacity is moderate $(\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)} < \xi \le \frac{2(a-c)}{3b(k_0+k_1+k_2)})$, we also observe that all the invested capacity in the supplier is used up. Intuitively, when the capacity type is first-priority and the realized capacity is at a moderate level, the firm with less capacity has incentive to order any leftover from the firm with more capacity because the benefit from satisfying more demand beyond its own invested capacity dominates the negative impact of the lower market price in a Cournot market. On the other hand, because firms compete in the same market, if the investing firm with more capacity has a unit of leftover capacity that the other firm does not want to tap into, the investing firm will also find it not profitable to satisfy the market with this additional unit of capacity. Thus, when the capacity type is first-priority, either both firms are constrained by the capacity, or neither firm is.

Following the same machinery in the exclusive capacity case, we can derive the firm's expected profit based on Lemma III.3 (the details are relegated to the appendix.) With the expected profits, we show the firm's equilibrium capacity investment in the following proposition.

Proposition III.4 (Firms' equilibrium capacity investment: First-priority capacity)**.** *There exist two equilibrium switching curves, $\underline{w}_0^f(w)$ and $\overline{w}_0^f(w)$, such that $\underline{w}_0^f(w) \leq \overline{w}_0^f(w)$ and*

i) When the fixed cost w_0 is low ($w_0 \leq \underline{w}_0^f(w)$), both firms invest in the supplier, where the equilibrium capacity, k^f, is

$$k^f \triangleq \left\{ k : \int_0^{\frac{a-c}{3b(\frac{k_0}{2}+k)}} \left[a - c - 3b\left(\frac{k_0}{2} + k\right)\xi \right] \xi f(\xi) d\xi - w = 0 \right\}, \quad (3.5)$$

which is decreasing in w. Furthermore, there exists a function $w_0^f(w)$ such that the equilibrium outcome is a prisoner's dilemma in the region between $w_0^f(w)$ and $\underline{w}_0^f(w)$, i.e., $V_i^f(k^f, k^f) \leq V_i^f(0,0)$ for $i = 1, 2$.

ii) When the fixed cost w_0 is intermediate ($\underline{w}_0^f(w) < w_0 \leq \overline{w}_0^f(w)$), the spillover effect occurs and only one firm (labeled as firm 1) invests in the

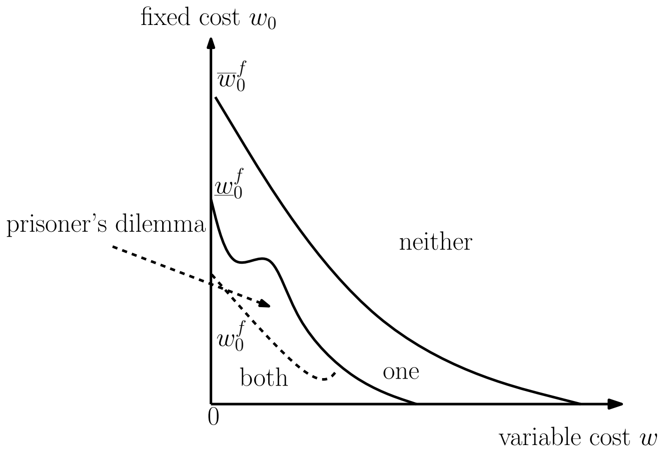

Figure 3.2: Equilibrium investment outcomes with first-priority capacity. "neither": neither firm invests. "one": one firm invests. "both": both firms invest.

supplier, where the equilibrium capacity, k_1^f, is such that

$$k_1^f \triangleq \left\{ k_1 : \begin{array}{l} \int_0^{\frac{a-c}{b\left(\frac{3k_0}{2}+2k_1\right)}} \left[a-c-b\left(\frac{3k_0}{2}+2k_1\right)\xi\right]\xi f(\xi)d\xi \\ +\int_{\frac{a-c}{b\left(\frac{3k_0}{2}+2k_1\right)}}^{\frac{2(a-c)}{3b(k_0+k_1)}} -2\left[a-c-b\left(k_0+k_1\right)\xi\right]\xi f(\xi)d\xi - w = 0 \end{array} \right\}, \tag{3.6}$$

and k_1^f decreases in w.

iii) When the fixed cost w_0 is high ($w_0 > \overline{w}_0^f(w)$), neither firm invests in the supplier.

Furthermore, $\overline{w}_0^f(w)$ decreases in w.

The general structure of the equilibrium is similar to the one in Proposition III.2: as the fixed cost w_0 increases, the equilibrium shifts from a regime where both firms invest, to a regime where only one firm invests, and finally to a regime where neither firm invests for a given variable cost. We also observe that the equilibrium switching curve $\overline{w}_0^f(w)$ decreases in w. However, comparing to the exclusive capacity case, we observe that the spillover effect occurs if one firm invests in the supplier with the first-priority capacity. That is, one firm is able to tap into the leftover capacity invested by the other firm.

To understand why the spillover effect does not occur under the other two regions (both firms investing or neither firm investing), we first observe that firms are able to access the same capacity level in both cases, as the firms commit to the same level of capacity investment in equilibrium and the random capacities available to the firms are perfectly correlated. In addition, both firms compete in the Cournot market. These facts imply that both firms will either exhaust all the available capacity or have some capacity leftover. Therefore, the spillover effect is observed only when *one* firm invests in the supplier's *first-priority* capacity.

Furthermore, the firm's capacity investment decision is affected by the spillover effect. This is reflected by the second term in the optimal condition (3.6). In this case, the realized capacity is moderately high ($\xi \in \left[\frac{a-c}{b(\frac{3k_0}{2}+2k_1)}, \frac{2(a-c)}{3b(k_0+k_1)}\right]$), and the benefit from the investment that allows the firm to access more capacity, is dominated by the downside which is the decreasing market price as the non-investing firm accesses the leftover and intensifies the quantity competition. As the benefit from accessing more capacity is dominated by the decreasing market price, the investing firm will have less incentive to invest in capacity.

Finally, we notice the equilibrium switching curve $\underline{w}_0^f(w)$ is not necessarily monotone in w due to the spillover effect. The only investing firm's profit decreases in the variable cost w and therefore the equilibrium switching curve $\overline{w}_0^e(w)$ that defines the one firm investing regime decreases in w. However, the non-investing firm's profit does not necessarily increase with respect to the variable cost (we observe that the profit of the non-investing firm increases in w in the exclusive capacity case.) While a higher w leads to a lower invested capacity and raises the price in the Cournot market, the non-investing firm may access the investing firm's leftover capacity. Thus when the investing firm lowers the capacity level, it could decrease the non-investing firm's capacity and resultant profit. Therefore, the non-investing firm's profit may increase or decrease in the variable capacity cost w, depending on which effect is stronger.

Thus we observe the equilibrium switching curve $\underline{w}_0^e(w)$ is not necessarily monotone in the variable cost w as shown in Figure 3.2.

3.5 Spillover effect: comparing exclusive and first-priority capacity

In Sections 3.3 and 3.4, we identified three equilibrium regimes – neither firm investing, one firm investing, and both firms investing, as well as how the equilibrium and its capacity level change in the capacity cost. We learned that the main difference between the two capacity types is that the first-priority capacity may result in the spillover effect, in which the non-investing firm benefits from access the investing firm's leftover capacity. In this section, we are interested in how the spillover effect affects the equilibrium outcomes, and as a result, the buying firms' and supplier's preference about the capacity type.

3.5.1 Impact on the equilibrium outcomes

At first glance, it seems that the spillover effect with the first-priority capacity will always intensify the competition because firms have greater flexibility to access the supplier's capacity. However, our analysis shows that due to this greater flexibility, the spillover effect indeed discourages some firms to invest in their supplier. Therefore, it could decrease both the number of investing firms and the capacity investment levels. As a result of the firms investing less aggressively, the spillover effect mitigates the risk of both firms being trapped in a prisoner's dilemma. Our next proposition formally summarizes these impacts.

Proposition III.5 (Impact of spillover effect on the equilibrium outcomes)**.**
i) The number of investing firms is higher with exclusive capacity than with first-priority capacity, i.e., $\underline{w}_0^f(w) \leq \underline{w}_0^e(w)$ *and* $\overline{w}_0^f(w) \leq \overline{w}_0^e(w)$.

ii) The total capacity level in the supplier is higher with exclusive capacity than with first-priority capacity.
iii) The prisoner's dilemma region is larger with exclusive capacity than with first-priority capacity, i.e., $[w_0^f(w), \underline{w}_0^f(w)] \subseteq [w_0^e(w), \underline{w}_0^e(w)]$ *for all* w.

Note that with the first-priority capacity, at first, it looks as if the end market competition is intensified as buying firms are able to access each other's leftover. If the supplier's total capacity level remains the same under both contracts, more units of the products will be available in the end market and the market clearing price will be lower with the first-priority capacity. In other words, competition between the two firms is intensified with the first-priority capacity. Thus, the investing firm has incentive to reduce its investment to reduce the competition intensity. On the other hand, because firms can access the other firm's leftover capacity and do not need as much capacity as before, they are also less motivated to invest in capacity. Therefore, firms invest less in the supplier with the first-priority capacity and the spillover effect indeed curbs competition between firms as indicated by Proposition III.5 (i) and (ii).

The reduced investment is manifested in two different ways. First, in Proposition III.5 (i), we find that the number of investing firms is lower with the first-priority capacity than with the exclusive capacity. Thus, in a region where both firms would invest with the exclusive capacity, it is possible that only one firm will invest with the first-priority capacity. Similarly, it is possible that neither firm will invest with the first-priority capacity in a region where one firm invests with the exclusive capacity. Second, in Proposition III.5 (ii), we observe that when there is only one firm investing, the invested capacity level is lower with the first-priority capacity, and when there are more firms investing with the exclusive capacity, the invested capacity level is also lower with the first-priority capacity, indicating that the spillover effect also decreases the total capacity.

We also observe that the prisoner's dilemma region is smaller with the first-priority capacity in Proposition III.5 (iii). From the buying firms' perspective, when only one firm invests, the non-investing firm (firm 2) is able to access the leftover capacity from the investing firm (firm 1). Therefore, firm 2 is less incentivized to invest aggressively to gain additional capacity. As a result, it follows that with the first-priority capacity, the prisoner's dilemma region is smaller, where both firms invest and compete intensively and earn a lower profit than what they earn when neither firm invests in the supplier.

3.5.2 Capacity type preference of buying firms and the supplier

We next analyze the implication of the spillover effect on the firms' and supplier's profit and their preference of the capacity types. In particular, we further investigate how the buying firms and supplier would prefer the capacity type.

Buying firms' preference. If both or neither firm invests in the supplier, firms are indifferent between the two capacity types. The difference in preference arises when only one firm invests in the supplier with at least one capacity type. The preference of buying firms is affected by two main effects: the *leading* effect associated with being the only investor, and the *spillover* effect that occurs with the first-priority capacity. When the investing firm (firm 1) becomes the only investor, the advantage of accessing more capacity allows the firm to extract more profit from the market than the non-investing firm. On the other hand, with the *spillover* effect the non-investing firm is able to access the leftover of the other firm's invested capacity. Consequently, the buying firm's preference depends on which of the two effects is stronger. We first summarize the buying firms' capacity type preference in the following proposition.

Proposition III.6 (Buying firms' preference about capacity type)**.**
i) When $0 \leq w_0 < \underline{w}_0^f(w)$, the firms are indifferent between the exclusive and first-priority capacity, i.e., $V_i^e(k^e, k^e) = V_i^f(k^f, k^f)$, as both firms invest the same amount with both types of capacity.

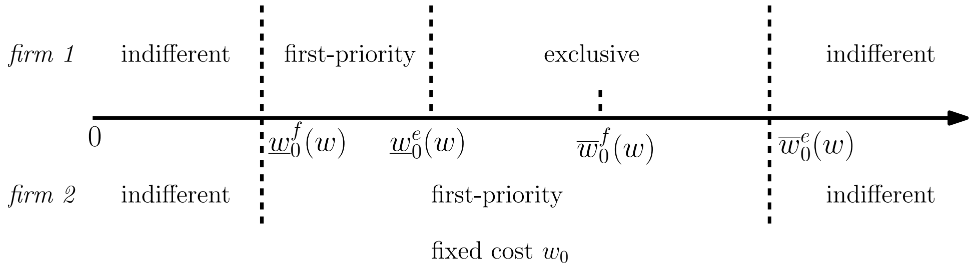

Figure 3.3: The buying firms' preference about the capacity types

ii) When $\underline{w}_0^f(w) \leq w_0 < \underline{w}_0^e(w)$, both firms prefer that one firm investing with first-priority capacity over both firms investing with exclusive capacity, i.e., $V_i^e(k^e, k^e) \leq V_i^f(k_1^f, 0)$.

iii) When $\underline{w}_0^e(w) \leq w_0 < \overline{w}_0^f(w)$ and $\overline{w}_0^f(w) \leq w_0 < \overline{w}_0^e(w)$, firm 1 prefers the exclusive capacity to prevent the spillover effect while firm 2 prefers the first-priority capacity, i.e., $V_1^e(k_1^e, 0) \geq V_1^f(k_1^f, 0)$, $V_1^e(k_1^e, 0) \geq V_1^f(0, 0)$, $V_2^e(k_1^e, 0) \leq V_2^f(k_1^f, 0)$ and $V_2^e(k_1^e, 0) \leq V_2^f(0, 0)$.

iv) When $w_0 \geq \overline{w}_0^e(w)$, the firms are indifferent between the exclusive and first-priority capacity, i.e., $V_i^e(0, 0) = V_i^f(0, 0)$, as neither firm invests.

The results are presented in Figure 3.3. One may expect that the investing firm should always prefer the exclusive capacity to mitigate the negative impact of the spillover effect and disallow the other firm to access its leftover capacity. However, our analysis indicates that this is not always the case. While the investing firm tends to prefer the exclusive capacity to prevent the spillover as shown in Proposition III.6 (iii), it prefers the first-priority capacity in the range $\underline{w}_0^f(w) \leq w_0 < \underline{w}_0^e(w)$ when the exclusive capacity will trigger the other firm to invest and the investing firm will lose the benefit from the leading effect, as shown in Proposition III.6 (ii). On the other hand, the non-investing firm always prefers the first-priority in the hope of being able to free-ride on the investing firm's leftover capacity. We next discuss the supplier's preference about the capacity types.

Supplier's preference. The supplier's expected profit with the exclusive and first-priority capacity, V_s^e and V_s^f, when firms invest in (k_1, k_2) are as follows. Recall that the supplier's realized capacity is denoted by $k_s = (k_0 + k_1 + k_2)\xi$, and c is the unit profit from production.

$$V_s^e(k_1, k_2) = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}} ck_s f(\xi)d\xi + \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}}^{\frac{a-c}{3b(\frac{k_0}{2}+k_2)}} c\left[\frac{a-c}{2b} + \left(\frac{k_0}{2} + k_2\right)\frac{\xi}{2}\right] f(\xi)d\xi + \int_{\frac{a-c}{3b(\frac{k_0}{2}+k_2)}}^{1} \frac{2c(a-c)}{3b} f(\xi)d\xi \tag{3.7}$$

$$V_s^f(k_1, k_2) = \int_0^{\frac{2(a-c)}{3b(k_0+k_1+k_2)}} ck_s f(\xi)d\xi + \int_{\frac{2(a-c)}{3b(k_0+k_1+k_2)}}^{1} \frac{2c(a-c)}{3b} f(\xi)d\xi \tag{3.8}$$

Comparing the supplier's profits under the two contacts leads to a few interesting observations. From the supplier's perspective, the spillover is a two-edged sword. While the spillover effect improves the capacity utilization of the supplier, it also reduces the supplier's total capacity as shown in Proposition III.5. Therefore, it is not obvious how the supplier's preference about capacity types changes as the type changes from exclusive to first-priority.

However, we show that if the fixed investment cost is low, both firms invest in the supplier and the supplier is indifferent between the two capacity types since both types induce the same amount of investment in the supplier. If the fixed cost is high, the supplier is also indifferent between the capacity types since neither firm is tempted to invest in the supplier anyway. In between, if more firms invest with exclusive capacity, the supplier benefits from the over-investment of buying firms more than the flexibility with first-priority capacity. These results are shown in the next proposition.

Proposition III.7 (Supplier's preference about capacity type)**.**

i) When $0 \leq w_0 < \underline{w}_0^f(w)$ *or* $w_0 \geq \overline{w}_0^e(w)$*, the supplier is indifferent between the exclusive and first-priority capacity, i.e.,* $V_s^e(k^e, k^e) = V_s^f(k^f, k^f)$ *and* $V_s^e(0,0) = V_s^f(0,0)$*.*

ii) When $\underline{w}_0^f(w) \leq w_0 < \underline{w}_0^e(w)$ *or* $\overline{w}_0^f(w) \leq w_0 < \overline{w}_0^e(w)$*, the supplier prefers the exclusive capacity, i.e.,* $V_s^e(k^e, k^e) \geq V_s^f(k_1^f, 0)$ *and* $V_s^e(k_1^e, 0) \geq V_s^f(0,0)$*.*

Proposition III.7 (i) shows that if both firms invest in the supplier, i.e., $0 \leq w_0 < \underline{w}_0^f(w)$, the supplier is indifferent between the two capacity types. To understand this, we notice that both firms compete in the Cournot market and both firms invest in the same capacity level. Therefore, both firms will either exhaust all the available capacity or have some capacity leftover for all realizations, and the first-priority capacity is *de facto* exclusive. Similarly, when neither firm invests in the supplier, i.e., $w_0 \geq \overline{w}_0^e(w)$, both firms will be able to access the same capacity level, and therefore the first-priority capacity is also *de facto* exclusive. In these two cases, the supplier is indifferent between the two capacity types.

Proposition III.7 (ii) implies that in regions $\underline{w}_0^f(w) \leq w_0 < \underline{w}_0^e(w)$ and $\overline{w}_0^f(w) \leq w_0 < \overline{w}_0^e(w)$, the supplier prefers the exclusive capacity to the first-priority capacity. In these two regions, the exclusive capacity results in strictly more firms investing. In the region where $\underline{w}_0^f(w) \leq w_0 < \underline{w}_0^e(w)$, both firms invest under the exclusive contract and one firm invests under the first-priority contract. Under the exclusive contract, both firms invest in the same capacity level in the supplier, and therefore either both firms will use up the invested capacity or both of them will have some leftover. That is, the supplier's capacity is utilized efficiently despite the exclusive claims. In addition, the total capacity invested by both firms with the exclusive capacity is higher than the total capacity invested by the one investing firm with the first-priority capacity. Therefore, the supplier is able to earn a higher profit with exclusive capacity and prefers the exclusive capacity to the first-priority capacity. In the region where $\overline{w}_0^f(w) \leq w_0 < \overline{w}_0^e(w)$, neither firm invests under the first-priority

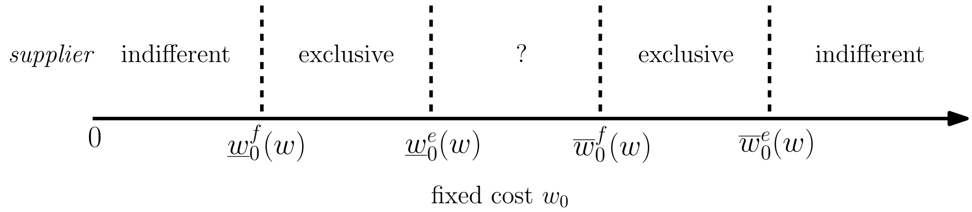

Figure 3.4: The supplier's preference about the capacity types

contract and one firm invests under the exclusive contract, so the supplier is able to extract more profit from leveraging the additional capacity invested by the investing firm and it also prefers the exclusive capacity. These results are also shown in Figure 3.4.

However, we note that the supplier's preference is not trivial when the fixed cost w_0 is between $\underline{w}_0^e(w)$ and $\overline{w}_0^f(w)$. In this case, only one firm invests under both contracts, and the supplier gains more capacity investment with the exclusive capacity but loses the flexibility in using it. Intuitively, if the realized yield is small, the supplier may benefit from the over-investment with exclusive capacity. If the yield is moderately high, the supplier may benefit from the flexibility in utilizing the first-priority capacity. If the yield is high, the supplier is indifferent between the two capacity types as there is enough capacity to produce with either type. Thus, depending on which of the two effects dominates, the supplier's preference may change. Its preference depends on the parameters such as the distribution of the yield and the variable cost to invest in the capacity. The supplier's *ex post* profit is illustrated in Figure 3.5 (a). To further explore the supplier's preference between the two capacity types in this case, we conduct a numerical study below.

In Figure 3.6, we present a numerical study showing how the difference in the supplier's profit between exclusive and first-priority contracts changes in the variable capacity cost w, and the distribution of the yield $\boldsymbol{\xi}$ when only one firm invests in the supplier. The first key observation is that the benefit from over-investment in the exclusive capacity tends to dominate the benefit from the flexibility in the first-priority

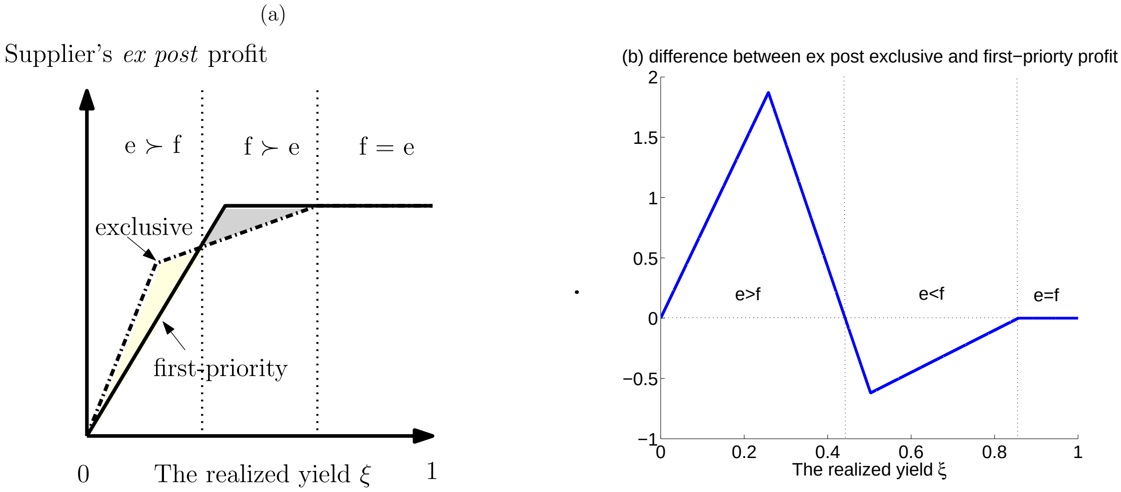

Figure 3.5: The supplier's *ex post* subgame perfect equilibrium profit when only one firm invests (a); a numerical example of the ex post profit difference (b). Parameters in (b): $a = 10$, $b = 1$, $c = 1$, $w = 0.1$, yield distribution $U[0, 1]$, and supplier's base capacity $k_0 = 7$.

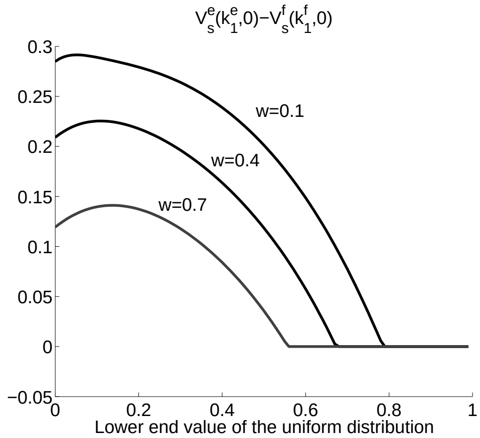

Figure 3.6: The supplier's expected profit difference $V_s^e(k_1^e, 0) - V_s^f(k_1^f, 0)$. Parameters: $a = 10$, $b = 1$, $c = 1$, supplier's base capacity $k_0 = 7$ and the yield distribution $U[low, 1]$ where the low value increases from 0 to 0.99.

capacity, i.e., $V_s^e(k_1^e, 0) - V_s^f(k_1^f, 0) \geq 0$. Let us take a closer look at one particular case of the *ex post* supplier's profit difference as shown in Figure 3.5 (b). It is clear that the region where the profit with exclusive capacity is greater than the profit with first-priority capacity, is greater than the other region, where the profit with exclusive capacity is smaller than the profit with first-priority capacity. This further confirms the conjecture that the over-investment benefit dominates the flexibility benefit. As the yield follows a uniform distribution between 0 and 1, we have that the ex ante expected profit with exclusive capacity is higher. Other cases are similar. While in this numerical example we illustrate that the benefit from the over-invested exclusive capacity often dominates the benefit from the flexibility in utilizing the first-priority capacity under the uniform yield distribution, the exact preference depends on the problem parameters and yield distribution in a general case.

To summarize the firms and supplier's preference about the capacity type, we note that the investing firm typically finds it beneficial to choose the exclusive capacity to prevent from the spillover, unless choosing the first-priority capacity allows it to access a larger portion of capacity and therefore it can enjoy the benefit from the leading effect. On the other hand, the non-investing firm typically prefers the first-priority capacity so that it can benefit from the spillover effect. The supplier, however, prefers the exclusive capacity when more buying firms invest under the exclusive capacity, driven by the benefit from firms' over-investment in the capacity. When only one firm invests under both contracts, the supplier's preference depends on which of the two effects is stronger, over-investment with the exclusive capacity or the smaller but flexible investment with the first-priority capacity.

3.6 Efficiency of the supply chain

So far we have derived the equilibrium outcomes with both types of capacity, analyzed the impact of spillover on the capacity investment decision, and characterized the firms' and supplier's preference about the capacity types. In this section, we focus on the efficiency of the supply chain. As a benchmark, we first analyze a first-best solution of the supply chain, where the downstream market is merged into a single firm in the end market and is integrated vertically with the supplier. We then compare the performance of the supply chain under both types of capacity with the first best solution and identify the inefficiency in the supply chain.

3.6.1 First-best benchmark

In the first-best benchmark, firms are not competing against each other. Instead, they serve the end market as a monopoly and are integrated vertically with the supplier. The sequence of events is still the same as in Section 4.2. In the first stage, the monopoly decides whether to invest to build extra capacity, and if so by how much. Because it is a vertically integrated monopoly, no decisions on exclusivity or first-priority need to be made. In the second stage, the monopoly produces based on the capacity realization and serves the end market with the market clearing price $p = a - bq_m$, where q_m is the total quantity the monopoly supplies to the market. Other parameters remain the same.

We solve this optimization problem using backward induction, and characterize the optimal solution in the following proposition.

Proposition III.8 (Optimal capacity investment)**.** *There exists an equilibrium switching curve $w_0^m(w)$ such that*

i) When the fixed cost is small ($w_0 \leq w_0^m(w)$), the monopoly prefers to invest in additional capacity, where the invested capacity k^m is such that

$$k^m \triangleq \left\{ k : \int_0^{\frac{a}{2b(k_0+k)}} \left[a - 2b\left(k_0 + k\right)\xi\right]\xi f(\xi)d\xi - w = 0 \right\}, \tag{3.9}$$

which decreases in w.

ii) When the fixed cost is high ($w_0 > w_0^m(w)$), the monopoly does not invest in additional capacity.

Furthermore, $w_0^m(w)$ decreases in w.

As before, when the fixed cost of the capacity investment is too high, the monopoly finds it not profitable to invest in additional capacity. Therefore, the investment may occur only if the fixed cost is low enough. As competition is absent in the first-best benchmark, if the monopoly decides to invest, it will simply choose the capacity level that maximizes its benefit, as reflected by equation (3.9). Based on the investment outcome, it is straightforward to derive the monopoly's expected profit. For the interest of space, we do not show the detailed formula.

3.6.2 Efficiency loss in the supply chain

Compared to the first-best benchmark, the supply chain with two firms sharing one supplier has two different features that may cause inefficiency: the *competition* (over-investment and spillover) among the buying firms and the *non-zero wholesale price*. The *competition* effect may lead to *over-investment* in the supplier, because the competing firms invest aggressively in order not to be shut out when serving the market. On the other hand, however, the *spillover* effect may partially mitigate the competition effect and result in a smaller total capacity. The *non-zero wholesale price* is another potential source of inefficiency for the supply chain, as when there are two tiers in the supply chain, a double marginalization effect distorts buying firms' incentive for capacity investment and order quantity. We next analyze the impact of the two types of inefficiency separately. We first examine the competition effect and then the effect of a non-zero wholesale price.

Impact of competition. In order to separate the impact of the two types of inefficiency, we first consider the supply chain in our base model, where two buying firms share a common supplier, but with the wholesale price set as $c = 0$. Doing so filters out any effect caused by the non-zero wholesale price, and comparing this case to the first-best benchmark allows us to isolate the impact of competition effect from the wholesale price. We highlight the impact of competition on the equilibrium capacity in the following proposition.

Proposition III.9 (Impact of competition on equilibrium capacity)**.** *When $c = 0$,*

i) With exclusive capacity, firms over-invest in the supplier's capacity. Specifically, the investing region is larger ($\overline{w}_0^e(w) \geq w_0^m(w)$), and the capacity investment level is higher ($2k^e \geq k^m$ and $k_1^e \geq k^m$).
ii) With first-priority capacity, firms may over-invest or under-invest in the supplier's capacity, but the spillover effect partially mitigates the over-investment in capacity. Thus, the total investment is lower under the first-priority contract than under the exclusive one.

When firms compete, they invest more aggressively in the supplier. In particular, when the capacity is exclusive, the region where there is at least one firm investing is larger than the region where the monopoly firm will invest in its capacity. Under the exclusive capacity, if both firms choose to invest, the competition effect dominates because either both firms have leftover or neither of them has. If only one firm invests in the supplier, the investing firm still over-invests in the supplier's capacity because it cannot access all the supplier's base capacity while it can in the first-best case. With the first-priority capacity, however, the investing firm has less incentive to invest in the supplier's capacity due to the spillover effect. Therefore, the spillover effect partially mitigates the over-investment with the exclusive capacity and may even result in under-investment in the supplier, as shown in Proposition III.9 (ii).

We next numerically investigate the impact of the competition on the equilibrium capacity investment and the supply chain efficiency. We use k_{sc}^i, $i = e, f$, to indicate the total capacity in equilibrium, and R_0^i to indicate the percentage of supply chain profit decrease relative to the first-best case. In Figure 3.7, we illustrate how the supply chain efficiency loss changes with the competition. We first observe in Figure 3.7 (a) that as the variable capacity cost w increases, the exclusive capacity is always higher than the first-best capacity, while the first-priority capacity can be higher than the first-best capacity (region I and III) or lower than the first-best capacity (region II). However, the first-priority capacity is always smaller than the exclusive capacity,

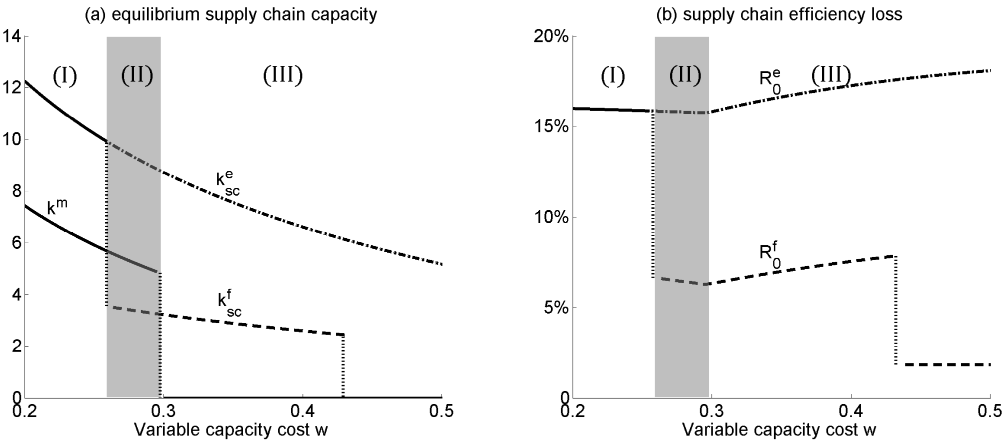

Figure 3.7: An example of efficiency loss of the supply chain due to competition. (a): k^m is the first-best supply chain capacity; k^e_{sc} (k^f_{sc}) is the total capacity when the capacity type is exclusive (first-priority). Region (I): $k^e_{sc} = k^f_{sc} > k^m$; Region (II): $k^e_{sc} > k^m > k^f_{sc}$; Region (III): $k^e_{sc} > k^f_{sc} \geq k^m$. (b): R^e_0 (R^f_0) is the supply chain efficiency loss relative to the first-best case when the capacity type is exclusive (first-priority). Parameters: $a = 10$, $b = 1$, $c = 0$, $w_0 = 1$, $k_0 = 7$, and the yield distribution $U[0, 1]$.

and this partial mitigation of the over-investment improves the supply chain efficiency as shown in Figure 3.7 (b). We observe that the supply chain efficiency loss is smaller under the first-priority capacity, and the improvement can be as large as 16.21%.

Impact of wholesale price. We next analyze the impact of the wholesale price by considering the supply chain with two competing firms sharing a supplier and a wholesale price c. We first show that when the capacity type is exclusive, a higher wholesale price disincentivizes buying firms to invest in the supplier in the proposition below.

Proposition III.10 (Impact of wholesale price on equilibrium exclusive capacity)**.** *When the capacity type is exclusive, both the number of investing firms and equilibrium capacity decrease as the wholesale price increases, i.e.,* $\overline{w}^e(w)$, $\underline{w}^e(w)$, k^e *and* k^e_1 *decrease in* c.

Note that a higher wholesale price is equivalent to a smaller market size. Therefore, with exclusive capacity, the investing firms will find it not necessary to invest much capacity when the market size is relatively small. This is why we observe both the number of investing firms and the sub-game perfect equilibrium capacity decreases in the wholesale price. One important implication of the result is that as the wholesale price increases, the over-investment observed in Proposition III.9 (i) may get mitigated. Therefore, the efficiency of the whole supply chain may be improved.

With first-priority capacity, however, while a smaller market size has the direct impact of disincentivizing the investment, it also diminishes the impact of the spillover effect and therefore may indirectly incentivize the investment. The is shown in equation (3.6), where the first term decreases in c reflecting the direct effect, and in the second term the integrand increases in c and the range of the integral decreases in c reflecting the indirect effect. Therefore, it is not clear analytically how the equilibrium capacity should change with respect to the wholesale price.

We next use numerical experiments to further explore the impact of the wholesale price on the equilibrium capacity and supply chain efficiency. In Figure 3.8 (a), we first observe that the equilibrium capacity decreases with the wholesale price. As a result, the over-investment in the supply chain can be mitigated and the efficiency of the supply chain may be improved. In Figure 3.8 (b), we observe that in the medium range of the wholesale cost, the supply chain efficiency loss is relatively small as the over-investment is significantly reduced from the level when $c = 0$. When $1.00 < c \leq 1.80$, we observe the first-priority capacity yields a lower efficiency loss compared to the exclusive capacity because only one firm invests under the first-priority contract and therefore the fixed cost is incurred only once. In contrast, both firms invest under the exclusive contract and therefore the fixed cost is incurred twice. When $1.80 < c \leq 2.80$, the exclusive capacity yields a lower efficiency loss than the first-priority capacity because firms are not incentivzed to invest under the first-priority contract as the wholesale cost is high, but they are still incentivized to invest under the exclusive contract. Therefore, the additional exclusive capacity

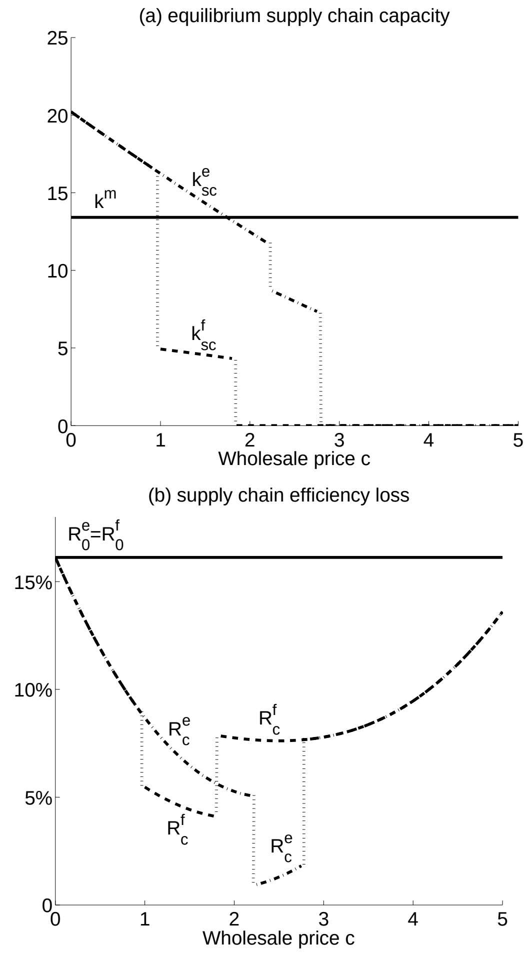

Figure 3.8: An example of efficiency loss of the supply chain due to the wholesale price. k^m is the first-best supply chain capacity; k^e_{sc} (k^f_{sc}) is the total capacity when the capacity type is exclusive (first-priority). R^e_0 (R^f_0) is the supply chain efficiency loss relative to the first-best case when the capacity type is exclusive (first-priority) and wholesale price is 0. R^e_c (R^f_c) is the supply chain efficiency loss relative to the first-best case when the capacity type is exclusive (first-priority) and wholesale price is c. Parameters: $a = 10$, $b = 1$, $k_0 = 7$, $w_0 = 1$, $w = 0.1$, and the yield distribution $U[0, 1]$.

investment, which is close to the first-best capacity as shown in Figure 3.8 (a), helps to improve the efficiency of the supply chain.

To summarize the discussion about the two types of inefficiency, interestingly, we find that both the spillover effect and the wholesale price can reduce the over-investment associated with the exclusive capacity. Therefore, when set appropriately, the wholesale price may echo the spillover effect of the first-priority capacity and improve the efficiency of the whole supply chain. However, in some cases, the combination of the wholesale price and the spillover effect may reduce the investment too much, and the exclusive contract performs better for the supply chain as a whole.

3.7 Conclusion

We investigate two capacity contract structures that firms may engage in when investing in expansion of a shared supplier's capacity. We characterize the equilibrium outcomes, identify conditions about when and to what extent the spillover effect and prisoner's dilemma occur, and analyze the impact of the spillover effect on the equilibrium outcomes, and firms' and supplier's capacity type preferences. We also investigated how the supply chain efficiency changes between exclusive and first-priority contracts.

Managerially, therefore, firms considering investing in suppliers who also supply their competitors must consider the consequences of their investment via the lens of a multi-player game, rather than myopically focusing on increased access to capacity. Placing restrictions on the supplier that are too tight may backfire in the form of competitors also jumping in with their own investments, which is reflected by the fact that more firms tend to invest and firms tend to over-invest with exclusive capacity in our model. We show that the spillover has both positive and negative effects on the investing firm. On the surface, allowing the spillover increases the end-market competition as more products are produced by both firms for a given capacity. On the other hand, allowing the non-investing firm to share the leftover capacity can actually disincentivizes a need for investment, and, consequently, both firms may avoid being trapped in a prisoner's dilemma. Depending on two effects, the leading effect and the spillover effect for buying firms, different capacity types may be preferable. While the non-investing firm always prefers the first-priority capacity, the investing firm does not always want to shut off the other firm from accessing its invested capacity. By allowing access to the leftover, both firms could be better off. We also show that the supplier's preference is driven by the tradeoff between the over-investment in the exclusive capacity versus the flexibility in utilizing the smaller first-priority capacity.

The results and insights of our work help us to understand a sequence of events that happened in the Foxconn-Sharp-Samsung case. Although there are many factors that affect Foxconn and Samsung's investment decisions, we highlight one particular factor in this case: Competing firms invest in a supplier to gain capacity. Despite its financial troubles, Sharp still maintains superior technology advantage in producing the LCD screens with the IGZO technology, and the factory in partnership with Hon Hai is the only one capable of producing the industry's largest sheets of glass panels (Osawa and Lee, 2013). Therefore, it is critical for firms to secure supply from Sharp to maintain competitive advantage in the future. Driven by this motivation, Foxconn invested to secure 50% capacity of Sharp's Sakai factory, while Samsung also invested to prevent competitors to gain too much control over Sharp, and secure a steady supply of LCD panels. This competition for access to capacity is precisely what this chapter considers, and this investment relationship is reflected in our model. As our analysis suggests, firms have to consider both direct and indirect consequences of their investment when they share a supplier with a competitor. Being too aggressive and claiming too much capacity may backfire as competitors may jump in with their own investment to prevent the firm from gaining a priority.

CHAPTER IV

To Share or Not to Share? Capacity Investments in a Shared Supplier

4.1 Introduction

Supply chains today are highly decentralized and consist of complex networks with many buying firms and suppliers. Often, the relationships between buyers and suppliers are not necessarily one-to-one and exclusive; multiple (competing) firms may share common suppliers, or a single firm may source the same component from multiple suppliers. The benefits of building and maintaining such supply chains include reduced cost (Li, 2013), increased reliability and resiliency (Wang et al., 2010), improved quality (Federgruen and Yang, 2009), and pooled resources (Plambeck and Taylor, 2005). In order to build and maintain relationships, firms share information and production technology with suppliers (Wang et al., 2014), offer mutually beneficial contracts that split the gain from coordination (Cachon and Lariviere, 2005), or invest production or financial resources in suppliers (Li, 2013). Rather surprisingly, we often observe that the firms engage in these activities even when their suppliers serve multiple firms, including the firms' competitors. Among these activities, we observe that many firms invest in or reserve the capacity in their shared suppliers. For example, Apple made sizable investments in its suppliers to expand their capacities

(Davis and Crothers, 1999; Ferrari, 2011). In 2011, Apple made a multi-billion dollar investment in LG Display for capacity expansion, although LG supplies its screen to other phone and tablet manufacturers which compete with Apple (LG also makes smartphones). According to IHS iSupply, instead of simply purchasing displays from its suppliers, Apple is providing financial support so that LG can build production capacity for new-generation LCD panels. A Fortune 100 company that the authors worked with, CISCO, invests in several suppliers to expand capacity in assembly and testing. A major cosmetics company, Johnson & Johnson's Neutrogena, purchased machines for a South Korean cosmetic supplier who also fulfilled orders for other cosmetic companies.

In practice, investing in a supplier does not always mean physical installation of new capacity at the supplier. Instead, the investment may also be an upfront monetary payment to reserve some of the supplier's capacity. In practice, these activities are also called supplier investment. Our model is general enough to account for both; in the rest of the chapter, the notion of "investing" and "reserving" can be used interchangeably, unless mentioned otherwise.

When a firm invests in a supplier who also serves another firm, one concern is that the supplier might use the invested (reserved) capacity to benefit other firms. This concern of leakage or spill-over is very pronounced when the supplier serves the investing firm's competitor. To avoid this, many investing firms impose constraints about how the capacity should be used, so that *the invested or reserved capacity be used exclusively or primarily* for the investing firm. For instance, in the cosmetics company that the authors worked with, machines invested by a buyer could be used only to satisfy her order. In another case of such exclusive capacity reservation, Foxconn invested $1.6 billion to buy a 46.5% stake in a single LCD factory owned by Sharp in Sakai, Japan, in early 2012, in order to receive a steady supply of LCD screens which Foxconn uses for Apple and other customers. (Dignan, 2012). Foxconn jointly runs operations with Sharp and earns a proportionate part of the revenues (Team, 2012). Under the joint agreement, Foxconn claims 50% of LCD outputs exclusively from the factory.

We note that these restrictions on the use of capacity are often enforced through physical devices like counters or fixtures, or audits by the investing firm or third parties. In some cases, buying firms send their employees to work at their supplier sites. As in the examples, buying firms often negotiate precise rules on the availability and use of contracted capacity.

Claiming exclusivity is not the only way to restrict the use of invested capacity. For example, from the authors' interview with the supply chain director of a global agriculture supply company, the firm invests in ingredient suppliers to reserve capacity, but it places no restrictions on what the supplier does with the fraction of the capacity that was not used. In this case, the firm only claims *first priority* in utilizing the supplier's capacity. In our literature review, we point out several papers that have considered one or both of these types of capacity investment and reservation.

When a supplier serves multiple firms, measuring the benefit of investment in the supplier becomes difficult. While investment allows a firm to reserve (secure) more capacity to serve uncertain demand, at the same time, the supplier and other buying firms can take advantage of the firm's investment and potentially intensify competition in the end market. In this chapter, we study how this trade-off affects capacity investment decisions: which capacity type firms (and the supplier) will choose so that they can strike a balance between the benefit from sharing the capacity and the adverse effects of investment.

To answer this question, we consider a supply chain (network) consisting of two buying firms sharing a supplier. Each firm, who faces uncertain demand, decides the terms of the capacity investment: capacity type—whether the invested capacity will be dedicated *exclusively* for the firm, or whether the investing firm will only demand *first priority* on the capacity leaving the supplier to use the unused capacity at its discretion— and the investment levels (costs of capacity installation/reservation vary depending on the terms and level of investment). Then, demands are realized and firms place orders to the supplier.

One of the critical factors that influence the benefit of investment in a shared supplier is how competitive the downstream market is. For instance, if the other firm is not directly competing with the investing firm in the end market, the impact of the supplier using leftover capacity to fulfill the other firm's order is not of concern to the investing firm. However, the use of leftover capacity can significantly affect the downstream competition if the other firm serves the same market as the investing firm. To study the impact of downstream market structure, we consider two different environments: (i) the two firms serve independent markets (firms' demands are independent), and (ii) the firms compete over quantity, *a la* Cournot.

Our contributions We model this situation as a multi-stage game and characterize its equilibrium. Our results highlight how firms should strategically choose the restrictions on the invested capacity as investment costs and/or the demands that firms face change. We show that three equilibria emerge: both firms investing in exclusive capacity, both firms investing in first-priority capacity, and one firm investing in exclusive and the other in first-priority capacity. We show that which of the three equilibria emerges depends on the capacity cost as well as the market environment that determines the demands of two firms. When demands of two buying firms are independent, the equilibrium changes from both choosing exclusive capacity to only one firm choosing exclusive capacity to no firm choosing exclusive capacity as the cost of reserving exclusive capacity increases. Rather surprisingly, we find that two *ex-ante* symmetric firms would end up in an asymmetric equilibrium where one firm allows the other to free-ride on its capacity. We note, however, that demand correlation makes the asymmetric investment equilibrium disappear in the Cournot market.

We also find that a prisoner's dilemma in which both firms overinvest only arises when both firms choose exclusive capacity. On the other hand, a free-rider equilibrium in which one firm invests in extra capacity and the other firm shares it even without any compensation occurs only when one firm chooses exclusive and the other firm chooses first-priority. Surprisingly these two outcomes only occur in the independent market. In the Cournot market, demand correlation makes these outcomes disappear: firms are never trapped in a prisoner's dilemma or a free-rider equilibrium. These results hold even if firms are asymmetric in terms of their wholesale price.

Although our base model considers the supplier to be a passive participant in this game, we also consider the model in which the strategic supplier is able to choose the capacity type that buying firms should invest. Once again we find that the outcome depends on the market environment, resulting in the different outcomes for the independent and Cournot market. For instance, when the capacity costs are the same, the supplier always prefers the exclusive capacity over the first-priority capacity in the independent market, but the two types are indifferent from the perspective of the supplier in the Cournot market. When the capacity costs are different, we show that the capacity type that the supplier prefers are different from the capacity type that the buyers prefer.

Our results and insights provide a plausible explanation for supplier investment in practice. It provides a rationale that why demanding to use the capacity exclusively does not necessarily serve the investing firm's best interest. When firms compete in a homogenous market, we show that firms may choose the first-priority capacity instead of more expensive exclusive capacity without worrying about spillover. On the other hand, in the independent market where there is no threat of competition, choosing exclusive capacity may become optimal even when doing so is more expensive than choosing first-priority capacity. Finally, we discuss the potential of using transfer payment to coordinate the network.

4.1.1 Literature review

The two main features of our model are sourcing from a shared supplier, and capacity investment. We review related literature from both strands of work below. We point out that ours is one of the first that combine sourcing from a shared supplier and *finite* capacity investment. Nevertheless, our contributions are best understood in light of the findings of the papers on sharing suppliers and capacity investment.

Our work is closely related to the outsourcing literature. Earlier works focus on analyzing the relationship between one buyer and one supplier. These include Iyer et al. (2005), Zhu et al. (2007) and Babich (2010). In recent years, the focus has moved toward more complicated relationships, which include dual sourcing (Li, 2013; Wang et al., 2010) and a back-up supplier (Yang et al., 2009).

There are several papers that considered (potentially competing) firms outsourcing to a common supplier. Cachon and Harker (2002) show that economy of scale makes outsourcing attractive even if the supplier does not have any direct cost advantage. Arya et al. (2008) find that for a retailer, outsourcing to a higher cost but uncapacitated supplier can be a strategic tool in raising its rival's costs and hence gain competitive advantage. Feng and Lu (2012) develop a multiunit bilateral bargaining framework and show that under both one-to-one and one-to-many channels, low cost outsourcing may hurt the manufacturers by changing manufacturers and supplier's bargaining position. Wadecki et al. (2011) analyze the setting where firms can invest to improve suppliers' reliability, and study the impact of downstream competition on firms' subsidies to (infinite capacity) suppliers in four supply chain structures differing in whether the firms share a supplier and whether the firms compete with each other. Wang et al. (2014) consider the effect of knowledge spillover (in contrast with capacity spillover) when one (or both) firms invest in a shared supplier, and derive conditions under which one or both firms invest. All these papers explore different incentives to outsource, identify possible adverse outcomes as a result of outsourcing, and find ways to coordinate misaligned incentives using either a non-cooperative or cooperative game framework. Most papers that examine multiple firms outsourcing to a common supplier do not consider capacity constraint. Discussions related to capacity investment in an outsourcing setting are relatively sparse.

Capacity management (within the firm) has been studied in industrial organization and operations management literature for decades; Dixit (1980) is an example of an early and influential paper in this area. Van Mieghem (2003) provides a review of earlier works. Among more recent papers in which firms build their own capacity and compete in an end market, Goyal and Netessine (2007) consider two competing firms' choice of flexible or dedicated technology and capacity investment, and show that flexible and dedicated technology can coexist in equilibrium.

There are a few recent papers that address capacity issues in an outsourcing setting. Plambeck and Taylor (2005) use a cooperative game framework to analyze the impact of contract manufacturing on innovation and capacity. They show that if firms outsource to a supplier, the supplier can increase utilization by pooling, which decreases firms' willingness to invest in innovation. If firms share capacity as a joint venture, this may increase or decrease innovation, but profitability always increases. Plambeck and Taylor (2007) analyze two firms outsourcing to a manufacturer through quantity flexible contracts. The firms invest in innovation, which determines the demand realization, while the supplier invests in capacity. In this setting, they show that renegotiation leads to better coordinating quantity flexible contracts. Ülkü et al. (2005) compare the timing of market entry between two supply chain structures: firms produce in-house, or outsource to a supplier. They show that firms and the supplier can coordinate the capacity and timing decision in equilibrium. Ülkü et al. (2007) analyze the case where a supplier and firms differ in demand forecast accuracy, and either one of them (or both) are responsible for capacity investment and bear the risk that the demand might be lower than the capacity built. If only firms invest, then the invested capacity is exclusively used by the investor. If only the supplier invests, then the invested capacity can be used to satisfy any firm's order. They explore the effectiveness of a premium-based contract which induces both parties to invest. Li et al. (2011) compare three capacity reservation options: no-transfer, supplier-transfer, and buyer-transfer, depending on whether the reserved capacity can be accessed by other firms and whether it is costly to access the other firm's reserved capacity. They identify which option is optimal from the supplier's perspective. Our work is different from these papers in two major ways (among many other minor points). None of these papers considers a situation where the buying firm can restrict how the reserved capacity should be used by the supplier. In addition, firms operate in independent markets in these papers (except for Li et al. (2011), where the firm's demands might be correlated but there are no price-dependency and competition), while we explore the impact of different levels of competition.

4.2 Model

The key model features are derived to frame our motivating examples, although some stylization is necessary. We consider a supply chain with a single supplier and two buying firms: we use $i = 1, 2$ to index the two downstream firms, and s to denote the supplier. When referring to firm i (she), we use j (he) to denote the other firm. Our game of capacity reservation and investment proceeds in three stages. In the first stage, firms select the type of restrictions, which we refer to as "capacity type", that specify how the supplier should use the invested (reserved) capacity. The capacity type determines how much capacity firms can use at their discretion, and resultantly, the output produced for the two firms. In the second stage, firms choose investment levels in the presence of demand uncertainty. In the third and last stage, the demand uncertainty is realized, firms fulfill their demands through the supplier (constrained by the capacity that firms can tap into) and accrue the revenues. We note that this three-stage model is equivalent to a two-stage game with contingent contract offers, and we discuss it in detail in *Remark 1* at the end of this section.

We consider the two most common types of capacity reservation for the first stage. Under the *exclusive* (denoted by e) restriction, the investing firm demands exclusive use of the reserved capacity and disallows the supplier from using the leftover capacity for any other orders. The second capacity type is *first priority* (denoted by f), where the investing firm claims the first right to use the invested capacity, but the supplier is free to use any leftover. Let w_e and w_f be the per-unit cost of exclusive and first-priority capacities respectively. That is, if firm i chooses to invest in K_i units of capacity under capacity type $\kappa_i \in \{e, f\}$, the cost to firm i is $w_{\kappa_i} K_i$. A portion of the capacity reservation cost, $\gamma w_{\kappa_i} K_i$ for some $\gamma \in [0, 1]$, accrues to the supplier as profit and the remaining portion, $(1-\gamma) w_{\kappa_i} K_i$, is spent in installing the capacity. The two boundary cases represent the scenarios where the entire capacity investment, $w_{\kappa_i} K_i$, is used to reserve the supplier's existing capacity as in the agriculture supply company case (i.e., $\gamma = 1$), and where the entire investment is used to build the capacity as in the Neutrogena case ($\gamma = 0$). The case, $0 < \gamma < 1$, is a mixture of the two.

In many cases, demanding exclusive use of capacity is costlier than demanding the first priority ($w_e > w_f$). This is particularly true if the supplier is willing to give a discount for investing in first-priority capacity because the supplier can potentially use it to meet other demand. However, there could be situations where $w_e \leq w_f$ (e.g., it may be costlier to build capacity that is more flexible; exclusive capacity may be cheaper because the buying firm provides proprietary technology and equipment). Thus, we analyze both cases and place no restrictions on the relationship between w_e and w_f.

As mentioned in the introduction, firms impose these restrictions to avoid direct or indirect spillover of their investment to other firms, and these restrictions can be enforced through the installation of physical devices or external auditing procedures.

The impact of such spillover depends on the nature of market competition as well as demand interaction between the firms. To study this, we consider two market environments with varying degree of competition: independent market and Cournot market. To describe the markets in consistent notation, let θ_i be a random variable representing the uncertain willingness to pay for firm i, and firm i's inverse demand be $p_i = \theta_i - b_i q_i - \eta_i b_i q_j$, where q_i and q_j are firm i and j's production quantities respectively, b_i is an exogenous parameter, and η_i measures the competition intensity between the two firms.

1. **Independent market**: In this environment, demands of two firms are independent. Examples include two firms serving geographically separated markets or producing different products from the same supplier. Therefore, each firm's demand is independent of the other firm's. That is, θ_1 and θ_2 are independent random variables, and there is no direct competition between the two firms, i.e., $\eta_i = 0$. For a given realization $\theta_i = a_i$, firm i's inverse-demand (price) p_i is given by $p_i = a_i - b_i q_i$.

2. **Cournot market**: In this case, the two firms produce homogeneous products and compete with quantities. That is, θ is a random variable that determines the inverse-demand for both firms (i.e, $\theta_1 = \theta_2 = \theta$), and firms directly compete via quantity, i.e., $\eta_i = 1$. Unlike the independent market, here demand signals are perfectly correlated ($\theta_1 = \theta_2$). Following the convention of Cournot competition, we also let $b_1 = b_2 = b$. For given production quantities q_1 and q_2, and realization $\theta = a$ (i.e, $a_1 = a_2 = a$), the market clearing price (inverse-demand) is $p_1 = p_2 = a - b(q_1 + q_2)$.

These two scenarios represent two extreme levels of competition and demand interaction, yet they are constructed to capture the key features of practice. In the independent market, the demand signal θ_i's are *independent*, and each firm's demand is independent of the other firm's. For example, both toothpaste and facial cleanser can be produced in the same tubing factory. In this case, it is reasonable to assume that demand for toothpaste is independent of demand for facial cleanser.

In contrast, in the Cournot market, the demand signals are *positively correlated*, but the two firms compete with quantities. To motivate this, consider the agriculture supply company and its ingredient supplier. For products such as insecticides, many of them share similar active ingredients and have similar functionality even if they are produced by different manufacturers. Therefore, the market for these products can be considered as (almost) perfect substitutes, which approximates the case of a Cournot market.

We next describe the general framework of how the equilibrium is determined, starting from the last stage of the game. Details for each market type follow in subsequent sections. For ease of notation, we suppress the dependence of p_i on (θ_i, θ_j) and (q_i, q_j) except for the case where the relationship must be shown explicitly. Let c_i represent the wholesale price per unit charged by the supplier for firm i.

In the last (third) stage, each firm observes the demand signal and decides on quantity. The maximum available capacity depends on the capacity types (e for exclusive or f for first-priority) and investment levels of both firms. To see this, first consider firm i's production decision when the other firm (firm j) reserves capacity exclusively. Thus, firm i is precluded from tapping into the capacity that firm j has reserved through capacity investment. If the two firms' investment levels and realized demand signals are (K_1, K_2) and $(\theta_1, \theta_2) = (a_1, a_2)$, respectively, firm i with capacity reservation type $\kappa_i \in \{e, f\}$ chooses the quantity that maximizes the following:

$$\pi_i^{\kappa_i e}(a_i, a_j, K_i, K_j) = \max_{q_i} (p_i - c_i)\, q_i, \quad \text{s.t.} \quad 0 \leq q_i \leq K_i \tag{4.1}$$

On the other hand, consider the case where firm j has reserved K_j units of capacity under first-priority reservation. Then, firm i can access firm j's leftover capacity. Thus, the optimal quantity of firm i is determined by solving the following decision problem. Throughout the analysis, we use x^+ to denote $\max\{x, 0\}$.

$$\pi_i^{\kappa_i f}(a_i, a_j, K_i, K_j) = \max_{q_i} (p_i - c_i) q_i, \quad \text{s.t.} \quad 0 \le q_i \le K_i + (K_j - q_j)^+ \tag{4.2}$$

We assume that the total supplier capacity that we consider is $K_1 + K_2$. This corresponds to the case where the supplier has no initial capacity. It also corresponds to the case where $K_1 + K_2$ are the total quantities of the supplier's existing capacity reserved by both firms.

We roll back the firms' optimal order quantities and solve for the firm's capacity investment decision in the second stage. For a given pair of reservation types, (κ_1, κ_2) where $\kappa_i \in \{e, f\}, i = 1, 2$, each firm sets the investment level K_i before the demand signals are realized. We denote firm i's expected profit given reservation types and investment levels by $\Pi_i^{\kappa_i \kappa_j}(K_i, K_j)$, that is

$$\Pi_i^{\kappa_i \kappa_j}(K_i, K_j) = E_{(\theta_i, \theta_j)}[\pi_i^{\kappa_i \kappa_j}(\theta_i, \theta_j, K_i, K_j)] - w_{\kappa_i} K_i. \tag{4.3}$$

A word about notation: we are following a convention where whenever we have arguments in subscripts, superscripts or brackets that define quantities for both firms, we first list the one for the focal firm, followed by the one for the other firm. So, in the equation above, we denote $\Pi_1^{\kappa_1 \kappa_2}(K_1, K_2)$ to represent firm 1's expected profit, and $\Pi_2^{\kappa_2 \kappa_1}(K_2, K_1)$ to represent firm 2's expected profit. Although somewhat unconventional, this allows us to distinguish clearly the impact of quantities (decision variables or demand parameters) that are specific to the focal firm from quantities that are specific to the other firm, and significantly simplifies the notation and analysis.

Then, the subgame-perfect capacity levels, $(K_1^{\kappa_1 \kappa_2}, K_2^{\kappa_2 \kappa_1})$, will satisfy the following equations simultaneously.

$$K_i^{\kappa_i\kappa_j} = \arg\max_{K_i\geq 0}\left\{\Pi_i^{\kappa_i\kappa_j}(K_i, K_j^{\kappa_j\kappa_i})\right\}, \text{ for } i,j\in\{1,2\} \text{ and } i\neq j \tag{4.4}$$

In the first stage, both firms choose the reservation type $\kappa_i \in \{e, f\}$ simultaneously. From equation (4.4), this game can be expressed as a 2×2 reduced form game with the following payoff matrix in which $\Pi_i^{\kappa_i\kappa_j}$ represents the sub-game perfect expected profit, $\Pi_i^{\kappa_i\kappa_j}(K_i^{\kappa_i\kappa_j}, K_j^{\kappa_i\kappa_j})$.

$(\kappa_1,\ \kappa_2)$	exclusive	first-priority
exclusive	$\Pi_1^{ee},\ \Pi_2^{ee}$	$\Pi_1^{ef},\ \Pi_2^{fe}$
first-priority	$\Pi_1^{fe},\ \Pi_2^{ef}$	$\Pi_1^{ff},\ \Pi_2^{ff}$

(4.5)

In what follows, we analyze the equilibrium and examine the resultant outcomes from both firm's perspective for the two market environments (independent and Cournot) in Sections 4.3 and 4.4 respectively. In Section 4.5, we also analyze the supplier's profit and compare the outcomes under which a strategic supplier can decide capacity types to the results in Sections 4.3 and 4.4.

Remark 1. We note that our model features a three-stage game with the capacity type choice and capacity investment decisions separated, i.e., the action space is $\{e, f\}$ in the first stage and $[0,\infty)$ in the second stage. However, this is equivalent to a two-stage game with contingent offers, where the action space for the first stage is $\{e, f\}\times[0,\infty)$. That is, in the first stage, both firms strategically offer a capacity investment plan, where the capacity investment level is contingent upon the other firm's capacity type. Then after the demand uncertainty is resolved, firms decide on quantity and satisfy the demand. This contingent offer is feasible because the action space of firm's capacity type choice $\{e, f\}$ has only two elements. Therefore, our model captures the case when firms simultaneously decide capacity type and capacity level in the first stage, followed by orders in the second stage.

Remark 2. Our model assumes that, for first-priority capacity, the supplier can access the leftover capacity without transfer payment. This assumption is consistent with our observations of the agriculture supply company and Li et al. (2011)'s observation about the computer hard-drive industry. If we explicitly include a transfer payment, either the supplier or the firm who wants to access the leftover capacity will pay the transfer payment. Since such payment can incentivize the use of the leftover capacity, one may conclude that, with transfer payment, firms are induced to choose the first-priority capacity more often. However, even without such arrangement, our results show that the firms may find it beneficial to share the capacity. As this chapter focuses primarily on how operational difference between the two types (i.e., whether the leftover can be used by other firms or not) results in different investment outcomes, we exclude the transfer payment from our model. But, we provide the impact of a positive transfer payment in Section 6.1.

4.3 Independent market

We begin with an analysis of the independent market scenario. Let $F_i(\cdot)$ be the distribution of firm i's demand signal. For a given realization, $\theta_i = a_i$, firm i's demand is endogenously determined with the inverse demand function, $p_i = a_i - b_i q_i$. We normalize $c_1 = c_2 = c \geq 0$; this is without loss of generality because the case where $c_1 \neq c_2$ is *equivalent* to the case where the unit cost is the same but the distribution of the firm's demand signal θ_i is shifted by the cost difference. As noted above, we solve this three-stage game via backward induction, starting with the third stage where firms decide order quantities after observing demand.

4.3.1 Third-stage quantity game

In the quantity game, the demand signal θ_i is realized, and firms choose production quantities to maximize their own profit. As stated in Section 4.2, the maximum available capacity depends on the capacity reservation type and investment level. If firm j reserves K_j units of exclusive capacity, from (4.1), firm i who has K_i units of capacity solves the following:

$$\pi_i^{\kappa_i e}(a_i, a_j, K_i, K_j) = \max_{q_i} (a_i - b_i q_i - c)\, q_i, \quad \text{s.t.} \quad 0 \le q_i \le K_i$$

Likewise, when the other firm reserves first-priority capacity, we have

$$\pi_i^{\kappa_i f}(a_i, a_j, K_i, K_j) = \max_{q_i} (a_i - b_i q_i - c)\, q_i, \quad \text{s.t.} \quad 0 \le q_i \le K_i + (K_j - q_j)^+$$

Since the profit function is concave in quantity, we obtain firms' optimal quantity or a best response function with respect to the other firm's order. Then we can obtain firms' equilibrium order quantities, stated in the following lemma. All proofs are relegated to the appendix.

Lemma IV.1 (Firms' equilibrium order quantities)**.** *If firm i has K_i units of type κ_i capacity and firm j has K_j units of type κ_j capacity, then firm i's equilibrium order quantity $q_i^{\kappa_i \kappa_j}$ for demand realization (a_i, a_j) is as follows, where $K_T = K_1 + K_2$.*

(κ_i, κ_j)	*Firm i's order quantity $q_i^{\kappa_i \kappa_j}$*
(e, e) *or* (f, e)	$\min\{\frac{(a_i-c)^+}{2b_i},\ K_i\}$
(e, f) *or* (f, f)	$\begin{cases} K_i & \text{if } K_i \le \frac{(a_i-c)^+}{2b_i} \text{ and } K_j \le \frac{(a_j-c)^+}{2b_j} \\ K_T - \frac{(a_j-c)^+}{2b_j} & \text{if } \frac{(a_j-c)^+}{2b_j} \le K_j \text{ and } \frac{(a_i-c)^+}{2b_i} + \frac{(a_j-c)^+}{2b_j} \ge K_T \\ \frac{(a_i-c)^+}{2b_i} & \text{otherwise} \end{cases}$

Lemma IV.1 states that when the other firm's capacity has been reserved exclusively and cannot be accessed, the optimal quantity is the minimum of the unconstrained profit-maximizing quantity and the firm's initial capacity K_i. If the other firm has reserved first-priority capacity, although the two firms' demands are independent, the optimal production quantity may depend on the other firm's demand. For instance, if firm i has high demand ($K_i \leq \frac{(a_i-c)^+}{2b_i}$) and firm j has low demand ($K_j > \frac{(a_j-c)^+}{2b_j}$), then firm i may use all or a portion of firm j's leftover capacity to meet her demand. Consequently, the optimal quantity is the minimum of $K_T - \frac{(a_j-c)^+}{2b_j}$ and $\frac{(a_i-c)^+}{2b_i}$.

4.3.2 Second-stage capacity investment game

We now move on to derive firms' expected profits, and then optimal capacity investment levels given capacity reservation types. Following the equilibrium order quantities, we express firms' profits as follows. If firm j has reserved capacity K_j exclusively, firm i cannot access any of it. Thus, firm i's expected profit when firm i reserves K_i units is:

$$\Pi_i^{\kappa_i e}(K_i, K_j) = \int_{c}^{2b_iK_i+c} \frac{(a_i-c)^2}{4b_i}\mathbf{d}F_i(a_i) + \int_{2b_iK_i+c}^{\infty} (a_i - c - b_iK_i)K_i\mathbf{d}F_i(a_i) - w_{\kappa_i}K_i \quad (4.6)$$

The two integrals above represent the case when the demand is low so that the unconstrained optimal order quantity is below the firm's capacity, and the case when the demand is high so that it is optimal to exhaust the capacity, respectively. If firm j has the first-priority capacity, firm i is able to access firm j's leftover. For this case, we obtain the following expression: there are several terms because we now have to account for various combinations of whether or not each firm's capacity is binding.

$$
\begin{aligned}
\Pi_i^{\kappa_i f}(K_i,K_j) = & \int_0^c \int_{2b_iK_T+c}^{\infty} (a_i - c - b_iK_T)K_T \mathbf{d}F_i\mathbf{d}F_j \\
& + \int_{2b_jK_j+c}^{\infty} \int_{2b_iK_i+c}^{\infty} (a_i - c - b_iK_i)K_i \mathbf{d}F_i\mathbf{d}F_j \\
& + \int_{c}^{2b_jK_j+c} \int_{2b_iK_T+c-\frac{b_i(a_j-c)}{b_j}}^{\infty} [a_i - c - b_i(K_T - \tfrac{a_j-c}{2b_j})](K_T - \tfrac{a_j-c}{2b_j}) \mathbf{d}F_i\mathbf{d}F_j \\
& + \int_{2b_jK_j+c}^{\infty} \int_{c}^{2b_iK_i+c} \frac{(a_i-c)^2}{4b_i} \mathbf{d}F_i\mathbf{d}F_j + \int_{c}^{2b_jK_j+c} \int_{c}^{2b_iK_T+c-\frac{b_i(a_j-c)}{b_j}} \frac{(a_i-c)^2}{4b_i} \mathbf{d}F_i\mathbf{d}F_j \\
& + \int_0^c \int_c^{2b_iK_T+c} \frac{(a_i-c)^2}{4b_i} \mathbf{d}F_i\mathbf{d}F_j - w_{\kappa_i}K_i
\end{aligned}
\tag{4.7}
$$

For each pair of reservation choices, (κ_i, κ_j), we can find the equilibrium investment level that maximizes the expected profits above. We find the following:

Lemma IV.2 (Firms' equilibrium capacities)**.**
(i) If $w_e > w_f$, then $K_i^{fe} \geq \max\{K_i^{ee},\ K_i^{ff}\} \geq \min\{K_i^{ee},\ K_i^{ff}\} \geq K_i^{ef}$.
(ii) If $w_e \leq w_f$, then $K_i^{ee} \geq K_i^{fe} \geq K_i^{ff}$ and $K_i^{ee} \geq K_i^{ef}$. In addition, if $w_e = w_f$, $K_i^{ee} = K_i^{fe}$.

The intuition behind this lemma is as follows. If firm j has reserved capacity exclusively, firm i is no longer able to tap into the leftover (thus, $K_i^{\kappa_i e}$ is firm i's final capacity). Because the two markets are independent (thus, there is no market competition to worry about), it is optimal for the firm to build a larger capacity with the cheaper option. Hence, if $w_e > w_f$, we have $K_i^{fe} \geq K_i^{ee}$; otherwise, $K_i^{fe} \leq K_i^{ee}$.

When firm j reserves capacity with first priority, firm i needs to build less capacity as it can access the other firm's leftover. This explains the orderings $K_i^{ee} \geq K_i^{ef}$ and $K_i^{fe} \geq K_i^{ff}$. When choosing the optimal capacity level, firm i needs to strike a balance between building its own capacity and taking a risk to rely for some of production on the other firm's leftover. When the exclusive capacity is more expensive than the first-priority capacity ($w_e > w_f$), firm i will build smaller capacity under the exclusive restriction than what she would build under the first-priority restriction ($K_i^{ff} \geq K_i^{ef}$) for two reasons. First, the unit cost of exclusive capacity is higher. Second, if firm i shuts off pooling by choosing the exclusive capacity, firm j (who reserved the first-priority capacity) needs to build a higher capacity. As demand is uncertain, increasing the capacity level will increase the chance of leftover too. Consequently, the available capacity for firm i increases and the firm builds less capacity. However, the ordering of $K_i^{ff} \geq K_i^{ef}$ may not hold when $w_e \leq w_f$. To see why this is the case, first note that firm i can still access the leftover of the other firm's capacity in both cases and $K_j^{fe} \geq K_j^{ff}$. This incentivizes the firm to build smaller capacity under exclusive capacity reservation. On the other hand, in this case, the cost of reserving a unit of exclusive capacity is cheaper than that of first-priority capacity. Therefore firm i has incentive to build more capacity. Depending on which of the two effects is stronger, the firm may build higher or lower capacity under exclusive capacity reservation.

4.3.3 First-stage capacity type choice game

Utilizing the results of the second and third stages, we derive firms' expected profits for a given pair of reservation types, (κ_i, κ_j), $\kappa_i = e, f$. From equation (3), if $(K_i^{\kappa_i\kappa_j}, K_j^{\kappa_j\kappa_i})$ represent the subgame perfect capacity levels for the two firms when firms i and j choose reservation types, κ_i and κ_j, respectively, then the firm i's subgame-perfect expected profit, $\Pi_i^{\kappa_i\kappa_j}$, is

$$\Pi_i^{\kappa_i\kappa_j} = \Pi_i^{\kappa_i\kappa_j}(K_i^{\kappa_i\kappa_j}, K_j^{\kappa_j\kappa_i}) = E_{(\theta_i,\ \theta_j)}[\pi_i^{\kappa_i\kappa_j}(\theta_i, \theta_j, K_i^{\kappa_i\kappa_j}, K_j^{\kappa_j\kappa_i})] - w_{\kappa_i}K_i^{\kappa_i\kappa_j} \tag{4.8}$$

The following theorem characterizes which type of capacity each firm chooses in equilibrium, which is the solution of the game characterized by the payoff matrix (4.5).

Theorem IV.3 (Equilibrium capacity type choices)**.**
(i) Suppose $w_e \leq w_f$*. Then, it is a dominant strategy to reserve the capacity exclusively.*
(ii) Suppose $w_e > w_f$*. Then, for given* w_f*, there exists a threshold* $\bar{w}_e{}^I(w_f)$ *such that, in equilibrium, one firm reserves the capacity with first priority and the other firm reserves the capacity exclusively for* $w_e \leq \bar{w}_e{}^I(w_f)$*, and both firms reserve first-priority capacity for* $w_e > \bar{w}_e{}^I(w_f)$*.*

To see why reserving capacity exclusively is a dominant strategy when $w_e \leq w_f$, suppose that firm j chooses the exclusive capacity. Then, firm i's best response is to choose the exclusive capacity, because it is cheaper and there is nothing to be gained by offering flexibility to the other firm. When firm j chooses the first-priority capacity, firm i is still better off with the exclusive capacity because in addition to being a cheaper option, the fact that the other firm (with first-priority capacity) is not allowed to tap into firm i's exclusive capacity forces the other firm to set a higher capacity level. Combining both cases, it is a dominant strategy for both firms to reserve capacity exclusively. However, as Corollary IV.4 will show, although exclusive capacity is dominant, an equilibrium in which both firms choose exclusive capacity is not always efficient, resulting in a prisoner's dilemma.

Now consider the case where the cost of first-priority capacity is lower: $w_f < w_e$. When w_e is slightly greater than w_f, i.e., $w_f < w_e < \bar{w}_e{}^I(w_f)$, one firm chooses first-priority capacity and the other firm chooses exclusive capacity in equilibrium. Considering the demands are independent, it is surprising that one firm chooses the exclusive capacity, which is more expensive. In this case, if firm j chooses first-priority capacity, firm i should choose exclusive capacity, because, as in discussions following Lemma IV.2, doing so forces firm j to build a larger capacity that firm i can tap

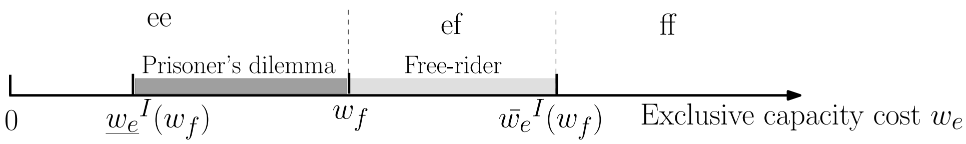

Figure 4.1: Equilibrium capacity type choices for independent market for given w_f

into, resulting in reduction of firm i's own investment. Likewise, if firm j chooses exclusive capacity, firm i should choose first-priority capacity only because the first-priority capacity is cheaper, and she cannot access firm j's leftover capacity anyway. Therefore, in equilibrium, one firm is willing to let the other firm free-ride on her unused capacity. However, when w_e is sufficiently high, $w_e > \bar{w_e}^I(w_f)$, the lower cost of the first-priority capacity dominates the benefit of choosing the exclusive capacity and forcing the other firm (with first priority) to build a larger capacity that firm i can tap into. Since the firm only gives up the unused portion of capacity, there is no profit loss from letting the other firm tap into the leftover when demands are independent. This equilibrium has the added benefit of allowing both firms to pool their demand uncertainties, resulting in higher profits and lower total capacity.

However, not all equilibria result in efficient outcomes. The equilibria may lead to adverse outcomes as shown in the following corollary.

Corollary IV.4 (Pitfalls in equilibrium capacity type choices)**.** *For given $w_f \geq 0$,*
(i) A prisoner's dilemma *equilibrium occurs, where both firms could have increased their profits by choosing the first-priority capacity, if $\underline{w_e}^I(w_f) \leq w_e \leq w_f$ for some $\underline{w_e}^I(w_f)$.*
(ii) A free-rider *equilibrium occurs, where the one firm invests in the first-priority capacity while the other firm invests in the exclusive capacity if $w_f < w_e \leq \bar{w_e}^I(w_f)$.*

When the cost of exclusive capacity is too low or too high relative to w_f, that is $w_e < \underline{w_e}^I(w_f)$ or $w_e > \bar{w_e}^I(w_f)$, the capacity cost becomes the dominant force in firm's decision between exclusive and first-priority capacity. In this case, the outcome that both firms choose the less expensive option is Pareto-efficient: both choose exclusive capacity when $w_e < \underline{w_e}^I(w_f)$ and first-priority capacity when $w_e > \bar{w_e}^I(w_f)$. When the cost of exclusive capacity falls in between the two thresholds, $\underline{w_e}^I(w_f)$ and $\bar{w_e}^I(w_f)$, outcomes become inefficient. For instance, if $\underline{w_e}^I(w_f) \leq w_e \leq w_f$, both firms choose the exclusive capacity, but both firms can improve their profits if they choose the (more expensive) first-priority capacity together. This is because, although the first-priority capacity is more expensive, both firms are able to access the other one's leftover capacity, leveraging capacity pooling. Hence, both firms can build smaller capacity (i.e., $K_i^{ff} \leq K_i^{ee}$ from the part (ii) of Lemma IV.2), and earn a higher profit with the first-priority capacity. It should be also noted from Corollary IV.4 that a prisoner's dilemma only occurs when both firms invest in exclusive capacity and the cost of exclusive capacity is slightly lower than that of first-priority capacity. On the other hand, when the cost of exclusive capacity is high, at least one firm is incentivized to reserve the first-priority capacity. If $w_f < w_e \leq \bar{w_e}^I(w_f)$, a free-rider equilibrium arises because one firm chooses the exclusive capacity so that the other firm is willing to choose the first-priority capacity (part (ii) of Theorem IV.3). Figure 4.1 illustrates Theorem 1 and Corollary IV.4.

Figure 4.2 shows how the firms' equilibrium profits and capacity levels change as w_e increases. As described by Theorem 1, the equilibrium changes from $(\kappa_1, \kappa_2) = (e, e)$ when $\frac{w_e}{w_f} \leq 1$, to (e, f) when $1 < \frac{w_e}{w_f} \leq 1.05$, to (f, f) when $\frac{w_e}{w_f} > 1.05$. When $w_e \leq w_f$, both firms invest in exclusive capacity in equilibrium. Unless w_e is significantly smaller than w_f ($w_e/w_f < 0.88$ in the example in Figure 2), this equilibrium (e, e) is inefficient: although the exclusive capacity is cheaper than the first-priority capacity, both firms still benefit more from investing in first-priority and allowing the other firm to share the leftover. This enables both firms to reduce the total capacity level, but, at the same time, to access more capacity. On the other hand, when firm 1 invests in (or reserves) exclusive capacity and firm 2 invests in first-priority capacity, a free-rider outcome occurs that firm 1 can tap into the leftover capacity of firm 2.

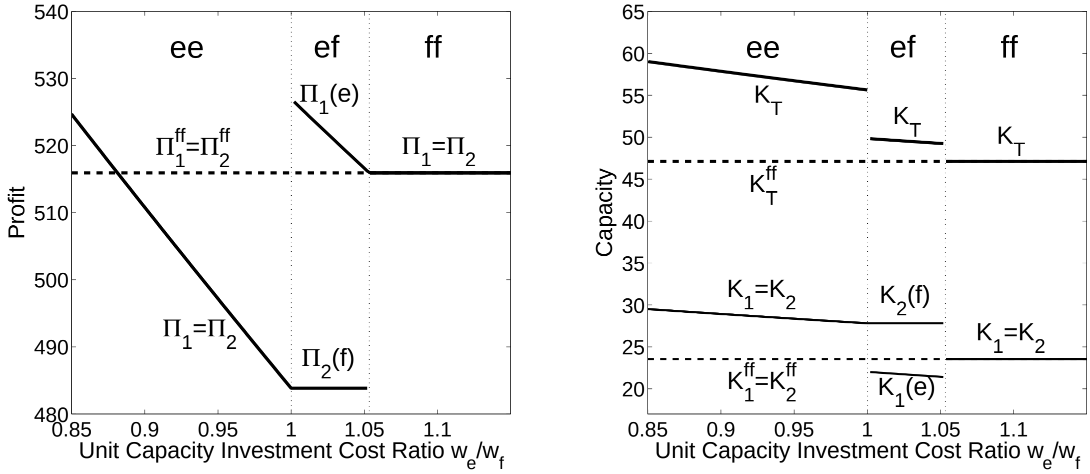

Figure 4.2: Independent market: Firms' profits (left panel) and capacities (right panel) under equilibrium (solid line) and non-equilibrium capacity type choices. Parameters: Price sensitivity $b_1 = b_2 = 1$; Wholesale cost $c = 1$; Market size θ_i follows a uniform distribution of $U[1, 100]$; Capacity cost $w_f = 9.5$.

In this situation, firm 1 invests in a smaller and more expensive exclusive capacity, and earns higher profit even when demands are symmetric. Here notice that this particular outcome occurs when exclusive capacity is slightly more expensive than first-priority capacity. This is illustrated in Figure 4.2 when $1 < \frac{w_e}{w_f} \leq 1.05$. Firm 2 invests in a larger first-priority capacity and earns lower profit. However, firm 2 still prefers this to reserving exclusive capacity because $w_f < w_e$ and the demands are independent.

The independent market case analyzed here illustrates the interplay of different forces in determining equilibrium outcomes, particularly, difference in capacity costs and the trade-off between sharing (unreliable) capacity and building its own capacity that the firm can use for certain. We now consider the Cournot market, where we find that these two forces continue to operate, but additionally, the dependence of the markets may influence the equilibrium outcome.

4.4 Cournot market

We now consider the Cournot market where both firms receive the perfectly correlated demand signal ($\theta_1 = \theta_2 = \theta$) and compete in quantity ($\eta_i = 1$). Specifically, if the two firms produce q_1 and q_2 upon observing the demand signal $\theta = a$ (following distribution $F(\cdot)$), the market clearing price is $p = a - b(q_1 + q_2)$. As before, firms may have different wholesale prices; without loss of generality we assume $c_1 \geq c_2$. If $c_1 = c_2$, firms are symmetric. If $c_1 > c_2$, two firms purchase the good at different prices. As in previous sections, we apply backward induction, starting from optimal quantities given capacity type, investment level, and realized demand signal. We present the subgame equilibrium when firm 1 has K_1 units of exclusive capacity and firm 2 has K_2 units of first-priority capacity. Other cases can be derived in a similar manner.

Lemma IV.5 (Firms' equilibrium order quantities)**.** *Suppose that firm 1 invests in K_1 units of exclusive capacity and firm 2 in K_2 units of first-priority capacity. The equilibrium order quantities for given demand signal $\theta = a$ are:*

	Conditions	*Firms' order quantity* (q_1^{ef}, q_2^{fe})
i)	$\max\{bK_1 + 2bK_2 + c_2, 2bK_1 + bK_2 + c_1\} \leq a$	$(K_1,\ K_2)$
ii)	$\max\{3bK_2 + 2c_2 - c_1, bK_2 + c_1\} \leq a < 2bK_1 + bK_2 + c_1$	$(\frac{a-c_1-bK_2}{2b},\ K_2)$
iii)	$2bK_2 + c_2 \leq a < bK_2 + c_1$	$(0,\ K_2)$
iv)	$\frac{3b(K_1+K_2)+c_1+c_2}{2} \leq a < bK_1 + 2bK_2 + c_2$	$(2(K_1 + K_2) - \frac{a-c_2}{b},\ \frac{a-c_2}{b} - K_1 - K_2)$
v)	$2c_1 - c_2 \leq a < \min\{\frac{3b(K_1+K_2)+c_1+c_2}{2}, 3bK_2 + 2c_2 - c_1\}$	$(\frac{a-2c_1+c_2}{3b},\ \frac{a-2c_2+c_1}{3b})$
vi)	$a < \min\{2c_1 - c_2, 2bK_2 + c_2\}$	$(0,\ \frac{a-c_2}{2b})$

For given capacity, the equilibrium quantities depend on demand signal $\theta = a$ and their wholesale prices c_1 and c_2 that firms need to pay. In the first three cases, firm 2 uses up all of his reserved capacity when the demand signal is very favorable ($\theta = a$ is large) relative to K_2. In this case, firm 1 uses her own capacity to produce (i) K_1, (ii) $\frac{a-c_1-bK_2}{2b}$, or (iii) 0, depending on the problem parameters. Case (iv) arises when a is moderately large and K_2 is not small compared to K_1 (notice that this case occurs only if $K_2 > K_1 + \frac{c_1}{b} - \frac{c_2}{b}$). Therefore, it is not optimal for firm 2 to use up his entire capacity, but firm 1 has incentive to use some leftover from firm 2. In case (v), the order quantities are not constrained by their initial capacities, thus both firms produce Cournot quantities. In case (vi), c_1 is too high or/and a is too low, only firm 2 produces a monopolistic quantity.

Once we derive optimal production quantities for each subgame, we determine firms' equilibrium capacity investments in the second-stage subgame as stated in the next lemma.

Lemma IV.6 (Firms' equilibrium capacities)**.** *Suppose that $K_1^{\kappa_1\kappa_2}$ and $K_2^{\kappa_2,\kappa_1}$ represent subgame-perfect equilibrium capacity when firm 1 chooses type κ_1 and firm 2 chooses type κ_2 capacity. Then, for $\kappa_1, \kappa_2 \in \{e, f\}$, $K_1^{\kappa_1\kappa_2}$ and $K_2^{\kappa_2\kappa_1}$ must satisfy the following:*

(i) If $w_e = w_f = w$, then $K_2^{\kappa_2\kappa_1}$ is any positive solution of

$$\int_{2bK+c_2+(bK+c_2-c_1)^+}^{\infty} (a - c_2 - 2bK - (bK + c_2 - c_1)^+)\mathbf{d}F(a) - w = 0$$

and $K_1^{\kappa_1\kappa_2} = (K_2^{\kappa_2\kappa_1} + \frac{c_2}{b} - \frac{c_1}{b})^+$.

(ii) If $w_e > w_f$, then $K_1^{ef} \leq (K_2^{fe} + \frac{c_2}{b} - \frac{c_1}{b})^+$, $K_1^{fe} \geq (K_2^{ef} + \frac{c_2}{b} - \frac{c_1}{b})^+$, and $K_i^{ee} \leq K_i^{ff}$.

(iii) If $w_e < w_f$, then $K_1^{ef} \geq (K_2^{fe} + \frac{c_2}{b} - \frac{c_1}{b})^+$, $K_1^{fe} \leq (K_2^{ef} + \frac{c_2}{b} - \frac{c_1}{b})^+$, and $K_i^{ee} \geq K_i^{ff}$.

It is interesting to note that when capacity costs are the same, $w_e = w_f$, the equilibrium capacity does not depend on capacity type. The result is quite different from the results of independent market case (Lemma IV.2): when two firms choose first-priority capacity, both invest to build smaller capacity than what they would when choosing exclusive capacity. To see why this difference occurs, notice that in Cournot market, demand signals are perfectly correlated. Hence, if it is better for firm i to produce one more unit by tapping in firm j's leftover capacity, firm j is also better off using that unit of capacity to produce one more unit for himself. Therefore, for any realization of the demand, either there exists leftover that no firms want to use, or the capacity is exhausted. As a result, the equilibrium capacity does not depend on the capacity type.

Building on the analysis of the first two stages, we present firms' capacity type choice.

Theorem IV.7 (Equilibrium capacity type choices)**.** *In equilibrium,*
(i) if $w_e > w_f$, both firms invest in first-priority capacity,
(ii) if $w_e = w_f$, firms are indifferent between exclusive and first-priority capacities,
(iii) if $w_e < w_f$, both firms invest in exclusive capacity.

In contrast to the independent market, choosing a cheaper capacity is a dominant strategy in the Cournot market. As a result, the asymmetric equilibria, (e, f) or (f, e), vanish. To see why this must be the case, consider the $w_e > w_f$ case first. Suppose firm j chooses the exclusive capacity; we show that firm i always prefers the first-priority capacity. At first, this is not immediately obvious. Although reserving first-priority capacity is cheaper, it may allow the other firm to feast on any leftover of firm i. However, notice that in the Cournot market, the demand signals that two firms receive are identical ($\theta_1 = \theta_2 = \theta$). The perfect positive correlation of demand signals mitigates most of the spill-over: when firm j faces a high demand and exhausts his capacity, firm i also faces a high demand, and therefore, firm i has either no or very little leftover. The only case that firm i has ample leftover is when θ is low; but in this case, firm j is not in need of excess capacity either. In addition, the first-priority capacity is cheaper, and firm i cannot access firm j's leftover anyway, so choosing the first-priority enables firm i to build more capacity for herself. Similar issues arise for all the other cases. In all cases, the perfect correlation of the demand signal, and the fact that in the Cournot market both players access the same market, make capacity pooling an unimportant factor when choosing the capacity type. As a result, the firms always choose the cheaper option.

In particular, we can show that even when both firms choose the first-priority capacity in equilibrium, neither firm will actually tap into the other firm's leftover for any realization of the demand signal. Consequently, there is no free rider effect in equilibrium, making the equilibrium outcome *de facto* exclusive. Combining this observation with the result of Theorem 5, we have a prisoner's dilemma does not occur in the Cournot market. We formally summarize the result in the following corollary.

Corollary IV.8. *Neither the free-rider equilibrium nor the prisoner's dilemma equilibrium arises in the Cournot market.*

4.5 Supplier's profit

Our analysis so far has focused on buying firms' decisions and profits while implicitly assuming that the supplier has no input on any decision. While this scenario is appropriate for a weak supplier, a supplier with strong market power may assume an active role in deciding the form of capacity investment that buyers invest in. In this section, we conduct a detailed analysis of the supplier's profit, preferred outcomes, and capacity type choices if the supplier can strategically offer the capacity type.

The supplier's profit has two components: profit from capacity investment or reservation (a γ fraction of the total investment amount, for exogenous $\gamma \in [0, 1]$), and profit from production (this equals c_i per unit). We denote these two components by Π_{sc} and Π_{sp} respectively, and the supplier's total profit as Π_s. For given investment types (κ_1, κ_2) and capacity level (K_i, K_j), the supplier's expected profit can be written as follows:

$$\begin{aligned}\Pi_s^{\kappa_1\kappa_2}(K_i,K_j) =& \Pi_{sc}^{\kappa_1\kappa_2}(K_i,K_j)+\Pi_{sp}^{\kappa_1\kappa_2}(K_i,K_j)\\ =&\gamma\sum_{i=1}^{2} w_{\kappa_i}K_i + E_{(\theta_i,\theta_j)}\left[\sum_{i=1}^{2} c_i q_i^{\kappa_i\kappa_j}(\theta_i,\theta_j,K_i,K_j)\right]\end{aligned} \tag{4.9}$$

where the reservation profit $\Pi_{sc}^{\kappa_1\kappa_2}(K_i,K_j) = \gamma\sum_{i=1}^2 w_{\kappa_i}K_i$, the production profit $\Pi_{sp}^{\kappa_1\kappa_2}(K_i,K_j) = E_{(\theta_i,\theta_j)}\left[\sum_{i=1}^2 c_i q_i^{\kappa_i\kappa_j}(\theta_i,\theta_j,K_i,K_j)\right]$, and the quantity, $q_i^{\kappa_i\kappa_j}(\theta_i,\theta_j,K_i,K_j)$, is the subgame-perfect production quantity after demand signals are realized.

We first examine the *passive* supplier's profit when the supplier accepts buyer's capacity type choices. We then compare this with the case where the supplier *actively* chooses capacity types.

4.5.1 The profit of a passive supplier

We first analyze the independent market case. As in Section 3, we assume $c_1 = c_2 = c$ without loss of generality.

Theorem IV.9 (Supplier's subgame-perfect expected profit: independent market)**.** *Suppose firm i chooses capacity type κ_i and firm j chooses type κ_j. Let $K_i^{\kappa_i\kappa_j}$ and $K_j^{\kappa_j\kappa_i}$ be the subgame perfect equilibrium capacity levels. If $K_i^{\kappa_i\kappa_j} > 0$ and $K_j^{\kappa_j\kappa_i} > 0$, then, the supplier's expected profit is as follows*[1]*:*

$$\Pi_s^{\kappa_1\kappa_2} = \Pi_{sc}^{\kappa_1\kappa_2} + \Pi_{sp}^{\kappa_1\kappa_2} = \gamma\textstyle\sum_{i=1}^2 w_{\kappa_i}K_i^{\kappa_i\kappa_j} + c\sum_{i=1}^2\left[\int_c^\infty \frac{a_i-c}{2b_i}\mathbf{d}F_i(a_i) - \frac{w_{\kappa_i}}{2b_i}\right]$$

where $\Pi_{sc}^{\kappa_1\kappa_2} = \gamma\sum_{i=1}^2 w_{\kappa_i}K_i^{\kappa_i\kappa_j}$ *and* $\Pi_{sp}^{\kappa_1\kappa_2} = c\sum_{i=1}^2\left[\int_c^\infty \frac{a_i-c}{2b_i}\mathbf{d}F_i(a_i) - \frac{w_{\kappa_i}}{2b_i}\right]$.

We observe that the supplier's total profit changes non-monotonically in the unit capacity cost w_{κ_i}. The first part of the supplier's profit, the reservation profit (Π_{sc}), changes non-monotonically with respect to w_{κ_i}, because an increase in capacity cost

[1] If $K_i^{\kappa_i\kappa_j} = K_j^{\kappa_j\kappa_i} = 0$, it is trivial. If $K_i^{\kappa_i\kappa_j} = 0$ and $K_j^{\kappa_j\kappa_i} > 0$, similar expressions can be derived.

results in a decrease in the amount of capacity reserved. Therefore, although the unit reservation profit (γw_{κ_i}) increases as the capacity cost increases, the total reservation profit may not necessarily increase. However, the supplier's profit from production (Π_{sp}) linearly decreases in capacity costs. Interestingly, we find that when the capacity costs are the same, the supplier's expected profit from production remains the same regardless of the capacity types the two firms choose, as highlighted in the following corollary.

Corollary IV.10. *Suppose $w_e = w_f$. If $K_i^{\kappa_i \kappa_j} > 0$ and $K_j^{\kappa_j \kappa_i} > 0$, $\Pi_{sp}^{\kappa_1 \kappa_2}$ remains the same for all (κ_1, κ_2), where $\kappa_i \in \{e, f\}$ for $i = 1, 2$.*

At first, this is counter-intuitive since the optimal investment level depends on the capacity type. To see why this happens, we first compare the supplier's profit from production between the two cases—$(\kappa_i, \kappa_j) = (e, e)$ and $(\kappa_i, \kappa_j) = (f, e)$. Since demands are independent and firm i cannot access firm j's leftover when firm j chooses the exclusive capacity, we have $K_i^{ee} = K_i^{fe}$ from Lemma IV.2(ii). From the same lemma, we also have $K_j^{ee} \geq K_j^{ef}$: firm j builds smaller capacity when firm j can access firm i's leftover. Although the total capacity is lower in the (f, e) case, it can be shown that the supplier's gain from pooling under the (f, e) case is exactly equal to the loss from the lowered total capacity in equilibrium. As a result, Π_{sp} remains the same in both cases although the firms' profits are different. We then compare the (e, e) and (f, f) cases. From lemma IV.2(ii), $K_i^{ee} \geq K_i^{ff}$. However, again the benefit from pooling is equal to the loss from the lowered total capacity, the supplier's profit from production remains the same.

Utilizing Theorem 1 and Theorem IV.9, we now examine how the supplier's equilibrium profit changes as the cost of exclusive capacity w_e increases. As illustrated in Figure 4.3(A), the supplier's profit changes non-monotonically in w_e. In fact, the supplier's profit is the largest under the (e, e) regime ($\frac{w_e}{w_f} \leq 1$) and the smallest under the (f, f) regime ($\frac{w_e}{w_f} \geq 1.05$) while the profit is non-decreasing within each of the

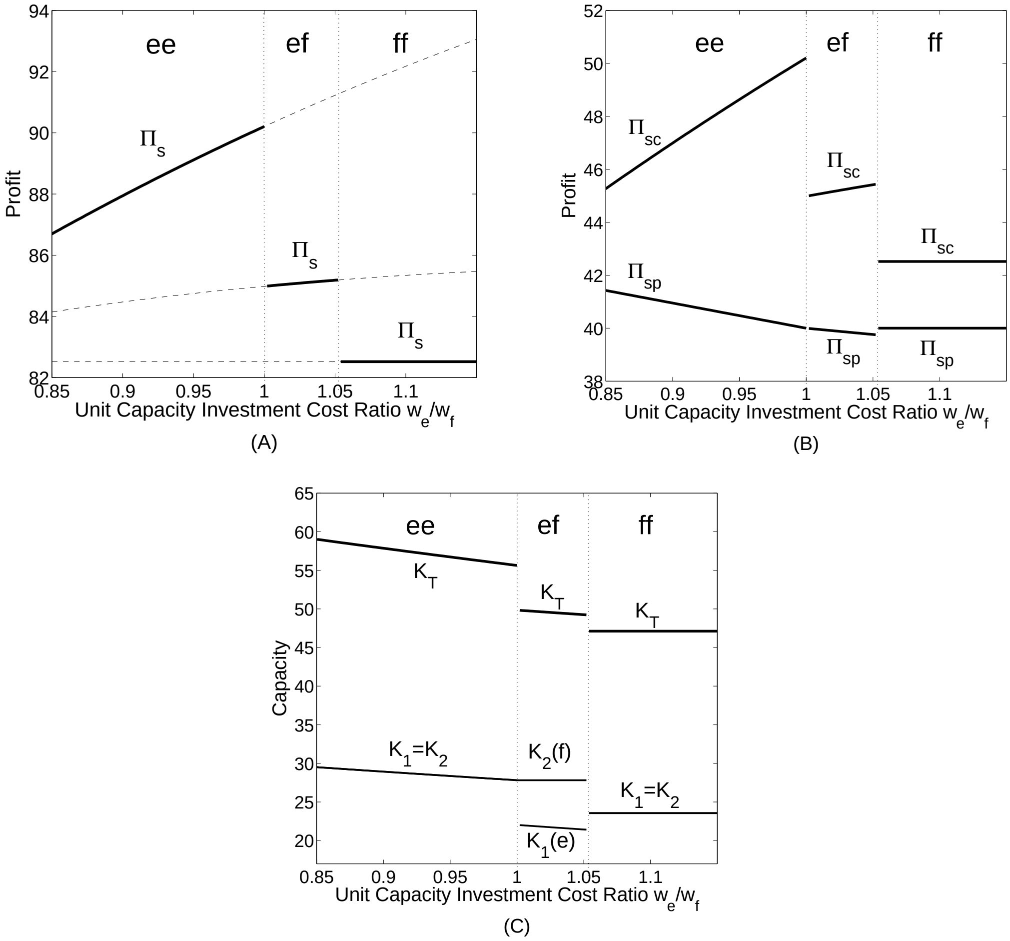

Figure 4.3: Independent market: (A) passive supplier's total equilibrium profit, (B) subgame-perfect production profit and capacity investment profit, and (C) equilibrium capacities. Parameters: Price sensitivity $b_1 = b_2 = 1$; Wholesale cost $c = 1$; $\gamma = 0.095$ fraction of capacity investment accounted as the supplier's profit; Market size θ_i follows a uniform distribution of $U[1, 100]$; Capacity cost $w_f = 9.5$.

three regimes (strictly increasing in (e, e) and (e, f) regimes). The figure also shows that the capacity types that the buying firms choose are not necessarily the ones that benefit the supplier. For instance, when $w_e > 1.05w_f$, the supplier could have earned higher profit if at least one firm chooses the exclusive capacity. However, the high cost of exclusive capacity forces both firms to choose the first-priority.

Figure 4.3(B) breaks down the supplier's total profit into its two components: reservation profit (Π_{sc}), and production profit (Π_{sp}). We first observe that the production profit decreases within the (e, e) and (e, f) region. This is because, as w_e increases, the firm(s) choosing the exclusive capacity will invest less, and reduce the quantity that the supplier can produce. On the other hand, the profit from production increases when the regime switches from (e, f) to (f, f) as the supplier gains significantly from pooling and makes the better use of capacity. Notice that in this case, the total capacity is indeed smaller under (f, f) than the total capacity under (e, f) (see Figure 4.3(C)), but the gain from pooling outweighs any loss from smaller capacity.

On the other hand, the profit from capacity reservation increases in w_e as long as the equilibrium capacity types remain the same. Although a higher cost of exclusive capacity forces firm(s) to choose a lower level of exclusive capacity (see the graph within each of the three regions in Figure 4.3(C)), the firms need to spend more to buy capacity, which increases the supplier's reservation profit. However, the capacity reservation profit drops sharply in w_e when the equilibrium shifts from (e, e) to (e, f) because firm 1 reduces her capacity level and free-rides on firm 2's leftover capacity: see Figure 4.3(C). The profit further drops when the equilibrium shifts from (e, f) to (f, f) as pooling reduces the total capacity level and resultant reservation profit: see Figure 4.3(C).

We observe similar phenomena in the Cournot markets: as capacity cost increases, the production profit weakly decreases within each region while the profit from capacity reservation can change non-monotonically. Also, as in the independent market, the supplier's profit is not necessarily monotone.

4.5.2 Strategic (active) supplier's capacity type choices

We now consider the scenario under which the supplier can choose the capacity type for buying firms in the first stage. Other than this change, everything else remains the same as in the base model. To highlight the result that arises from difference in the capacity type and avoid the convoluted impact of capacity costs, we first focus on the case $w_e = w_f$: the capacity costs are the same for both types. It turns out that the supplier prefers different capacity types depending on the market.

Theorem IV.11 (Strategic supplier's capacity type choices).
Suppose $w_e = w_f = w$. If $K_i^{\kappa_i \kappa_j} > 0$ and $K_j^{\kappa_j \kappa_i} > 0$,
(i) (Independent market) The supplier always offers the exclusive capacity to both firms.
(ii) (Cournot market) The supplier is indifferent between the two capacity types.

In the independent market, the supplier is better off by offering exclusive capacity because doing so forces both firms to build large capacity (Lemma IV.2), increasing the supplier's profit. However, it should be noted that the increase in profit is entirely from capacity reservation. In fact, as Corollary IV.10 shows, the production profit remains the same, and is independent of capacity type as long as the capacity costs are the same. In contrast, the capacity type does not play any role in the Cournot market. Because the demand signal is perfectly correlated (i.e., $\theta_1 = \theta_2 = \theta$), both firms choose the same capacity level and produce the same quantity regardless of the capacity types.

Next, when the capacity costs are different, the supplier's capacity type choice becomes more complicated as production profit and reservation profit (which changes non-linearly in capacity cost) interplay. We illustrate the supplier's capacity choices in the independent market using a numerical example. We fix all the market environments the same and only change the capacity cost w_e. In this case, the supplier may offer (e, e), (e, f), or (f, f) to the firms as shown in the left panel of Figure 4.4. When w_e is sufficiently small ($\frac{w_e}{w_f} \leq 0.40$) and/or the first-priority capacity is expensive, the supplier gains the most from capacity investment if both firms reserves the first-priority capacity. Therefore, the supplier should offer the first-priority capacity to both firms. In the middle region ($0.40 < \frac{w_e}{w_f} \leq 1.38$), the supplier prefers offer-

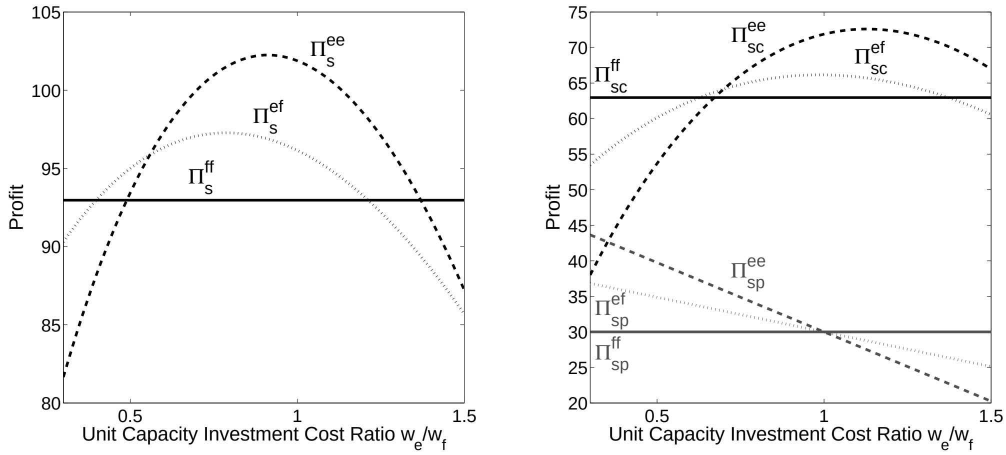

Figure 4.4: Independent demand: Supplier's total profits (left panel) and profits from production and capacity investment (right panel). Parameters: Price sensitivity $b_1 = b_2 = 1$; Wholesale cost $c = 1$; $\gamma = 0.1$; Market size θ_i follows a uniform distribution of $U[1,\ 100]$; Per unit capacity investment cost $w_f = 19.5$.

ing the exclusive capacity to one firm and the first-priority capacity to another firm, then offering the exclusive capacity to both firms as w_e increases. When w_e becomes sufficiently large ($\frac{w_e}{w_f} > 1.38$), the exclusive capacity is too expensive to build. In this case, the supplier should offer the first-priority capacity to both firms, because any other scenarios where at least one firm reserves exclusive capacity will result in a significant drop in the production quantity.

To further understand this, from the right panel of Figure 4.4, we first observe that the supplier earns the highest production profit ($\Pi_{sp}^{\kappa_i\kappa_j}$) when both firms are offered the cheaper capacity type as shown in Theorem IV.9. However, the ranking of the supplier's reservation profit ($\Pi_{sc}^{\kappa_i\kappa_j}$) under different capacity types changes with respect to w_e as a result of the interaction between the unit capacity cost and resultant capacity levels that firms choose in equilibrium. Specifically, as w_e increases, the unit reservation profit (i.e., γw_e) increases but the firm with exclusive capacity will decrease the invested capacity accordingly. When w_e is sufficiently small, relative to w_f, the supplier gains the most when both invest in first-priority capacity (Π_{sc}^{ff}), because the unit capacity profit from exclusive capacity is too small and therefore the reservation profit (Π_{sc}^{ee} and Π_{sc}^{ef}) is small. Likewise, Π_{sc}^{ff} is also the largest when w_e is very large, because the firm(s) will build very small exclusive capacity, making Π_{sc}^{ee} and Π_{sc}^{ef} small. In the middle region, offering exclusive capacity to at least one firm (Π_{sc}^{ee} and Π_{sc}^{ef}) achieves the largest reservation profit for the supplier. Therefore, the tradeoff between production profit and reservation profit yields the outcomes observed for the supplier's capacity choices in the left panel of Figure 4.4.

Finally, we compare the outcomes with a strategic supplier to the outcomes with a passive supplier. In both cases, we first consider $w_e = w_f$. In the independent market, the outcome remains to be (e, e). Note that, in the base model, this is when both firms fall in the prisoner's dilemma by over-investing in exclusive capacity. While this is not a desired outcome for the buying firms, the supplier benefits from the firms' overinvestment. In the Cournot market, both firms and the supplier are indifferent between the two capacity types as the perfectly correlated demand signal and symmetric nature of competition force both firms to choose the same capacity level and order the same amount under different capacity choices (the capacity usage is *de facto* exclusive).

Now consider when capacity costs are different. In the independent market, when $w_e \leq w_f$, the strategic supplier may offer the first-priority capacity to at least one firm while both firms prefer to choose the exclusive capacity. On the other hand, when $w_e > w_f$, the equilibrium with the passive supplier is that at least one firm should prefer the first-priority capacity, but the strategic supplier may offer exclusive capacity to both firms because of high capacity reservation profit. Similar results are observed in the Cournot market. The supplier's preference of capacity types depends on the tradeoff between production profit and reservation profit, thus the outcomes with the passive supplier is different from the outcomes with the strategic supplier.

4.6 Conclusion

In this chapter, we analyzed capacity investment decisions of firms sharing a common supplier. We identify three equilibria and show that, as the cost of reserving capacity exclusively increases, the equilibrium shifts from both firms choosing the exclusive capacity, to one firm choosing the exclusive capacity and the other firm choosing the first-priority capacity, to both firms choosing the first-priority capacity. While this transition is quite robust to the change in firm's demand or procurement cost, the underlying relationship between two firms' demands can influence which equilibrium arises. For example, when both firms compete in the Cournot market, an asymmetric equilibrium does not arise even when the two firms have different costs. In contrast, such equilibrium can arise even for two symmetric firms when demands are independent.

We identify two cases where the firms' investment can be inefficient. When the cost of exclusive capacity is slightly higher than the cost of first-priority capacity and demands are independent, a *free-rider* equilibrium arises in which one firm builds large first-priority capacity and the other firm with small exclusive capacity utilizes the leftover capacity. This is surprising since one firm deliberately chooses the more expensive capacity (exclusive) to induce the other firm to build larger capacity (first-priority) even when the demand of one firm is independent of the demand of the other firm. On the other hand, if the cost of exclusive capacity is the same or slightly lower than the first-priority capacity, firms build large exclusive capacity instead of pooling capacity, resulting in a *prisoner's dilemma*. Surprisingly, this occurs when exclusive capacity is cheaper than first-priority capacity. Therefore, demanding the exclusive right to use the capacity does not necessarily serve the investing firm's best interest. Once again, the interdependence of demands plays a critical role: While both adverse outcomes arise when demands are independent, neither arises in the Cournot market. In this case, demand correlation allows firms to choose first-priority capacity instead of the more expensive exclusive capacity without worrying about spillover.

We find that when the supplier can choose the capacity type to buying firms, the capacity type that the supplier prefers can be different from the capacity type that firms prefer. Again, how two firms' demands are related plays a crucial role. When capacity costs are the same and demands are independent, the supplier offers exclusive capacity so that he can induce firms to overinvest. In the Cournot market, surprisingly, the supplier becomes indifferent between two capacity types as the value of capacity pooling diminishes when two firms receive positively correlated demand signals. When the capacity costs are different, we illustrate with the independent market that the strategic supplier may offer exclusive capacity to both firms, while both firms prefer first-priority capacity, or offer first-priority capacity to both firms while at least one firm prefers exclusive capacity. Our results indicate that the difference in capacity costs between the two types critically determines the nature of the equilibrium. Thus, even if the supplier cannot enforce the capacity type for the firms, the supplier may still be be able to induce the the desired outcomes by offering a discount or charging a premium on exclusive (or first-priority) capacity.

4.6.1 Managerial insights and discussion

Our work yields important managerial insights that have potential impact on practitioners. First, linking back to our motivating examples, our results provide a parsimonious explanation about firms' capacity type choices when investing in a shared supplier. Under Cournot competition as in the agriculture supply company case, the firm competes in an almost homogeneous market and the first-priority capacity acts as the exclusive capacity, because depending on the demand realizations, either all firms are short on capacity or all firms have too much capacity. Therefore, although the firm chooses the first-priority capacity, it is not likely that the other firm will access her leftover capacity. In contrast, in the independent market, firms may choose the exclusive capacity even when it is more expensive than the first-priority capacity, if doing so forces the other firm to build a large first-priority capacity that the firm may tap into. Second, our work highlights the importance of the capacity cost. If the supplier and firms do not distinguish the two types of capacity, i.e., the costs for the two capacity types are the same, firms tend to choose the exclusive contract. However, our results identify the inefficiency that occurs in this very case: firms overinvest in capacity in the independent market, and could have earned a higher profit if they choose the first-priority capacity together. Therefore, if the firms are able to negotiate for a discount on the first-priority capacity, at least one and possibly both firms will turn to the first-priority capacity, and profit can be improved.

Our work raises the following question: besides offering discount on the first-priority capacity, can the firms and the supplier form an agreement to (partially) remove these inefficiencies, and find an investment strategy which yields a better outcome for both firms (or even all three parties)? One potential way to achieve such a "co-ordination" is *transfer payment*: if firm i invests in the first-priority capacity which then gets used by firm j, then firm j pays firm i a per-unit rate τ for accessing this capacity. This transfer payment is essentially the *tradable capacity option* in the capacity reservation setting, where the tradable capacity option implies that buying firms with capacity reservation have the right to trade the reserved capacity among themselves. As mentioned in Plambeck and Taylor (2007), Taiwan Semiconductor Manufacturing Corporation, the largest semiconductor contract manufacturer, pioneered selling tradable capacity options (LaPedus, 1995; *Economist*, 1996).

Such transfer payments, while on the surface seemingly very attractive, are not a straightforward solution to the problem. On one hand, the transfer payment certainly encourages firms to share capacity. It encourages firms to choose the first-priority capacity, because of the additional profit from the transfer payment when the other firm uses the leftover capacity. Similarly, it discourages firms from choosing the exclusive capacity, because firms need to pay an extra fee to access the other firm's leftover, while the firm cannot earn this extra income as the other firm cannot access her leftover. In fact, we can analytically show that when an exogenous positive transfer fee τ is charged per unit of the invested capacity used by another firm, the region of both firms choosing the exclusive capacity, as well as the region of prisoner's dilemma, shrinks.

Paradoxically, however, the transfer payment also makes it more difficult for firms to share capacity. Compared with the base case where there is no transfer payment (which is consistent with our observation of the agriculture supply company case and the computer hard drive industry case identified in Li et al. (2011)), allowing a transfer payment decreases the efficiency of the system because firms now need to pay to access the other firm's leftover capacity. So, they may instead choose to increase their own capacity which they can access without making an additional transfer payment, resulting in greater potential for installing wasted capacity. In fact, it can be shown that when the transfer price is sufficiently high, firms will find it non-profitable to use the other firm's leftover, and therefore both capacity types end up becoming exclusive.

An additional level of complication arises from the fact that ideally the transfer price should be set endogenously, rather than exogenously. With the aforementioned two conflicting forces operating, and given the complexity of the current model, adding an additional layer to decide the optimal transfer payment renders our model too intractable to analyze. In fact, an appropriate approach to decide the optimal transfer payment might be a three-party negotiation among the two buying firms and the supplier. Although this is beyond the scope of this chapter, we believe this open research question to determine the optimal transfer payment has significant potential to contribute to the stream of literature. We believe the literature on structuring investments in the supplier, in the presence of market competition, is still developing, and offers several research questions of potentially high impact to theory and practice.

In Chapter III, we consider the setting where the demand is deterministic and firms compete in a Cournot market, while the capacity is random and subject to production yield. In addition, the capacity cost has two parts: the fixed cost and the variable cost. In Chapter IV, we consider another setting where the demand is uncertain and firms may serve two independent markets or compete in a Cournot market, while the capacity to build is deterministic. In this case, the capacity cost is linear. While these two settings capture the characteristics of different market structures and provide clean characterizations of the equilibrium outcomes, we conjecture that the insights remain intact in more general settings.

For example, one may consider a setting where both the demand and capacity are stochastic and the capacity investment cost has both the fixed and variable parts. In this setting, firms still face the following key tradeoffs among others. First, between the exclusive and first-priority capacity, firms need to balance the benefit from pooling the demand uncertainty and the cost to build up or reserve each type of capacity. Second, the combination of the fixed and variable capacity costs determines whether firms should invest in the supplier, and if so, how much capacity should be invested in. As a result, the spillover when one firm decides to invest while the other decides not to under the first-priority capacity is also expected. Finally, the stochastic capacity is expected to motivate firms to invest more aggressively than when the capacity is deterministic. However, as the setting is rather complicated with a mixture of the stochastic demand, stochastic capacity, and fixed and variable capacity costs, one may need to carefully identify the driving force of each phenomenon identified.

CHAPTER V

Conclusions

To summarize the dissertation, we investigate capacity management problems in a supply chain setting. At an intra-firm level, we use stochastic dynamic programming and data-driven optimization to develop tools to facilitate a firm's capacity adjustment decisions. At an inter-firm level, we use game-theoretic methods to analyze firms' interactions with their shared supplier leveraging capacity decisions as a strategic tool to gain competitiveness. Analytically, the dissertation pushes the boundary of analytics to support firms' capacity adjustment decisions. Managerially, the dissertation further extends insights about potential adverse outcomes in capacity management in a network setting, and highlights directions to fix these inefficiencies.

More specifically, in Chapter II, we study a firm's optimal strategy to adjust its capacity using demand information. We consider two scenarios: the firm has a *single* or *multiple* opportunities to adjust its capacity, reflecting the impact of different capacity adjustment leadtime and costs. For both scenarios, we first formulate the problem as a stochastic dynamic program, and then characterize the firm's *optimal* policy: when to adjust and by how much. We show that the optimal policy can be counter-intuitive and complex. For example, when the firm can change the capacity only once, the optimal decision on when and by how much to change the capacity is not monotone in the likelihood of high demand. This phenomenon is particularly surprising as the corresponding problem is an optimal stopping problem, where monotonicity of the optimal policy is quite common. When the firm can adjust the capacity multiple times, we characterize the optimal policy as a control band policy. As it is challenging to compute and implement the optimal policy, we develop a data-driven heuristic for each scenario. In the single adjustment scenario, we show that a heuristic which explores demand for an appropriately chosen length of time and adjusts the capacity based on the observed demand is *asymptotically optimal*, and characterize the *convergence rate*. In the multiple adjustment scenario, we show that a heuristic under which the firm adjusts its capacity at a predetermined set of periods with exponentially increasing gap between two consecutive decisions is asymptotically optimal. We finally apply our heuristics to a numerical study inspired by real data and business scenario and demonstrate the performance and robustness of the heuristics.

In Chapter III, we consider what happens when two competing firms invest in a shared supplier's capacity. A firm that invests in the supplier may gain additional capacity, but market competition and spillover of the investment may wipe out any gain from the investment. To protect firms' investments, many impose restrictions on how the supplier's capacity is used. We model firms' investment and production decisions as a two-period game, and analyze the equilibrium outcomes under two common forms of restrictions: the investing firm has *exclusive* use of the invested capacity, or *first priority* in having the firm's order fulfilled. We characterize the equilibrium capacity investment outcomes in terms of the number of investing firms and capacity investment levels, and identify conditions under which the spillover effect occurs, where one firm is able to tap into the other firm's invested capacity. We find that although on the surface the spillover effect seems to intensify competition, it indeed demotivates firms from investing in the supplier. Therefore, the spillover effect results in fewer investing firms and lower total capacity investment, which consequently mitigates the risk of both firms being trapped in a prisoner's dilemma, where both firms invest but both earn a lower profit than they do when neither invests. We also characterize the firms' and supplier's preference about the capacity types, and analyze the efficiency loss in the supply chain. While the non-investing firm always prefers the first-priority capacity hoping to benefit from the spillover effect, the investing firm does not always want to shut off the non-investing firm's access to its leftover capacity, especially when allowing spillover results in strictly fewer investing firms. The supplier's preference is driven by the tradeoff between the over-investment with exclusive capacity and the smaller but flexible investment with first-priority capacity. We also find that both the spillover effect and the wholesale price reduce over-investment in the supplier.

In Chapter IV, we study the capacity decisions (types and capacity levels) that firms choose when multiple firms invest in a shared supplier. We examine two specific capacity reservation agreements that are widely used in practice: *exclusive* (capacity can only be used by the investing firm exclusively) and *first-priority* (the leftover capacity can be used to fill other firms' order). We model this relationship as a multi-stage game and characterize the capacity type and investment level that firms choose in equilibrium. We identify three equilibria—both firms investing in exclusive capacity, both firms investing in first-priority capacity, and one firm investing in exclusive and the other in first-priority capacity—and explain the conditions under which each of the three equilibria arises as a function of capacity costs and market parameters that govern uncertain demands. We also provide conditions under which two inefficient outcomes occur: a prisoners' dilemma in which firms over-invest in exclusive capacity and a free-rider equilibrium under which one firm under-invests in anticipation of feasting on the other firm's leftover. We find that the capacity type that the supplier prefers is determined by the interplay between two opposite forces: gain from overinvesting and gain from capacity pooling. We show that the capacity types that the supplier prefers do not coincide with that firms prefer. There are some unexpected findings. First, in contrast to conventional intuition, demanding exclusive use is not necessarily optimal for the firm even when firms have independent demands. Second, even when demands are independent, it is possible for a firm to invest in more expensive capacity to gain exclusivity (as opposed to cheaper first-priority capacity) in order to induce the other firm to build a larger capacity. Third, surprisingly, these inefficiencies entirely disappear when demands are highly correlated and firms are competing (e.g., Cournot setting). This means that in some competitive settings, paying premium to gain exclusive capacity may be a waste of money.

The dissertation suggests several other directions for future research. Along the stream of intra-firm capacity management in Chapter II, we believe there are a lot more opportunities in the area combing learning and capacity management. For example, how does the firm's learning opportunity affect the joint decision of capacity and inventory? How should the firm compute its capacity management strategy efficiently when the product life cycle is not stationary? Along the stream of inter-firm capacity management in Chapter III and IV, one could consider equilibrium outcomes if investments are used to reduce the uncertainty in the yield, rather than increase capacity. At a more general level, with the increasing trend of decentralized, networked, yet cooperative, supply chains, the inter-dependency of the contractual relationship for one pair of agents with other agents could lead to unexpected outcomes. For example, what are the consequences of acquiring a supplier in such an environment? We believe there are several fruitful opportunities for research in both areas.

APPENDICES

APPENDIX A

Proofs and Technical Details for Chapter II: Capacity Investment with Demand Learning

In the proofs we focus on the case where the demand distribution is discrete. When the demand distribution is continuous, similar proofs hold.

Proof of Lemma II.1. We first show that for any j, we have $E[\boldsymbol{\Pi_{j+1}}|\boldsymbol{\Pi_j}] = \boldsymbol{\Pi_j}$. From equation (2.1), we have

$$\begin{aligned} E[\Pi_{j+1,i}|\boldsymbol{\Pi_j}] &= \sum_{d_j=0}^{\infty} \frac{\Pi_{j,i} f_j(d_j|\theta_i)}{\sum_{k=1}^{I} [\Pi_{j,k} f_j(d_j|\theta_k)]} Pr(D_j = d_j|\boldsymbol{\Pi_j}) \\ &= \sum_{d_j=0}^{\infty} \frac{\Pi_{j,i} f_j(d_j|\theta_i)}{\sum_{k=1}^{I} [\Pi_{j,k} f_j(d_j|\theta_k)]} \sum_{k=1}^{I} [\Pi_{j,k} f_j(d_j|\theta_k)] \\ &= \Pi_{j,i} \sum_{d_j=0}^{\infty} f_j(d_j|\theta_i) = \Pi_{j,i} \end{aligned} \tag{A.1}$$

That is, $E[\boldsymbol{\Pi_{j+1}}|\boldsymbol{\Pi_j}] = \boldsymbol{\Pi_j}$. Then for any $j_1 < j_2$, we have

$$\begin{aligned} E[\boldsymbol{\Pi_{j_2}}|\boldsymbol{\Pi_{j_1}}] &= E[E[\boldsymbol{\Pi_{j_2}}|\boldsymbol{\Pi_{j_2-1}}, \boldsymbol{\Pi_{j_1}}]|\boldsymbol{\Pi_{j_1}}] = E[E[\boldsymbol{\Pi_{j_2}}|\boldsymbol{\Pi_{j_2-1}}]|\boldsymbol{\Pi_{j_1}}] \\ &= E[\boldsymbol{\Pi_{j_2-1}}|\boldsymbol{\Pi_{j_1}}] \end{aligned} \tag{A.2}$$

Applying the above equations iteratively, we have $E[\mathbf{\Pi_{j_2}}|\mathbf{\Pi_{j_1}}] = \mathbf{\Pi_{j_1}}$. $\square$

Proof of Proposition II.3. For ease of exposition we define

$$G_j(\boldsymbol{\pi}, \mu) \triangleq \sum_{i=j+l}^{J} h_i(\boldsymbol{\pi}, \mu) - \hat{C}(\mu_0, \mu) \tag{A.3}$$

It is observed that for given μ, we have $G_j(\boldsymbol{\pi}, \mu)$ is linear in $\boldsymbol{\pi}$. For given $\boldsymbol{\pi}$, if $\mathcal{K} = \mathbb{R}^+$ and $\mu \in \mathcal{K}$, we have $G_j(\boldsymbol{\pi}, \mu)$ is concave in μ; if $\mathcal{K} = \{\delta_k, k = 1, 2, ..., |\mathcal{K}|\}$, we define $\Delta G_{j,k}(\boldsymbol{\pi})$ as follows:

$$\Delta G_{j,k}(\boldsymbol{\pi}) \triangleq \frac{G_j(\boldsymbol{\pi}, \delta_{k+1}) - G_j(\boldsymbol{\pi}, \delta_k)}{\delta_{k+1} - \delta_k} \text{ for } k = 1, 2, ..., |\mathcal{K}| - 1. \tag{A.4}$$

For given j and $\boldsymbol{\pi}$, we have $\Delta G_{j,k}(\boldsymbol{\pi})$ is a decreasing sequence in k. In addition, we have $G_j(\boldsymbol{\pi}, 0) < \infty$ and $\lim_{\mu\to\infty} G_j(\boldsymbol{\pi}, \mu) = -\infty$. In the following, we use $\boldsymbol{x}^T$ to denote the transpose of $\boldsymbol{x}$.

(i). We prove the convexity by induction. By equation (2.15) and (2.16), we have $L^a_{J-l+1}(\boldsymbol{\pi_{J-l+1}})$ and $L^s_{J-l+1}(\boldsymbol{\pi_{J-l+1}})$ are linear and therefore convex in $\boldsymbol{\pi_{J-l+1}}$. As the maximum of convex functions is convex, we have $V_{J-l+1}(\boldsymbol{\pi_{J-l+1}})$ is convex.

For $j < J - l$, assume $L^a_{j+1}(\boldsymbol{\pi_{j+1}})$, $L^s_{j+1}(\boldsymbol{\pi_{j+1}})$ and $V_{j+1}(\boldsymbol{\pi_{j+1}})$ are convex. By equation (2.8) and (2.15), we have

$$L^a_j(\boldsymbol{\pi_j}) = \sup_k \{G_j(\boldsymbol{\pi_j}, \delta_k)\} \tag{A.5}$$

For each k, we have $G_j(\boldsymbol{\pi_j}, \delta_k)$ is linear in $\boldsymbol{\pi_j}$. As the supremum of convex functions is convex and a positive linear combination of convex functions is convex, we have $L^a_j(\boldsymbol{\pi_j})$ is convex in $\boldsymbol{\pi_j}$.

From the induction hypothesis, we have $V_{j+1}(\boldsymbol{\pi_{j+1}})$ is convex, then we can write $V_{j+1}(\boldsymbol{\pi_{j+1}}) = \sup_{k\in K_{j+1}}\{\boldsymbol{a}_k\boldsymbol{\pi}^T_{\boldsymbol{j+1}} + b_k\}$, where K_{j+1} represents an index set, $\boldsymbol{a}_k$ is a constant vector of dimensions $1 \times I$, and b_k is a constant. Then define a $1 \times I$ vector

$\boldsymbol{e} \triangleq (1, ..., 1)$ and a $I \times I$ diagonal matrix $P_j(d_j) \triangleq diag(f_j(d_j|\theta_1), ..., f_j(d_j|\theta_I))$, and following equation (2.16), we have

$$\begin{aligned}
L_j^s(\boldsymbol{\pi}_j) &= h_{j+l}(\boldsymbol{\pi}_j, \mu_0) + E\left[V_{j+1}(\boldsymbol{\Pi_{j+1}})|\boldsymbol{\pi_j}\right] \\
&= h_{j+l}(\boldsymbol{\pi}_j, \mu_0) + E\left[\sup_{k \in K_{j+1}} \left\{\boldsymbol{a}_k \boldsymbol{\Pi}_{j+1}^T + b_k\right\}\middle| \boldsymbol{\pi_j}\right] \\
&= h_{j+l}(\boldsymbol{\pi}_j, \mu_0) + \sum_{d_j=0}^{\infty}\left[\sup_{k \in K_{j+1}} \left\{\boldsymbol{a}_k \frac{P_j(d_j)\boldsymbol{\pi_j}^T}{\boldsymbol{e}P_j(d_j)\boldsymbol{\pi_j}^T} + b_k\right\}\right] \boldsymbol{e}P_j(d_j)\boldsymbol{\pi_j}^T \\
&= h_{j+l}(\boldsymbol{\pi}_j, \mu_0) + \sum_{d_j=0}^{\infty}\left[\sup_{k \in K_{j+1}} \left\{\tilde{\boldsymbol{a}}_k(d_j)\boldsymbol{\pi_j}^T\right\}\right] \qquad \text{(A.6)}
\end{aligned}$$

where $\tilde{\boldsymbol{a}}_k(d_j) \triangleq \boldsymbol{a}_k P_j(d_j) + b_k \boldsymbol{e} P_j(d_j)$.

Once again, as the supremum of convex functions is convex and a positive linear combination of convex functions is convex, we have $L_j^s(\boldsymbol{\pi_j})$ is convex in $\boldsymbol{\pi_j}$. It follows that $V_j(\boldsymbol{\pi_j})$ is convex in $\boldsymbol{\pi_j}$.

(ii). We show that $\mathbb{P}_j$ is a convex partition of $\mathcal{P}$ by verifying the four conditions in Definition II.2.

- Condition (i): By the construction of $\mathbb{P}_j$, we have $\emptyset \notin \mathbb{P}_j$.

- Condition (ii): Let $\bigcup_k \mathcal{P}_{jk}$ denote the union of all sets in $\mathbb{P}_j$. For any $\boldsymbol{\pi} \in \bigcup_k \mathcal{P}_{jk}$, it is trivial that $\boldsymbol{\pi} \in \mathcal{P}$. Therefore, we have $\bigcup_k \mathcal{P}_{jk} \subseteq \mathcal{P}$. For any $\boldsymbol{\pi} \in \mathcal{P}$, we have $\Delta G_{j,k}(\boldsymbol{\pi})$ decreases in k. As we have $|G_j(\boldsymbol{\pi}, 0)| < \infty$ and $\lim_{\mu\to\infty} G_j(\boldsymbol{\pi}, \mu) = -\infty$, there exists a δ_k such that $\delta_k = \arg\max_{\mu \in \mathcal{K}} G_j(\boldsymbol{\pi}, \mu)$. Therefore, we have $\pi \in \bigcup_k \mathcal{P}_{jk}$. It follows that $\mathcal{P} \subseteq \bigcup_k \mathcal{P}_{jk}$. Then we have proved that $\bigcup_k \mathcal{P}_{jk} = \mathcal{P}$.

- Condition (iii): Assume there exist $k_1 < k_2$ such that $\mathcal{P}_{jk_1} \in \mathbb{P}_j$, $\mathcal{P}_{jk_2} \in \mathbb{P}_j$, and $\mathcal{P}_{jk_1} \bigcap \mathcal{P}_{jk_2} \neq \emptyset$. Then for $\boldsymbol{\pi} \in \mathcal{P}_{jk_1} \bigcap \mathcal{P}_{jk_2}$, we have $\hat{\mu}_j^a(\boldsymbol{\pi}) = \delta_{k_1}$ and $\hat{\mu}_j^a(\boldsymbol{\pi}) = \delta_{k_2}$. However, this contradicts the fact that $\hat{\mu}_j^a(\boldsymbol{\pi})$ is uniquely-defined.

- Condition (iv): Let $\boldsymbol{\pi_1} \in \mathcal{P}_{j\hat{k}}$ and $\boldsymbol{\pi_2} \in \mathcal{P}_{j\hat{k}}$. We have $\hat{\mu}_j^a(\boldsymbol{\pi_1}) = \delta_{\hat{k}}$ and $\hat{\mu}_j^a(\boldsymbol{\pi_2}) = \delta_{\hat{k}}$, and

$$L_j^a(\boldsymbol{\pi_i}) = G_j(\boldsymbol{\pi_i}, \delta_{\hat{k}}),\ i = 1, 2 \tag{A.7}$$

We observe that $L_j^a(\boldsymbol{\pi_i})$ is linear in $\boldsymbol{\pi_i}$. From part (i) we have $L_j^a(\boldsymbol{\pi})$ is convex in $\boldsymbol{\pi}$. Therefore, for any $\alpha \in (0, 1)$, we have

$$L_j^a(\alpha\boldsymbol{\pi_1} + (1-\alpha)\boldsymbol{\pi_2}) \le \alpha L_j^a(\boldsymbol{\pi_1}) + (1-\alpha)L_j^a(\boldsymbol{\pi_2}) = G_j(\alpha\boldsymbol{\pi_1} + (1-\alpha)\boldsymbol{\pi_2}, \delta_{\hat{k}}) \tag{A.8}$$

By the definition of $L_j^a(\alpha\boldsymbol{\pi_1} + (1-\alpha)\boldsymbol{\pi_2})$, we have

$$\begin{aligned} L_j^a(\alpha\boldsymbol{\pi_1} + (1-\alpha)\boldsymbol{\pi_2}) &= \sup_k \{G_j(\alpha\boldsymbol{\pi_1} + (1-\alpha)\boldsymbol{\pi_2}, \delta_k)\} \\ &\ge G_j(\alpha\boldsymbol{\pi_1} + (1-\alpha)\boldsymbol{\pi_2}, \delta_{\hat{k}}) \end{aligned} \tag{A.9}$$

By equation (A.8) and (A.9), we have

$$L_j^a(\alpha\boldsymbol{\pi_1} + (1-\alpha)\boldsymbol{\pi_2}) = G_j(\alpha\boldsymbol{\pi_1} + (1-\alpha)\boldsymbol{\pi_2}, \delta_{\hat{k}}) \tag{A.10}$$

which implies that $\hat{\mu}_j^a(\alpha\boldsymbol{\pi_1} + (1-\alpha)\boldsymbol{\pi_2}) = \delta_{\hat{k}}$.

(iii). Consider $\mathcal{P}_{jk} \in \mathbb{P}_j$. For $\boldsymbol{\pi} \in \mathcal{P}_{jk}$, we have $L_j^a(\boldsymbol{\pi_j})$ is linear in $\boldsymbol{\pi_j}$, and $L_j^s(\boldsymbol{\pi_j})$ is convex in $\boldsymbol{\pi_j}$. Therefore, the difference $\Delta L_j(\boldsymbol{\pi_j}) \triangleq L_j^a(\boldsymbol{\pi_j}) - L_j^s(\boldsymbol{\pi_j})$ is concave in $\boldsymbol{\pi_j}$. Therefore, if $\Delta L_j(\boldsymbol{\pi_j}) \le 0$ for all $\boldsymbol{\pi} \in \mathcal{P}_{jk}$, we have $S_{jk} = \emptyset$. Otherwise, define $S_{jk} \triangleq \{\boldsymbol{\pi} : \boldsymbol{\pi} \in \mathcal{P}_{jk}, \Delta L_j(\boldsymbol{\pi_j}) > 0\}$. It follows that S_{jk} is a convex set and for all $\boldsymbol{\pi} \in S_{jk}$, it is optimal for the firm to stop observing the demand and adjust the capacity.

(iv). By equation (A.5), we have $L_j^a(\boldsymbol{\pi_j}) = \sup_k \{G_j(\boldsymbol{\pi_j}, \delta_k)\}$. As $\Delta G_{j,k}(\boldsymbol{\pi})$ (defined in equation (A.4)) is a decreasing sequence in k for given $\boldsymbol{\pi}$ and j, to prove the result, it is sufficient to show that for given k, if $\boldsymbol{\pi_{j_1}} \preceq \boldsymbol{\pi_{j_2}}$, then $\Delta G_{j,k}(\boldsymbol{\pi_{j_1}}) \leq \Delta G_{j,k}(\boldsymbol{\pi_{j_2}})$. We first prove the result for the case where for $\hat{i}_1 < \hat{i}_2$ and $\epsilon > 0$, we have $\pi_{j_2,\hat{i}_1} = \pi_{j_1,\hat{i}_1} - \epsilon$, $\pi_{j_2,\hat{i}_2} = \pi_{j_1,\hat{i}_2} + \epsilon$, and $\pi_{j_2,\hat{i}} = \pi_{j_1,\hat{i}}$ for all $\hat{i} \neq \hat{i}_1, \hat{i}_2$.

First, by equation (A.3) and (2.3), we have

$$\begin{aligned}
\Delta G_{j,k}(\boldsymbol{\pi_{j_2}}) - \Delta G_{j,k}(\boldsymbol{\pi_{j_1}}) &= \frac{\sum_{i=j+l}^{J} \left[h_i(\boldsymbol{\pi_{j_2}}, \delta_{k+1}) - h_i(\boldsymbol{\pi_{j_2}}, \delta_k)\right]}{\delta_{k+1} - \delta_k} \\
&\quad - \frac{\sum_{i=j+l}^{J} \left[h_i(\boldsymbol{\pi_{j_1}}, \delta_{k+1}) - h_i(\boldsymbol{\pi_{j_1}}, \delta_k)\right]}{\delta_{k+1} - \delta_k} \\
&= \frac{\epsilon}{\delta_{k+1} - \delta_k} \sum_{i=j+l}^{J} \left\{ E\left[g(D_i)|\theta_{\hat{i}_2}\right] - E\left[g(D_i)|\theta_{\hat{i}_1}\right] \right\} \quad \text{(A.11)}
\end{aligned}$$

where

$$g(D_i) \triangleq -c_1(D_i - \delta_{k+1}\tau)^+ + c_1(D_i - \delta_k\tau)^+ - c_0(\delta_{k+1} - \delta_k)\tau.$$

Because $D_i|\theta_{\hat{i}_1} \preceq_{st} D_i|\theta_{\hat{i}_2}$, and $g(D_i)$ increases in D_i, we have

$$\Delta G_{j,k}(\boldsymbol{\pi_{j_2}}) - \Delta G_{j,k}(\boldsymbol{\pi_{j_1}}) \geq 0. \tag{A.12}$$

For an arbitrary pair of $\boldsymbol{\pi_{j_1}}$ and $\boldsymbol{\pi_{j_2}}$ such that $\boldsymbol{\pi_{j_1}} \preceq \boldsymbol{\pi_{j_2}}$, we observe that $\boldsymbol{\pi_{j_2}}$ can be obtained from $\boldsymbol{\pi_{j_1}}$ within finite steps using the operations above (subtract ϵ_j from an element with a lower index and add ϵ_j to an element with a higher index). $\square$

Proof of Proposition II.4. The proof of part (i) is similar to the proof for Proposition II.3(i). We only prove part (ii) here.

When the feasible set for the capacity adjustment is continuous, we have

$$L_j^a(\boldsymbol{\pi_j}) = \max_{\mu \in \mathbb{R}^+} \left\{G_j(\boldsymbol{\pi_j}, \mu)\right\}. \tag{A.13}$$

We first define

$$G_j^{j_1,j_2}(\mu) \triangleq G_j(\boldsymbol{\pi_{j_2}}, \mu) - G_j(\boldsymbol{\pi_{j_1}}, \mu). \tag{A.14}$$

Then to prove the result, it is sufficient to show that for given μ, if $\boldsymbol{\pi_{j_1}} \preceq \boldsymbol{\pi_{j_2}}$, we have $\frac{dG_j^{j_1,j_2}}{d\mu}(\mu) \geq 0$. Following a similar step as in the proof of Proposition II.3(iv), it is sufficient to prove for the following case: for $\hat{i}_1 < \hat{i}_2$ and $\epsilon > 0$, we have $\pi_{j_2,\hat{i}_1} = \pi_{j_1,\hat{i}_1} - \epsilon$, $\pi_{j_2,\hat{i}_2} = \pi_{j_1,\hat{i}_2} + \epsilon$, and $\pi_{j_2,\hat{i}} = \pi_{j_1,\hat{i}}$ for all $\hat{i} \neq \hat{i}_1, \hat{i}_2$.

Following equation (A.3) and (2.3), we have

$$\begin{aligned} \frac{dG_j^{j_1,j_2}}{d\mu}(\mu) &= \sum_{i=j+l}^{J} \left[\frac{\partial h_i}{\partial \mu}(\boldsymbol{\pi_{j_2}}, \mu) - \frac{\partial h_i}{\partial \mu}(\boldsymbol{\pi_{j_1}}, \mu) \right] \\ &= \epsilon c_1 \tau \sum_{i=j+l}^{J} \left[F_i(\mu\tau|\theta_{\hat{i}_1}) - F_i(\mu\tau|\theta_{\hat{i}_2}) \right] \geq 0 \end{aligned} \tag{A.15}$$

It follows that $\hat{\mu}_j^a(\boldsymbol{\pi_{j_1}}) \leq \hat{\mu}_j^a(\boldsymbol{\pi_{j_2}})$, which completes the proof. □

We state the following proposition from Gallego (1992) before proving Proposition II.6.

Proposition A.1 (Proposition 1 in Gallego (1992))**.** *Let $\mathcal{F}$ denote the class of cumulative distributions with finite mean μ and variance σ^2, and R be a finite constant.*

$$\max_{F \in \mathcal{F}} \int (x - R)^+ dF(x) = \frac{1}{2}(\sqrt{\Delta^2 + \sigma^2} - \Delta) \tag{A.16}$$

where $\Delta = R - \mu$.

Essentially, this is a one-sided deviation bound. Following a similar proof, we have

$$\max_{F \in \mathcal{F}} \int (R - x)^+ dF(x) = \frac{1}{2}(\sqrt{\Delta^2 + \sigma^2} - \Delta) \tag{A.17}$$

where $\Delta = \mu - R$.

Proof of Proposition II.6. We derive an upper bound of the regret as follows. We first observe that for any x, y, and z, we have

$$(x-y)^{+} \leq (x-z)^{+} + (z-y)^{+}. \tag{A.18}$$

We have $\hat{\lambda}_{i,\tau_n} = \frac{N(n\lambda_i\tau_n)}{n\tau_n}$ from equation (2.22). To simplify the notations, we similarly define

$$\hat{\lambda}_{i,j\tau_n} \triangleq \frac{D_j|\theta_{i,n}}{n\tau_n} = \frac{N(n\lambda_i j\tau_n) - N(n\lambda_i(j-1)\tau_n)}{n\tau_n} \text{ for } j = 2, 3, ..., J_n \tag{A.19}$$

We observe that $\hat{\lambda}_{i,j\tau_n}$ for $j = 1, 2, ..., J_n$ is a sequence of i.i.d. random variables with $E\left(\hat{\lambda}_{i,j\tau_n}\right) = \lambda_i$ and $\mathrm{Var}\left(\hat{\lambda}_{i,j\tau_n}\right) = \frac{\sigma^2\lambda_i}{n\tau_n}$. From equation (2.23), we have

$$V_{0,n}^{ts}(\boldsymbol{\pi_1}) = \sum_{i=1}^{I} \pi_{1,i} E\left\{ \begin{array}{l} pn\hat{\lambda}_{i,(l_n+1)\tau_n}\tau_n - c_1 n\left(\hat{\lambda}_{i,(l_n+1)\tau_n} - \mu_0\right)^{+}\tau_n - c_0 n\mu_0\tau_n \\ -\hat{C}(n\mu_0, n\hat{\lambda}_{i,\tau_n}) + \sum\limits_{j=l_n+2}^{J_n}\Big[pn\hat{\lambda}_{i,j\tau_n}\tau_n \\ -c_1 n\left(\hat{\lambda}_{i,j\tau_n} - \hat{\lambda}_{i,\tau_n}\right)^{+}\tau_n - c_0 n\hat{\lambda}_{i,\tau_n}\tau_n\Big] \end{array} \middle| \theta_{i,n} \right\} \tag{A.20}$$

The expected operating profit from the period l_n+1, $pn\lambda_i\tau_n - c_1 nE(\hat{\lambda}_{i,(l_n+1)\tau_n} - \mu_0)^{+}\tau_n - c_0 n\mu_0\tau_n$, is positive, as $p \geq c_1 > c_0$. Therefore, we have

$$\text{RHS of (A.20)} \geq \sum_{i=1}^{I} \pi_{1,i} \left\{ \begin{array}{l} (p-c_0)n\lambda_i(T-\tau_n-l_t) - E\left[\hat{C}(n\mu_0, n\hat{\lambda}_{i,\tau_n})\right] \\ -c_1 n\tau_n \sum_{j=l_n+2}^{J_n} E\left(\hat{\lambda}_{i,j\tau_n} - \hat{\lambda}_{i,\tau_n}\right)^{+} \end{array} \right\} \tag{A.21}$$

By equation (2.25), we have the deterministic upper bound

$$V_{0,n}^{d} = \sum_{i=1}^{I} \pi_{1,i}\Big\{(p-c_0)n\lambda_i(T-l_t) - \hat{C}(n\mu_0, n\lambda_i)\Big\}.$$

Therefore, we have the regret

$$\begin{aligned} R_n^{ts} &= 1 - V_{0,n}^{ts}/V_{0,n}^{d} \\ &\leq \frac{1}{V_{0,n}^{d}} \sum_{i=1}^{I} \pi_{1,i} \left\{ \begin{array}{l} (p-c_0)n\lambda_i\tau_n - \hat{C}(n\mu_0, n\lambda_i) + E\left[\hat{C}(n\mu_0, n\hat{\lambda}_{i,\tau_n})\right] \\ +c_1 n\tau_n \sum_{j=l_n+2}^{J_n} E\left(\hat{\lambda}_{i,j\tau_n} - \hat{\lambda}_{i,\tau_n}\right)^+ \end{array} \right\}. \end{aligned} \tag{A.22}$$

Recall that for the initial capacity position μ and target capacity position μ', we have $\hat{C}(\mu,\mu') = c_a(\mu'-\mu)^+ + \gamma_a(\mu-\mu')^+$. We notice when $\gamma_a \geq 0$, we can directly apply (A.18) and $E[-\gamma_a(n\mu_0 - n\lambda_i)^+ + \gamma_a(n\mu_0 - n\hat{\lambda}_{i,\tau_n})^+] \leq E[\gamma_a n(\lambda_i - \hat{\lambda}_{i,\tau_n})^+]$. When $\gamma_a < 0$, we have $E[-\gamma_a(n\mu_0 - n\lambda_i)^+ + \gamma_a(n\mu_0 - n\hat{\lambda}_{i,\tau_n})^+] \leq 0$ by Jensen's inequality. Therefore, applying (A.18), we have

$$\text{RHS of (A.22)} \leq \frac{1}{V_{0,n}^{d}} \sum_{i=1}^{I} \pi_{1,i} \left\{ \begin{array}{l} (p-c_0)n\lambda_i\tau_n + c_a nE(\hat{\lambda}_{i,\tau_n} - \lambda_i)^+ \\ +\gamma_a^+ nE(\lambda_i - \hat{\lambda}_{i,\tau_n})^+ \\ +c_1 n\tau_n \sum\limits_{j=l_n+2}^{J_n} \left[E\left(\hat{\lambda}_{i,j\tau_n} - \lambda_i\right)^+ + E\left(\lambda_i - \hat{\lambda}_{i,\tau_n}\right)^+\right] \end{array} \right\} \tag{A.23}$$

From equation (A.23), it is clear that to derive an upper bound of R_n^{ts}, we need to find upper bounds for $E\left(\hat{\lambda}_{i,j\tau_n} - \lambda_i\right)^+$ and $E\left(\lambda_i - \hat{\lambda}_{i,\tau_n}\right)^+$ respectively. Recall that $E\left(\hat{\lambda}_{i,j\tau_n}\right) = \lambda_i$ and $\text{Var}\left(\hat{\lambda}_{i,j\tau_n}\right) = \frac{\sigma^2\lambda_i}{n\tau_n}$. In the following, we use C_i to represent a constant for all i, which is independent of n and τ_n.

We first find an upper bound for $E\left(\hat{\lambda}_{i,j\tau_n} - \lambda_i\right)^+$. By equation (A.16), we have

$$E\left(\hat{\lambda}_{i,j\tau_n} - \lambda_i\right)^+ \leq \frac{\sigma\sqrt{\lambda_i}}{2\sqrt{n\tau_n}} \text{ for } j = 1, 2, ..., J_n. \tag{A.24}$$

For $E\left(\lambda_i - \hat{\lambda}_{i,\tau_n}\right)^+$, by equation (A.17), we have the following

$$E\left(\lambda_i - \hat{\lambda}_{i,\tau_n}\right)^+ \leq \frac{\sigma\sqrt{\lambda_i}}{2\sqrt{n\tau_n}}. \tag{A.25}$$

From equation (A.24) and (A.25), we have

$$\text{RHS of (A.23)} \leq C_1\tau_n + \frac{C_2}{\sqrt{n\tau_n}} \tag{A.26}$$

Then the result follows by setting that $\tau_n \asymp n^{-\frac{1}{3}}$. $\square$

Proof of Proposition II.7. The proof follows in two steps: (1) show $V_j^m(\boldsymbol{\pi_j}, \hat{\mu}_{j-1})$ is concave in the capacity position $\hat{\mu}_{j-1}$ for all $j \leq J-l+1$; (2) show the optimal policy follows the control band structure and find the lower and upper thresholds.

We first show the concavity by induction.

For $j = J - l + 1$, we have $V_{J-l+1}^m(\boldsymbol{\pi_{J-l+1}}, \hat{\mu}_{J-l}) = 0$, and therefore is concave in $\hat{\mu}_{J-l}$. For $j + 1$, assume $V_{j+1}^m(\boldsymbol{\pi_{j+1}}, \hat{\mu}_j)$ is concave in $\hat{\mu}_j$. It follows that $E\left[V_{j+1}^m(\boldsymbol{\Pi_{j+1}}, \hat{\mu})|\boldsymbol{\pi_j}\right]$ is concave in $\hat{\mu}$ as the positive combination of concave functions is concave. Therefore, we have $h_{j+l}\left(\boldsymbol{\pi_j}, \hat{\mu}\right) - c_a\left(\hat{\mu} - \hat{\mu}_{j-1}\right)^+ - \gamma_a\left(\hat{\mu}_{j-1} - \hat{\mu}\right)^+ + E\left[V_{j+1}^m(\boldsymbol{\Pi_{j+1}}, \hat{\mu})|\boldsymbol{\pi_j}\right]$ is jointly concave in $(\mu, \hat{\mu})$. For a jointly-concave function $f(x, y)$ and a convex set $\mathcal{Y}$, we have $g(x) = \max_{y\in\mathcal{Y}} f(x, y)$ is concave in x. Then it follows that

$$\begin{aligned} V_j^m(\boldsymbol{\pi_j}, \hat{\mu}_{j-1}) = \max_{\hat{\mu}\in\mathbb{R}^+} \Big\{ & h_{j+l}\left(\boldsymbol{\pi_j}, \hat{\mu}\right) - c_a\left(\hat{\mu} - \hat{\mu}_{j-1}\right)^+ - \gamma_a\left(\hat{\mu}_{j-1} - \hat{\mu}\right)^+ \\ & + E\left[V_{j+1}^m(\boldsymbol{\Pi_{j+1}}, \hat{\mu})|\boldsymbol{\pi_j}\right] \Big\} \end{aligned}$$

is concave in $\hat{\mu}_{j-1}$.

We next show the optimal policy follows the control band structure. For $\hat{\mu}_{j-1} = 0$, we define

$$\underline{\mu}_j(\boldsymbol{\pi_j}) \triangleq \arg\max_{\hat{\mu}\in\mathbb{R}^+} \left\{ h_{j+l}\left(\boldsymbol{\pi_j}, \hat{\mu}\right) - c_a\hat{\mu} + E\left[V_{j+1}^m(\boldsymbol{\Pi_{j+1}}, \hat{\mu})|\boldsymbol{\pi_j}\right] \right\} \tag{A.27}$$

For any $\hat{\mu}_{j-1} < \underline{\mu}_j(\boldsymbol{\pi_j})$, it is optimal to adjust the capacity up to the level $\underline{\mu}_j(\boldsymbol{\pi_j})$.

For an arbitrary large $\hat{\mu}_{j-1}$, we define

$$\overline{\mu}_j(\boldsymbol{\pi_j}) \triangleq \arg\max_{\hat{\mu}\in\mathbb{R}^+} \left\{h_{j+l}\left(\boldsymbol{\pi_j},\hat{\mu}\right) + \gamma_a\hat{\mu} + E\left[V^m_{j+1}(\boldsymbol{\Pi_{j+1}},\hat{\mu})|\boldsymbol{\pi_j}\right]\right\} \tag{A.28}$$

It is optimal for the firm to disinvest its capacity to $\overline{\mu}_j(\boldsymbol{\pi_j})$ for all $\hat{\mu}_{j-1} > \overline{\mu}_j(\boldsymbol{\pi_j})$. As $c_a \geq 0$ and $c_a \geq -\gamma_a$, it follows that $\overline{\mu}_j(\boldsymbol{\pi_j}) \geq \underline{\mu}_j(\boldsymbol{\pi_j})$. Following the concavity of the value-to-go function, it is optimal for the firm to stay put when $\underline{\mu}_j(\boldsymbol{\pi_j}) \leq \mu \leq \overline{\mu}_j(\boldsymbol{\pi_j})$. Therefore, we have proved the optimal policy is a control band policy. □

Proof of Lemma II.8. When there are only two demand types, we can reduce the information state to $\pi_j = \pi_{j,2}$ as $\pi_{j,1} = 1-\pi_{j,2}$, and therefore $\boldsymbol{\pi_j} = (1-\pi_j, \pi_j)$. We define the two functions

$$G^a_j(\pi_j,\hat{\mu}_{j-1},\hat{\mu}) = h_{j+l}\left(\boldsymbol{\pi_j},\hat{\mu}\right) - c_a\left(\hat{\mu}-\hat{\mu}_{j-1}\right)^+ - \gamma_a\left(\hat{\mu}_{j-1}-\hat{\mu}\right)^+ + E\left[V^m_{j+1}(\Pi_{j+1},\hat{\mu})|\pi_j\right]$$
$$G^s_j(\pi_j,\hat{\mu}) = h_{j+l}\left(\boldsymbol{\pi_j},\hat{\mu}\right) + E\left[V^m_{j+1}(\Pi_{j+1},\hat{\mu})|\pi_j\right]. \tag{A.29}$$

We observe that $V^m_j(\pi_j,\hat{\mu}_{j-1}) = \max_{\hat{\mu}} G^a_j(\pi_j,\hat{\mu}_{j-1},\hat{\mu})$. In addition, from Proposition II.7, we have $G^a_j(\pi_j,\hat{\mu}_{j-1},\hat{\mu})$ is concave in $\hat{\mu}$. Therefore, to show the two thresholds increase in the information state π_j, it is sufficient to show that for $j = 1,...,J-l$, we have $\frac{\partial G^a_j(\pi_j,\hat{\mu}_{j-1},\hat{\mu})}{\partial\hat{\mu}}$ increases in π_j for any $\hat{\mu} \neq \hat{\mu}_{j-1}$, and $\frac{\partial G^s_j(\pi_j,\hat{\mu})}{\partial\hat{\mu}}$, which is a special case when $\hat{\mu} = \hat{\mu}_{j-1}$, increases in π_j. We prove this by induction. We present the result for $G^a_j(\cdot)$ as the proof for $G^s_j(\cdot)$ is identical.

To establish induction basis, let $j = J-l$. If $\hat{\mu} \leq \hat{\mu}_{J-l-1}$, we have

$$\frac{\partial G^a_{J-l}(\pi_{J-l},\hat{\mu}_{J-l-1},\hat{\mu})}{\partial\hat{\mu}} = c_1\tau[1-F_J(\hat{\mu}\tau|\theta_1)] - c_0\tau + \pi_{J-l}c_1\tau[F_J(\hat{\mu}\tau|\theta_1) - F_J(\hat{\mu}\tau|\theta_2)] + \gamma_a \tag{A.30}$$

Therefore, as $D_j|\theta_1 \preceq_{st} D_j|\theta_2$, we have $\frac{\partial G^a_{J-l}(\pi_{J-l},\hat{\mu}_{J-l-1},\hat{\mu})}{\partial\hat{\mu}}$ increases in π_{J-l} when $\hat{\mu} < \hat{\mu}_{J-l-1}$. A similar argument establishes the result for the case $\hat{\mu} > \hat{\mu}_{J-l-1}$.

Suppose that the result hold for all $t = j+1,...,J-l$. Thus, at period $j+1$,

$\frac{\partial G^a_{j+1}(\pi_{j+1},\hat{\mu}_j,\hat{\mu})}{\partial \hat{\mu}}$ increases in π_{j+1} for all $\hat{\mu} \neq \hat{\mu}_j$. From the induction hypothesis, the two switching curves $\underline{\mu}_{j+1}(\pi_{j+1})$ and $\bar{\mu}_{j+1}(\pi_{j+1})$ increase in π_{j+1} as well. We now show that $\frac{\partial G^a_j(\pi_j,\hat{\mu}_{j-1},\hat{\mu})}{\partial \hat{\mu}}$ increases in π_j for $\hat{\mu} \neq \hat{\mu}_{j-1}$.

For any $\hat{\mu} < \hat{\mu}_{j-1}$, we have

$$\begin{aligned}\frac{\partial G^a_j(\pi_j,\hat{\mu}_{j-1},\hat{\mu})}{\partial \hat{\mu}} =& c_1\tau[1 - F_{j+l}(\hat{\mu}\tau|\theta_1)] - c_0\tau + \pi_j c_1\tau[F_{j+l}(\hat{\mu}\tau|\theta_1) - F_{j+l}(\hat{\mu}\tau|\theta_2)] + \gamma_a \\ &+ E\left[\frac{\partial V^m_{j+1}}{\partial \hat{\mu}}(\Pi_{j+1},\hat{\mu})\middle|\, \pi_j\right] \end{aligned} \tag{A.31}$$

The expression for $\frac{\partial G^a_j}{\partial \hat{\mu}}(\pi_j,\hat{\mu}_{j-1},\hat{\mu})$ when $\hat{\mu} > \hat{\mu}_{j-1}$ is similar. In order to show equation (A.31) increases in π_j, we need to show that $E\left[\frac{\partial V^m_{j+1}}{\partial \hat{\mu}}(\Pi_{j+1},\hat{\mu})\middle|\, \pi_j\right]$ increases in π_j, which is shown in two steps below.

We first observe that

$$\frac{\partial V^m_{j+1}}{\partial \hat{\mu}_j}(\pi_{j+1},\hat{\mu}_j) = \begin{cases} c_a & \text{if } \hat{\mu}_j < \underline{\mu}_{j+1}(\pi_{j+1}) \\ \frac{\partial G^s_{j+1}}{\partial \hat{\mu}}(\pi_{j+1},\hat{\mu}_j) & \text{if } \underline{\mu}_{j+1}(\pi_{j+1}) \le \hat{\mu}_j \le \bar{\mu}_{j+1}(\pi_{j+1}) \\ -\gamma_a & \text{if } \hat{\mu}_j > \bar{\mu}_{j+1}(\pi_{j+1}) \end{cases} \tag{A.32}$$

Notice that $\frac{\partial V^m_{j+1}}{\partial \hat{\mu}_j}(\pi_{j+1},\hat{\mu}_j)$ is continuously decreasing in $\hat{\mu}_j$, and $\underline{\mu}_{j+1}(\pi_{j+1})$ and $\bar{\mu}_{j+1}(\pi_{j+1})$ increase in π_{j+1}. From the induction hypothesis, $\frac{\partial V^m_{j+1}}{\partial \hat{\mu}_j}(\pi_{j+1},\hat{\mu}_j)$ increases in π_{j+1}.

Next, we show that $\Pi_{j+1}|\pi_j$ increases in π_j in the first order stochastic dominance sense. Notice that by equation (2.1), given information state π_j, for realized demand d_j, we have

$$\pi_{j+1} = \frac{\pi_j f_j(d_j|\theta_2)}{\pi_j f_j(d_j|\theta_2) + (1-\pi_j) f_j(d_j|\theta_1)}.$$

Then following any sample path of the random demand, π_{j+1} increases in π_j. We notice that the demand density functions bear the monotone likelihood ratio property, so we have π_{j+1} increases in d_j. Therefore, we have $\Pi_{j+1}|\pi_j$ increases in π_j in the first order stochastic dominance sense. Thus, $E\left[\frac{\partial V^m_{j+1}}{\partial \hat{\mu}}(\Pi_{j+1},\hat{\mu})\middle|\, \pi_j\right]$ must increase in π_j.

It then follows that $\frac{\partial G_j^a}{\partial \hat{\mu}}(\pi_j, \hat{\mu}_{j-1}, \hat{\mu})$ increases in π_j for $\hat{\mu} < \hat{\mu}_{j-1}$. A similar argument proves the case for $\hat{\mu} > \hat{\mu}_{j-1}$. Therefore, the result holds for period j. $\square$

Proof of Proposition II.9. Similar to the proof of Proposition II.6, we first find an upper bound of the regret by finding a lower bound of $V_{0,n}^{ms}$. To simplify the notations, we define the firm's expected profits under the multi-step heuristic (given the demand type λ_i) in different periods as follows. We still use $\hat{\lambda}_{i,j\tau_n}$ to denote $\frac{D_j|\theta_{i,n}}{n\tau_n}$. First, in period $l_n + 1$, the firm's capacity is still the initial capacity μ_0, and we have

$$W_{s,n}(\lambda_i) \triangleq E\Big\{pn\hat{\lambda}_{i,(l_n+1)\tau_n}\tau_n - c_1 n\left(\hat{\lambda}_{i,(l_n+1)\tau_n} - \mu_0\right)^+ \tau_n - c_0 n\mu_0\tau_n\Big\} \tag{A.33}$$

Second, during period $l_n + 2$ and $l_n + 2^{K_n} - 1$, the firm's capacity level is updated according to the heuristic. Observing that $E[\bar{\lambda}_{i,\kappa}] = \lambda_i$, we have the firm's expected profits as

$$\begin{aligned} W_{m,n}(\lambda_i) \triangleq & E\sum_{\kappa=1}^{K_n-1}\left\{\sum_{j=l_n+2^\kappa}^{l_n+2^{\kappa+1}-1}\left[pn\hat{\lambda}_{i,j\tau_n}\tau_n - c_1 n\left(\hat{\lambda}_{i,j\tau_n} - \bar{\lambda}_{i,\kappa}\right)^+\tau_n - c_0 n\bar{\lambda}_{i,\kappa}\tau_n\right]\right. \\ & \left. - \hat{C}\left(n\bar{\lambda}_{i,\kappa-1}, n\bar{\lambda}_{i,\kappa}\right)\right\} \\ = & \sum_{\kappa=1}^{K_n-1}\left\{(p-c_0)n\lambda_i 2^\kappa\tau_n - c_1 n\tau_n\sum_{j=l_n+2^\kappa}^{l_n+2^{\kappa+1}-1}E\left(\hat{\lambda}_{i,j\tau_n} - \bar{\lambda}_{i,\kappa}\right)^+\right. \\ & \left. - E\left[\hat{C}\left(n\bar{\lambda}_{i,\kappa-1}, n\bar{\lambda}_{i,\kappa}\right)\right]\right\} \end{aligned} \tag{A.34}$$

Finally, during period $l_n + 2^{K_n}$ and J_n, the firm makes the last adjustment, and the capacity maintains at this level for the rest of the time horizon. Then we have

$$
\begin{aligned}
W_{l,n}(\lambda_i) \triangleq & E\left\{ \sum_{j=l_n+2^{K_n}}^{J_n} \left[pn\hat{\lambda}_{i,j\tau_n}\tau_n - c_1 n \left(\hat{\lambda}_{i,j\tau_n} - \bar{\lambda}_{i,K_n}\right)^+ \tau_n - c_0 n\bar{\lambda}_{i,K_n}\tau_n \right] \right. \\
& \left. \qquad - \hat{C}\left(n\bar{\lambda}_{i,K_n-1}, n\bar{\lambda}_{i,K_n}\right) \right\} \\
= & (p-c_0)n\lambda_i \left(T - l_t - \left(2^{K_n}-1\right)\tau_n\right) - c_1 n\tau_n \sum_{j=l_n+2^{K_n}}^{J_n} E\left(\hat{\lambda}_{i,j\tau_n} - \bar{\lambda}_{i,K_n}\right)^+ \\
& - E\left[\hat{C}\left(n\bar{\lambda}_{i,K_n-1}, n\bar{\lambda}_{i,K_n}\right)\right] \qquad \text{(A.35)}
\end{aligned}
$$

Because $W_{s,n}(\lambda_i) \geq 0$ as $p \geq c_1 > c_0$, we have

$$
V_{0,n}^{ms} = \sum_{i=1}^{I} \pi_{1,i}\Big\{W_{s,n}(\lambda_i) + W_{m,n}(\lambda_i) + W_{l,n}(\lambda_i)\Big\} \geq \sum_{i=1}^{I} \pi_{1,i}\Big\{W_{m,n}(\lambda_i) + W_{l,n}(\lambda_i)\Big\} \qquad \text{(A.36)}
$$

Therefore, by equation (2.32), we have an upper bound of the regret as follows

$$
\begin{aligned}
R_n^{\pi_n} &= 1 - V_{0,n}^{ms}/V_{0,n}^d \\
&\leq \frac{1}{V_{0,n}^d}\sum_{i=1}^{I}\pi_{1,i}\left\{(p-c_0)n\lambda_i(T-l_t) - \hat{C}(n\mu_0, n\lambda_i) - W_{m,n}(\lambda_i) - W_{l,n}(\lambda_i)\right\} \\
&= \frac{1}{V_{0,n}^d}\sum_{i=1}^{I}\pi_{1,i}\left\{
\begin{array}{l}
(p-c_0)n\lambda_i\tau_n - \hat{C}(n\mu_0, n\lambda_i) + \sum_{\kappa=1}^{K_n} E\left[\hat{C}\left(n\bar{\lambda}_{i,\kappa-1}, n\bar{\lambda}_{i,\kappa}\right)\right] \\
+c_1 n\tau_n \left[\sum_{\kappa=1}^{K_n-1}\sum_{j=l_n+2^{\kappa}}^{l_n+2^{\kappa+1}-1} E\left(\hat{\lambda}_{i,j\tau_n} - \bar{\lambda}_{i,\kappa}\right)^+ \right. \\
\left. + \sum_{j=l_n+2^{K_n}}^{J_n} E\left(\hat{\lambda}_{i,j\tau_n} - \bar{\lambda}_{i,K_n}\right)^+ \right]
\end{array}
\right\} \qquad \text{(A.37)}
\end{aligned}
$$

To find an upper bound for the right hand side of equation (A.37), we need to find an upper bound for $E\left(\hat{\lambda}_{i,j\tau_n} - \bar{\lambda}_{i,\kappa}\right)^+$, $E(\bar{\lambda}_{i,\kappa} - \bar{\lambda}_{i,\kappa-1})^+$, and $E(\bar{\lambda}_{i,\kappa-1} - \bar{\lambda}_{i,\kappa})^+$ respectively. Note that $E[\bar{\lambda}_{i,\kappa}] = \lambda_i$ and $Var[\bar{\lambda}_{i,\kappa}] = \frac{\sigma^2\lambda_i}{n(2^\kappa-1)\tau_n}$. We use C_i to represent a constant which is independent of n and τ_n for all i. By Proposition A.1 and the inequality of (A.18), we have

$$\begin{aligned} E\left(\hat{\lambda}_{i,j\tau_n} - \bar{\lambda}_{i,\kappa}\right)^+ &\le E\left(\hat{\lambda}_{i,j\tau_n} - \lambda_i\right)^+ + E\left(\lambda_i - \bar{\lambda}_{i,\kappa}\right)^+ \\ &\le \frac{\sigma\sqrt{\lambda_i}}{2\sqrt{n\tau_n}} + \frac{\sigma\sqrt{\lambda_i}}{2\sqrt{n(2^\kappa - 1)\tau_n}} \le \frac{C_3}{\sqrt{n\tau_n}} \end{aligned} \tag{A.38}$$

For $\kappa = 1$, we have

$$E\left(\bar{\lambda}_{i,1} - \bar{\lambda}_{i,0}\right)^+ = E\left(\bar{\lambda}_{i,1} - \mu_0\right)^+ \le E\left(\bar{\lambda}_{i,1} - \lambda_i\right)^+ + (\lambda_i - \mu_0)^+ \le \frac{\sigma\sqrt{\lambda_i}}{2\sqrt{n\tau_n}} + (\lambda_i - \mu_0)^+ \tag{A.39}$$

$$E\left(\bar{\lambda}_{i,0} - \bar{\lambda}_{i,1}\right)^+ = E\left(\mu_0 - \bar{\lambda}_{i,1}\right)^+ \le (\mu_0 - \lambda_i)^+ + E\left(\lambda_i - \bar{\lambda}_{i,1}\right)^+ \le (\mu_0 - \lambda_i)^+ + \frac{\sigma\sqrt{\lambda_i}}{2\sqrt{n\tau_n}} \tag{A.40}$$

By equation (2.30), for $\kappa \ge 2$, we have

$$\begin{aligned} &E\left(\bar{\lambda}_{i,\kappa} - \bar{\lambda}_{i,\kappa-1}\right)^+ \\ &= E\left(\frac{\bar{\lambda}_{i,\kappa-1} n(2^{\kappa-1} - 1)\tau_n + N(n\lambda_i(2^\kappa - 1)\tau_n) - N(n\lambda_i(2^{\kappa-1} - 1)\tau_n)}{n(2^\kappa - 1)\tau_n} - \bar{\lambda}_{i,\kappa-1}\right)^+ \\ &= E\left(\frac{N(n\lambda_i(2^\kappa - 1)\tau_n) - N(n\lambda_i(2^{\kappa-1} - 1)\tau_n) - \bar{\lambda}_{i,\kappa-1} n 2^{\kappa-1}\tau_n}{n(2^\kappa - 1)\tau_n}\right)^+ \\ &\le E\left(\frac{N(n\lambda_i(2^\kappa - 1)\tau_n) - N(n\lambda_i(2^{\kappa-1} - 1)\tau_n)}{n2^{\kappa-1}\tau_n} - \lambda_i\right)^+ + E\left(\lambda_i - \bar{\lambda}_{i,\kappa-1}\right)^+ \\ &\le \frac{\sigma\sqrt{\lambda_i}}{2\sqrt{n2^{\kappa-1}\tau_n}} + \frac{\sigma\sqrt{\lambda_i}}{2\sqrt{n(2^{\kappa-1} - 1)\tau_n}} \le \frac{C_5}{\sqrt{n2^\kappa\tau_n}} \end{aligned} \tag{A.41}$$

$$E\left(\bar{\lambda}_{i,\kappa-1} - \bar{\lambda}_{i,\kappa}\right)^+ \le \frac{C_6}{\sqrt{n2^\kappa\tau_n}} \tag{A.42}$$

We next apply the inequality (A.38) to (A.42) to the right hand side of equation (A.37), and gather the items by the outsourcing costs and capacity adjustment costs respectively. Then we obtain an upper bound of the regret as follows

$$
\begin{aligned}
\text{RHS of (A.37)} &\leq C_7\tau_n + C_8\frac{c_1}{\sqrt{n\tau_n}} + C_9\sum_{\kappa=1}^{K_n}\frac{\max(c_a, \gamma_a^+)}{\sqrt{n2^{\kappa}\tau_n}} \\
&\leq C_7\tau_n + \frac{C_{10}}{\sqrt{n\tau_n}} \qquad \text{(A.43)}
\end{aligned}
$$

The last inequality follows the fact that K_n satisfies that $(2^{K_n+1}-1)\tau_n \leq T - l_t$.

By setting $\tau_n \asymp n^{-\frac{1}{3}}$, we obtain an upper bound of the regret on the order of $n^{-\frac{1}{3}}$. $\square$

Remark: The exponentially increasing time between two consecutive adjustments is important in establishing the upper bound in the order of $n^{-1/3}$. To illustrate this, we alternatively consider another heuristic, where the time between two consecutive adjustments is fixed as $\eta\tau_n$, $\eta \in \mathbb{N}^+$. We denote the regret under this heuristic as R_n^{fa}. Following the same logic as in the proof of Proposition II.9, it can be obtained that the upper bound of the regret satisfies the following

$$
\begin{aligned}
R_n^{fa} &\leq C_7\tau_n + C_{11}\frac{c_1}{\sqrt{n\tau_n}} + C_{12}\frac{\max(c_a, \gamma_a^+)}{\sqrt{n\eta\tau_n}}\frac{T}{\eta\tau_n} \\
&\leq C_7\tau_n + C_{13}\frac{1}{\sqrt{n\tau_n^3}} \qquad \text{(A.44)}
\end{aligned}
$$

In this case, the firm should set $\tau_n \asymp n^{-1/5}$ and yield an upper bound in the order of $n^{-1/5}$. It cannot tighten the upper bound to the order of $n^{-1/3}$, because the capacity adjustment is too frequent and the adjustment cost is too high.

Profit and cost parameter estimations in numerical examples.

We use *Production* to indicate the total production volume of Ford in 2012, which is approximated by the wholesale volume of 5,668 thousands units (operating highlights, Ford Motor Company 2012). As estimated by IHS Automotive (P.12, Ford Motor Company 2012), the global automotive industry production capacity for light vehicles is about 108 million units, which exceeds the global production by 26 million units.

We therefore use the industry capacity utilization $Utilization = \frac{108-26}{108} = 75.93\%$ to estimated Ford's total capacity (including all types of products) in 2012 as

$$\begin{aligned} Capacity = \frac{Production}{Utilization} &= \frac{5,668 \times 10^3}{75.93\%} = 7,465 \times 10^3 units/year \\ &= 622.1 \times 10^3 units/month \end{aligned}$$

- Capacity adjustment cost c_a and γ_a. The capacity adjustment cost c_a is estimated from the *Amortization of special tools*(AST) (P.102, Ford Motor Company 2012). As Ford generally amortizes special tools over the expected life of a product program using a straightline method, we calculate the expected cost to install one unit of capacity c_a as

$$\begin{aligned} c_a = \frac{AST \times \frac{T}{2}}{Capacity} &= \frac{1,861 \times 10^6 \times \frac{3}{2}}{622.1 \times 10^3} \\ &= 4,487 \ dollars \cdot month/units. \end{aligned}$$

  As the capacity adjustment is often irreversible, we use a coefficient γ to measure the irreversibility and assume $\gamma_a = \gamma c_a$. In the base analysis, we assume $\gamma = 0.1$, i.e., it is costly for the firm to downsize its capacity.

- Capacity overhead cost c_0. The overhead cost is estimated from *Maintenance and rearrangement* expense (MR) (P.102, Ford Motor Company 2012). This cost reflects the firm's expense to conduct routine maintenance and repair to keep up its capacity level, and is incurred regardless of the production location. Therefore, we calculate c_0 as

$$c_0 = \frac{MR}{Capacity} = \frac{1,352 \times 10^6}{7,465 \times 10^3} = 181.1 \ dollars/units.$$

- Capacity outsourcing cost c_1. The outsourcing cost is incurred when the demand exceeds the installed capacity and therefore has to be satisfied by another facility. Therefore, the capacity outsourcing cost includes the cost to maintain the extra unit of outsourcing capacity and additional machine setup and transportation costs, and we denote the cost $c_1 = (1+\beta)c_0$ with $\beta > 0$. In the base case, we assume $\beta = 1$.

- Unit profit p. The unit profit is the profit the firm earns from selling a car, excluding the capacity related cost. We denote the gross revenue by *Revenue* and the total operating cost by *Cost*. Then we estimate the unit profit as

$$\begin{aligned} p &= \frac{Revenue - Cost + MR + AST}{Production} \\ &= \frac{125,567 - 121,584 + 1,352 + 1,861}{5,668} \times 10^3 \\ &= 1,270 \; dollars/units. \end{aligned}$$

  We observe that from Ford Focus's official website[1], a simple average of the starting manufacturer suggested retail price (MSRP) for the seven current focus models yields a value of $(16,200 + 18,200 + 19,200 + 23,200 + 23,799 + 24,200 + 39,200)/7 = \$23,414$. We observe that this value is close to the average retail price estimated from the financial data, $Revenue/Production = 125,567/5,668 \times 10^3 = \$22,154$.

In the numerical analysis, we perform robustness checks with respect to these estimated parameters.

[1]Ford Focus: http://www.ford.com/cars/focus/

APPENDIX B

Proofs and Technical Details for Chapter III: Investing in a Shared Supplier in a Competitive Market: The Stochastic Capacity Case

Proof of Lemma III.1. Without loss of generality, we assume $k_1 \geq k_2$. Using concavity of the objective function, we obtain the best response function of firm 1 and firm 2 as:

$$q_1^*(q_2) = \begin{cases} \frac{a-c-bq_2}{2b} & \text{if } \frac{a-c-bq_2}{2b} < (\frac{k_0}{2} + k_1)\xi, \\ (\frac{k_0}{2} + k_1)\xi & \text{if } \frac{a-c-bq_2}{2b} \geq (\frac{k_0}{2} + k_1)\xi. \end{cases}$$

$$q_2^*(q_1) = \begin{cases} \frac{a-c-bq_1}{2b} & \text{if } \frac{a-c-bq_1}{2b} < (\frac{k_0}{2} + k_2)\xi, \\ (\frac{k_0}{2} + k_2)\xi & \text{if } \frac{a-c-bq_1}{2b} \geq (\frac{k_0}{2} + k_2)\xi. \end{cases}$$

Solving for the intersection of the best response functions $q_1^*(q_2)$ and $q_2^*(q_1)$, we can obtain the equilibrium order quantities and hence the equilibrium profits shown in Lemma III.1. □

Discussion about when $\frac{a-c}{3b(\frac{k_0}{2}+k_2)} > 1$. We observe that when $\frac{a-c}{3b(\frac{k_0}{2}+k_2)} > 1$, there are two cases.

Case 1. When $\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)} \leq 1$, we have the equilibrium order quantity and ex post profit as follows:

realized yield ξ	order quantity (q_1^*, q_2^*)	ex post profit $(\pi_1^e,\ \pi_2^e)$
$0 \leq \xi \leq \frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}$	$\big((\frac{k_0}{2}+k_1)\xi, (\frac{k_0}{2}+k_2)\xi\big)$	$\big(m_2(\frac{k_0}{2}+k_1)\xi, m_2(\frac{k_0}{2}+k_2)\xi\big)$
$\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)} < \xi \leq 1$	$\left(\frac{a-c-b(\frac{k_0}{2}+k_2)\xi}{2b}, (\frac{k_0}{2}+k_2)\xi\right)$	$\left(\frac{m_1[a-c-b(\frac{k_0}{2}+k_2)\xi]}{2b}, m_1(\frac{k_0}{2}+k_2)\xi\right)$

Case 2. When $\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)} > 1$, we have the equilibrium order quantity and ex post profit as follows:

realized yield ξ	order quantity (q_1^*, q_2^*)	ex post profit $(\pi_1^e,\ \pi_2^e)$
$0 \leq \xi \leq 1$	$\big((\frac{k_0}{2}+k_1)\xi, (\frac{k_0}{2}+k_2)\xi\big)$	$\big(m_2(\frac{k_0}{2}+k_1)\xi, m_2(\frac{k_0}{2}+k_2)\xi\big)$

With the ex post profit derived, the firm's expected profit can be obtained similarly as in equation (3.1) and (3.2), and the analysis follows. The discussions for the first-priority case (Lemma III.3) are similar and omitted for space.

Proof of Proposition III.2. We prove the proposition in two steps. We first illustrate that given the number of investing firms, the equilibrium capacity investment satisfies equation (3.3) or (3.4) respectively, and show the monotonicity of the equilibrium capacity with respect to the variable capacity cost w. Then we show the monotonicity of the number of investing firms with respect to the fixed capacity cost w_0 by constructing the equilibrium switching curves, and finally show the monotonicity of the equilibrium switching curves.

Both firms investing: If both firms decide to invest in the supplier, we first show that the equilibrium capacity satisfies $k_1 = k_2$, and then the equilibrium capacity is the k^e as shown in condition (3.3).

We first show that if $k_1 > k_2$, then (k_1, k_2) cannot be the equilibrium capacity. Following equation (3.1) and (3.2), we have

$$\frac{\partial V_1^e}{\partial k_1} = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}} \left[a - c - b\left(\frac{3k_0}{2} + 2k_1 + k_2\right)\xi\right]\xi f(\xi)d\xi - w \tag{B.1}$$

$$\frac{\partial V_2^e}{\partial k_2} = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}} \left[a - c - b\left(\frac{3k_0}{2} + k_1 + 2k_2\right)\xi\right]\xi f(\xi)d\xi$$
$$+ \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}}^{\frac{a-c}{3b(\frac{k_0}{2}+k_2)}} \frac{a - c - b(k_0 + 2k_2)\xi}{2}\xi f(\xi)d\xi - w \tag{B.2}$$

Let $k_1 = k + \epsilon$ and $k_2 = k$ where $\epsilon > 0$, then we have

$$\left.\frac{\partial V_1^e}{\partial k_1}\right|_{(k+\epsilon,k)} - \left.\frac{\partial V_2^e}{\partial k_2}\right|_{(k+\epsilon,k)} = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+3k+2\epsilon)}} -b\epsilon\xi^2 f(\xi)d\xi$$
$$- \int_{\frac{a-c}{b(\frac{3k_0}{2}+3k+2\epsilon)}}^{\frac{a-c}{3b(\frac{k_0}{2}+k)}} \frac{a - c - b(k_0 + 2k)\xi}{2}\xi f(\xi)d\xi \leq 0 \tag{B.3}$$

Therefore, at least one of the two firms will have incentive to deviate from the current capacity investment level, and we have (k_1, k_2) where $k_1 > k_2$ cannot be an equilibrium. Similarly, we have (k_1, k_2) where $k_1 < k_2$ cannot be an equilibrium.

We next show that $k_1 = k_2 = k^e$ is indeed an equilibrium by showing that neither firm has incentive to deviate in this case. We focus on the analysis for firm 1 as the analysis for firm 2 is similar. If firm 1 deviates from k^e to $k^e + \epsilon$, then we have

$$\begin{aligned}\left.\frac{\partial V_1^e}{\partial k_1}\right|_{(k^e+\epsilon,k^e)} - \left.\frac{\partial V_1^e}{\partial k_1}\right|_{(k^e,k^e)} = &\int_0^{\frac{a-c}{b(\frac{3k_0}{2}+3k^e+2\epsilon)}} -2b\epsilon\xi^2 f(\xi)d\xi \\ &- \int_{\frac{a-c}{b(\frac{3k_0}{2}+3k^e+2\epsilon)}}^{\frac{a-c}{b(\frac{3k_0}{2}+3k^e)}} \left[a-c-b\left(\frac{3k_0}{2}+3k^e\right)\xi\right]\xi f(\xi)d\xi \le 0\end{aligned} \tag{B.4}$$

That is, firm 1 will have incentive to decrease its capacity investment from $k^e+\epsilon$. On the other hand, if firm 1 deviates from k^e to $k^e - \epsilon$. Then we have

$$\begin{aligned}\left.\frac{\partial V_1^e}{\partial k_1}\right|_{(k^e-\epsilon,k^e)} - \left.\frac{\partial V_1^e}{\partial k_1}\right|_{(k^e,k^e)} = &\int_0^{\frac{a-c}{b(\frac{3k_0}{2}+3k^e)}} 2b\epsilon\xi^2 f(\xi)d\xi \\ &+ \int_{\frac{a-c}{b(\frac{3k_0}{2}+3k^e)}}^{\frac{a-c}{b(\frac{3k_0}{2}+3k^e-\epsilon)}} \left[a-c-b\left(\frac{3k_0}{2}+3k^e-2\epsilon\right)\xi\right]\xi f(\xi)d\xi \\ &+ \int_{\frac{a-c}{b(\frac{3k_0}{2}+3k^e-\epsilon)}}^{\frac{a-c}{3b(\frac{k_0}{2}+k^e-\epsilon)}} \frac{a-c-b(k_0+2k^e-2\epsilon)\xi}{2}\xi f(\xi)d\xi \ge 0\end{aligned} \tag{B.5}$$

Therefore, firm 1 will have incentive to increase its capacity investment from $k^e - \epsilon$. To conclude, we have shown that if both firms invest in the supplier, the equilibrium capacity investment k^e is the same for both firms and is characterized by condition (3.3).

Then we implicitly differentiate k^e with respect to w in equation (3.3), and obtain that

$$\frac{\partial k^e}{\partial w} = -\frac{1}{3b\int_0^{\frac{a-c}{3b(\frac{k_0}{2}+k^e)}} \xi^2 f(\xi)d\xi} \leq 0 \tag{B.6}$$

Therefore, it follows that k^e decreases in w.

<u>One firm investing</u>: If only one firm invests, without loss of generality assuming the investing firm is firm 1, we have the first order derivative of firm 1's expected profit as

$$\left.\frac{\partial V_1^e}{\partial k_1}\right|_{(k_1,0)} = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1)}} \left[a - c - b\left(\frac{3k_0}{2} + 2k_1\right)\xi\right] \xi f(\xi)d\xi - w \tag{B.7}$$

It follows that the second order derivative

$$\left.\frac{\partial^2 V_1^e}{\partial k_1^2}\right|_{(k_1,0)} = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1)}} -2b\xi^2 f(\xi)d\xi \leq 0 \tag{B.8}$$

Therefore, firm 1 will choose k_1^e which satisfies the condition that $\frac{\partial V_1^e}{\partial k_1} = 0$.

Similarly, we implicitly differentiate k_1^e with respect to w in equation (3.4), and obtain that

$$\frac{\partial k_1^e}{\partial w} = -\frac{1}{2b\int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^e)}} \xi^2 f(\xi)d\xi} \leq 0 \tag{B.9}$$

Therefore, it follows that k_1^e decreases in w.

<u>Monotonicity of number of investing firms</u>: To simplify the notations, we define that $L_i^e(k_1, k_2; w) \triangleq E\left[\pi_i^e(k_1, k_2, \boldsymbol{\xi})\right] - wk_i$. For given w, we write $L_i^e(k_1, k_2; w)$ as $L_i^e(k_1, k_2)$ when there is no confusion. Then we have $V_i^e(k_1, k_2) = L_i^e(k_1, k_2) - w_0\mathbf{1}_{\{k_i>0\}}$. When only one firm invests, we still label the investing firm as firm 1. We prove the results

in three steps. First, we show in a technical lemma that $k^e \leq k_1^e \leq 2k^e$. Then we show that $L_1^e(k_1^e, 0) \geq L_1^e(0,0)$ and $L_2^e(k^e, k^e) \geq L_2^e(k_1^e, 0)$. Finally, we prove the monotonicity in the number of investing firms by deriving the switching curves, and then show the monotonicity of the equilibrium switching curves.

Lemma B.1 (Monotonicity in the capacity investment)**.** $k^e \leq k_1^e \leq 2k^e$.

Proof of Lemma B.1. By comparing equation (3.3) and (3.4), we have $3k^e = 2k_1^e$. Therefore, the results follow. □

We next show that $L_1^e(k_1^e, 0) \geq L_1^e(0,0)$ and $L_2^e(k^e, k^e) \geq L_2^e(k_1^e, 0)$. First, as $L_1^e(k_1^e, 0) \triangleq \max_{k \geq 0} L_1^e(k, 0)$, it follows that $L_1^e(k_1^e, 0) \geq L_1^e(0,0)$. Second, using equation (3.3) and (3.4) we have

$$
\begin{aligned}
L_2^e(k^e, k^e) &= \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+3k^e)}} m_2(\frac{k_0}{2} + k^e)\xi f(\xi)d\xi + \int_{\frac{a-c}{3b(\frac{k_0}{2}+k^e)}}^{1} \frac{m_0(a-c)}{3b} f(\xi)d\xi - wk^e \\
&= \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+3k^e)}} \left[(a-c-b(k_0+k^e)\xi)\frac{k_0}{2} + b(k^e)^2\xi\right] \xi f(\xi)d\xi \\
&\quad + \int_{\frac{a-c}{3b(\frac{k_0}{2}+k^e)}}^{1} \frac{m_0(a-c)}{3b} f(\xi)d\xi \\
&\triangleq E[f_2^e(k^e, k^e, \boldsymbol{\xi})] \qquad \text{(B.10)}\\
L_2^e(k_1^e, 0) &= \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^e)}} m_2\frac{k_0}{2}\xi f(\xi)d\xi + \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^e)}}^{\frac{2(a-c)}{3bk_0}} m_1\frac{k_0}{2}\xi f(\xi)d\xi + \int_{\frac{2(a-c)}{3bk_0}}^{1} \frac{m_0(a-c)}{3b} f(\xi)d\xi \\
&\triangleq E[f_2^e(k_1^e, 0, \boldsymbol{\xi})] \qquad \text{(B.11)}
\end{aligned}
$$

For any realization of ξ, the integrand $f_2^e(k^e, k^e, \xi) \geq f_2^e(k_1^e, 0, \xi)$ following Lemma B.1. Therefore, we have $L_2^e(k^e, k^e) \geq L_2^e(k_1^e, 0)$.

Finally, we define $\overline{w}_0^e(w) \triangleq L_1^e(k_1^e, 0; w) - L_1^e(0, 0; w)$, $\underline{w}_0^e(w) \triangleq \min\{L_2^e(k^e, k^e; w) - L_2^e(k_1^e, 0; w), \overline{w}_0^e(w)\}$, and $w_0^e(w) \triangleq V_i^e(k^e, k^e) - V_i^e(0, 0)$. It follows that when $w_0 \geq \overline{w}_0^e(w)$, neither firm has incentive to deviate from the status quo (neither firm invests in the supplier); when $\underline{w}_0^e(w) \leq w_0 < \overline{w}_0^e(w)$, only one firm invests in the supplier; when $w_0 < \underline{w}_0^e(w)$, both firms invest in the supplier. When $w_0^e(w) \leq w_0 < \underline{w}_0^e(w)$, both firms invest in the supplier but both firms earn a lower profit than they do when neither firms in the supplier. Therefore, both firms are trapped in a prisoner's dilemma.

For the monotonicity of the equilibrium switching curves, by envelope theorem, we have

$$\frac{\partial \overline{w}_0^e(w)}{\partial w} = -k_1^e \leq 0. \tag{B.12}$$

That is $\overline{w}_0^e(w)$ decreases in w. We also define $\hat{w}_0^e(w) \triangleq L_2^e(k^e, k^e; w) - L_2^e(k_1^e, 0; w)$. Then we have

$$\begin{aligned}
\frac{\partial \hat{w}_0^e(w)}{\partial w} &= \int_0^{\frac{a-c}{3b(\frac{k_0}{2}+k^e)}} -\frac{b(k_0+2k^e)}{2}\xi^2 f(\xi) d\xi \frac{\partial k^e}{\partial w} - k^e + \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^e)}} \frac{bk_0}{2}\xi^2 f(\xi) d\xi \frac{\partial k_1^e}{\partial w} \\
&= -\frac{8k^e + k_0}{12} \leq 0
\end{aligned} \tag{B.13}$$

The second equality follows equation (B.6) and (B.9). As $\underline{w}_0^e(w) \triangleq \min\{\hat{w}_0^e(w), \overline{w}_0^e(w)\}$, we have $\underline{w}_0^e(w)$ decreases in w. $\square$

Remark: One may observe that when $\overline{w}_0^e(w) < L_2^e(k^e, k^e) - L_2^e(k_1^e, 0)$ and $w_0 \in [\overline{w}_0^e(w), L_2^e(k^e, k^e) - L_2^e(k_1^e, 0)]$, in theory there exist two equilibria: both firms investing in the supplier, and neither firm investing in the supplier. However, as the status quo of this game is that neither firm invests in the supplier at the first place (and both firms are deciding simultaneously about whether they should invest in the supplier), the final equilibrium outcome of this game is still that neither firm invests in the supplier. This is why we define $\underline{w}_0^e(w) \triangleq \min\{L_2^e(k^e, k^e) - L_2^e(k_1^e, 0), \overline{w}_0^e(w)\}$. That is, we always have $\underline{w}_0^e(w) \leq \overline{w}_0^e(w)$.

The proof in this section is similar to the proofs in Section 3.3. Therefore, we will sketch the proof and illustrate details of the important steps for the interest of space.

Proof of Lemma III.3. The proof is similar to the proof of Lemma III.1. Without loss of generality, we assume $k_1 \geq k_2$. By concavity of the objective function, we obtain the best response functions as:

$$q_1^*(q_2) = \begin{cases} \frac{a-c-bq_2}{2b} & \text{if } \frac{a-c-bq_2}{2b} < (\frac{k_0}{2} + k_1)\xi, \\ (\frac{k_0}{2} + k_1)\xi & \text{if } \frac{a-c-bq_2}{2b} \geq (\frac{k_0}{2} + k_1)\xi. \end{cases}$$

$$q_2^*(q_1) = \begin{cases} \frac{a-c-bq_1}{2b} & \text{if } \frac{a-c-bq_1}{2b} < k_s - q_1, \\ k_s - q_1 & \text{if } \frac{a-c-bq_1}{2b} \geq k_s - q_1. \end{cases}$$

Similarly, we solve for the intersection of the best response functions $q_1^*(q_2)$ and $q_2^*(q_1)$, and obtain the equilibrium order quantities and profits in Lemma III.3. $\square$

Before we prove Proposition III.4, we first derive the expressions for firms' profit based on Lemma III.3, assuming $k_1 \geq k_2$.

$$V_1^f(k_1,k_2) = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}} m_2\left(\frac{k_0}{2}+k_1\right)\xi f(\xi)d\xi + \int_{\frac{2(a-c)}{3b(k_0+k_1+k_2)}}^{1} \frac{m_0(a-c)}{3b} f(\xi)d\xi$$

$$+ \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}}^{\frac{2(a-c)}{3b(k_0+k_1+k_2)}} m_2\left(\frac{a-c}{b}-k_s\right) f(\xi)d\xi - w_0\mathbf{1}_{\{k_1>0\}} - wk_1 \qquad \text{(B.14)}$$

$$V_2^f(k_1,k_2) = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}} m_2\left(\frac{k_0}{2}+k_2\right)\xi f(\xi)d\xi + \int_{\frac{2(a-c)}{3b(k_0+k_1+k_2)}}^{1} \frac{m_0(a-c)}{3b} f(\xi)d\xi$$

$$+ \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}}^{\frac{2(a-c)}{3b(k_0+k_1+k_2)}} m_2\left(2k_s-\frac{a-c}{b}\right) f(\xi)d\xi - w_0\mathbf{1}_{\{k_2>0\}} - wk_2 \qquad \text{(B.15)}$$

We also prove the following technical lemma that will be used in the proof of Proposition III.4.

Lemma B.2. *For a finite differentiable function $h(x)$ defined on $x \in \mathbb{R}^+$, if $h'(0) \geq 0$ and $\lim_{x\to\infty} h'(x) = 0$, then $x^*(w) = \arg\max_x\{h(x) - wx\}$ decreases in w.*

Proof of Lemma B.2. If $h(x)$ is monotone, the proof is trivial. We prove the lemma for the case where $g(x;w) \triangleq h(x) - wx$ may have two local maxima, $\hat{x}_1(w) \leq \hat{x}_2(w)$. For the cases where $g(x;w)$ has more than two local maxima, the lemma can be proved similarly. We denote the local minimum between the two local maxima as $\underline{x}(w)$.

To prove the lemma, it is sufficient to show that if $g(\hat{x}_1(w);w) \geq g(\hat{x}_2(w);w)$, then for $\hat{w} > w$, we have $g(\hat{x}_1(\hat{w});\hat{w}) \geq g(\hat{x}_2(\hat{w});\hat{w})$. First, as the function $g(x;w)$ has at most two local maxima, it follows that $\hat{x}_1(\hat{w})$ and $\hat{x}_2(\hat{w})$ decrease in w, and

$\underline{x}(w)$ increases in w. Second, we have

$$
\begin{aligned}
g(\hat{x}_2(\hat{w})) - g(\hat{x}_1(\hat{w})) &= \int_{\hat{x}_1(\hat{w})}^{\underline{x}(\hat{w})} g'(x;\hat{w})dx + \int_{\underline{x}(\hat{w})}^{\hat{x}_2(\hat{w})} g'(x;\hat{w})dx \\
&= \int_{\hat{x}_1(\hat{w})}^{\underline{x}(\hat{w})} (h'(x) - \hat{w})dx + \int_{\underline{x}(\hat{w})}^{\hat{x}_2(\hat{w})} (h'(x) - \hat{w})dx \\
&\leq \int_{\hat{x}_1(w)}^{\underline{x}(w)} (h'(x) - w)dx + \int_{\underline{x}(w)}^{\hat{x}_2(w)} (h'(x) - w)dx \\
&= g(\hat{x}_2(w)) - g(\hat{x}_1(w)) \leq 0 \qquad \text{(B.16)}
\end{aligned}
$$

Therefore, we have proved that $g(\hat{x}_1(\hat{w})) \geq g(\hat{x}_2(\hat{w}))$. □

Proof of Proposition III.4. The proof is similar to the proof of Proposition III.2 with two key steps. We first derive the subgame perfect equilibrium capacity investment given the number of investing firms. Then we characterize the monotonicity of number of investing firms.

Both firms investing: If both firms investing in the supplier, we first obtain the first order derivative of firms' profit with respect to its capacity investment as follows, assuming $k_1 > k_2$.

$$
\begin{aligned}
\frac{\partial V_1^f}{\partial k_1} = &\int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}} \left[a - c - b\left(\frac{3k_0}{2} + 2k_1 + k_2\right)\xi\right] \xi f(\xi) d\xi \\
&+ \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}}^{\frac{2(a-c)}{3b(k_0+k_1+k_2)}} -2[a - c - b(k_0 + k_1 + k_2)\xi]\xi f(\xi)d\xi - w \qquad \text{(B.17)}
\end{aligned}
$$

$$\frac{\partial V_2^f}{\partial k_2} = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}} \left[a - c - b\left(\frac{3k_0}{2} + k_1 + 2k_2\right)\xi\right]\xi f(\xi)d\xi$$
$$+ \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}}^{\frac{2(a-c)}{3b(k_0+k_1+k_2)}} [3(a-c) - 4b(k_0+k_1+k_2)\xi]\xi f(\xi)d\xi - w \tag{B.18}$$

Then following similar steps as in proof of Proposition III.2, we have (k_1, k_2) where $k_1 \neq k_2$ cannot be an equilibrium, and $k_1 = k_2 = k^f$ where k^f is defined in Proposition III.4 is indeed an equilibrium. The details are omitted for space. Then we implicitly differentiate k^f with respect to w in equation (3.5), and obtain that

$$\frac{\partial k^f}{\partial w} = -\frac{1}{3b\int_0^{\frac{a-c}{3b(\frac{k_0}{2}+k^f)}} \xi^2 f(\xi)d\xi} \leq 0 \tag{B.19}$$

Therefore, it follows that k^f decreases in w.

Only one firm investing: If only one firm invests, we assume the investing firm is firm 1 and obtain the following first order derivative of firm 1's expected profit with respect to its capacity investment.

$$\left.\frac{\partial V_1^f}{\partial k_1}\right|_{(k_1,0)} = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1)}} \left[a - c - b\left(\frac{3k_0}{2} + 2k_1\right)\xi\right]\xi f(\xi)d\xi$$
$$+ \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1)}}^{\frac{2(a-c)}{3b(k_0+k_1)}} -2[a-c-b(k_0+k_1)\xi]\xi f(\xi)d\xi - w \tag{B.20}$$

We observe that $\lim_{k_1\to\infty} \left.\frac{\partial V_1^f}{\partial k_1}\right|_{(k_1,0)} = -w$. In addition, $V_1^f(k_1, 0)$ is a continuous and differentiable function in k_1. Therefore, there exists a finite $k_1 = k_1^f$ where $\max_{k_1\geq 0} V_1^f(k_1, 0)$ is attained, and the k_1^f satisfies the condition specified by the first order condition as shown in equation (3.6). The decrease of k_1^f with respect to w

follows Lemma B.2.

Monotonicity of number of investing firms: To simplify the notations, we similarly define $L_i^f(k_1, k_2) \triangleq E\left[\pi_i^f(k_1, k_2, \boldsymbol{\xi})\right] - wk_i$. Then we have $V_i^f(k_1, k_2) = L_i^f(k_1, k_2) - w_0 \mathbf{1}_{\{k_i > 0\}}$. When only one firm invests, we still label the investing firm as firm 1. We prove the results in two steps. First, we show that $L_1^f(k_1^f, 0) \geq L_1^f(0, 0)$. Then we define $\overline{w}_0^f(w) \triangleq L_1^f(k_1^f, 0) - L_1^f(0, 0)$, $\underline{w}_0^f(w) \triangleq \min\{(L_2^f(k^f, k^f) - L_2^f(k_1^f, 0))^+, \overline{w}_0^f(w)\}$, and $w_0^f(w) \triangleq V_i^f(k^f, k^f) - V_i^f(0, 0)$. It follows that when $w_0 \geq \overline{w}_0^f(w)$, neither firm has incentive to invest in the supplier; when $\underline{w}_0^f(w) \leq w_0 < \overline{w}_0^f(w)$, only one firm invests in the supplier; when $w_0 < \underline{w}_0^f(w)$, both firms invest in the supplier. When $w_0^f(w) \leq w_0 < \underline{w}_0^f(w)$, both firms invest in the supplier but both firms earn a lower profit than they do when neither firms in the supplier. Therefore, both firms are trapped in a prisoner's dilemma.

For the monotonicity of the equilibrium switching curve $\overline{w}_0^f(w)$, by envelope theorem, we have

$$\frac{\partial \overline{w}_0^f(w)}{\partial w} = -k_1^f \leq 0. \tag{B.21}$$

That is $\overline{w}_0^f(w)$ decreases in w. □

When only one firm invests, we label the investing firm as firm 1. Before we proceed to prove propositions in this section, we first prove two technical lemmas.

Lemma B.3 (Over-investment with exclusive capacity). $k_1^f \leq k_1^e$.

Proof of Lemma B.3. Assume $k_1^f > k_1^e$, then following equation (3.6) and (3.4), we have

$$\int_{0}^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^f)}} \left[a-c-b\left(\frac{3k_0}{2}+2k_1^f\right)\xi\right]\xi f(\xi)d\xi$$
$$+\int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^f)}}^{\frac{2(a-c)}{3b(k_0+k_1^f)}} -2\left[a-c-b\left(k_0+k_1^f\right)\xi\right]\xi f(\xi)d\xi - w$$
$$\leq \int_{0}^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^f)}} \left[a-c-b\left(\frac{3k_0}{2}+2k_1^f\right)\xi\right]\xi f(\xi)d\xi - w$$
$$< \int_{0}^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^e)}} \left[a-c-b\left(\frac{3k_0}{2}+2k_1^e\right)\xi\right]\xi f(\xi)d\xi - w = 0 \tag{B.22}$$

This contradicts the condition specified in equation (3.6). Therefore, we conclude that $k_1^f \leq k_1^e$. $\square$

Lemma B.4. *$V_1^e(k_1^e, 0) \geq V_1^f(k_1^f, 0)$; $V_2^f(k_1^f, 0) \geq V_2^e(k_1^e, 0)$.*

Proof of Lemma B.4. We first prove that $V_1^e(k_1^e, 0) \geq V_1^f(k_1^f, 0)$ by showing that $V_1^e(k_1^e, 0) \geq V_1^e(k_1^f, 0) \geq V_1^f(k_1^f, 0)$. We observe that $V_1^e(k_1^e, 0) \triangleq \max_{k\geq 0} V_1^e(k, 0)$, so we have $V_1^e(k_1^e, 0) \geq V_1^e(k_1^f, 0)$. Next, by equation (3.1) and (B.14), we have

$$V_1^e(k_1^f, 0) - V_1^f(k_1^f, 0) = \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^f)}}^{\frac{2(a-c)}{3b(k_0+k_1^f)}} \left\{\frac{[a-c-b(\frac{k_0}{2})\xi]^2}{4b} - \frac{(a-c-bk_s)^2}{b}\right\} f(\xi)d\xi$$
$$+\int_{\frac{2(a-c)}{3b(k_0+k_1^f)}}^{\frac{2(a-c)}{3bk_0}} \left\{\frac{[a-c-b(\frac{k_0}{2})\xi]^2}{4b} - \frac{(a-c)^2}{9b}\right\} f(\xi)d\xi \tag{B.23}$$

We note that $\frac{[a-c-b(\frac{k_0}{2})\xi]^2}{4b} - \frac{(a-c-bk_s)^2}{b} \geq 0$ when $\frac{a-c}{b(\frac{3k_0}{2}+2k_1^f)} \leq \xi \leq \frac{2(a-c)}{3b(k_0+k_1^f)}$, and

$\frac{[a-c-b(\frac{k_0}{2})\xi]^2}{4b} - \frac{(a-c)^2}{9b} \geq 0$ when $\frac{2(a-c)}{3b(k_0+k_1^f)} \leq \xi \leq \frac{2(a-c)}{3bk_0}$. Therefore, it follows that $V_1^e(k_1^f, 0) \geq V_1^f(k_1^f, 0)$. Then we have proved that $V_1^e(k_1^e, 0) \geq V_1^f(k_1^f, 0)$.

We next prove that $V_2^f(k_1^f, 0) \geq V_2^e(k_1^e, 0)$ by showing that $V_2^f(k_1^f, 0) \geq V_2^e(k_1^f, 0) \geq V_2^e(k_1^e, 0)$. Similarly, by equation (3.2) and (B.15), we have

$$\begin{aligned}
& V_2^f(k_1^f, 0) - V_2^e(k_1^f, 0) \\
&= \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^f)}}^{\frac{2(a-c)}{3b(k_0+k_1^f)}} \left[(a-c-bk_s)\left(2k_s - \frac{a-c}{b}\right) - \frac{(a-c-b\frac{k_0}{2}\xi)k_0}{4}\xi\right] f(\xi)d\xi \\
&+ \int_{\frac{2(a-c)}{3b(k_0+k_1^f)}}^{\frac{2(a-c)}{3bk_0}} \left[\frac{(a-c)^2}{9b} - \frac{(a-c-b\frac{k_0}{2}\xi)k_0}{4}\xi\right] f(\xi)d\xi \qquad \text{(B.24)}
\end{aligned}$$

We note that $(a-c-bk_s)\left(2k_s - \frac{a-c}{b}\right) - \frac{(a-c-b\frac{k_0}{2}\xi)k_0}{4}\xi \geq 0$ when $\frac{a-c}{b(\frac{3k_0}{2}+2k_1^f)} \leq \xi \leq \frac{2(a-c)}{3b(k_0+k_1^f)}$, and $\frac{(a-c)^2}{9b} - \frac{(a-c-b\frac{k_0}{2}\xi)k_0}{4}\xi \geq 0$ when $\frac{2(a-c)}{3b(k_0+k_1^f)} \leq \xi \leq \frac{2(a-c)}{3bk_0}$. Therefore, we obtain that $V_2^f(k_1^f, 0) \geq V_2^e(k_1^f, 0)$. Then from equation (3.2), we have

$$\left.\frac{\partial V_2^e}{\partial k_1}\right|_{(k_1,0)} = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1+k_2)}} -\frac{bk_0\xi^2}{2} f(\xi)d\xi \leq 0 \qquad \text{(B.25)}$$

In addition, by Lemma B.3, we have $k_1^f \leq k_1^e$ and hence $V_2^e(k_1^f, 0) \geq V_2^e(k_1^e, 0)$. Then we have proved $V_2^f(k_1^f, 0) \geq V_2^e(k_1^e, 0)$. □

Proof of Proposition III.5. i and iii) We prove this part by showing that $w_0^e(w) = w_0^f(w)$ and $\underline{w}_0^e(w) \geq \underline{w}_0^f(w)$. First, if $k_1 = k_2$, we have $V_i^e(k_1, k_2) = V_i^f(k_1, k_2)$. In addition, by definition we have $w_0^e(w) = V_i^e(k^e, k^e) - V_i^e(0, 0)$ and $w_0^f(w) = V_i^f(k^f, k^f) - V_i^f(0, 0)$. Therefore, we have $w_0^e(w) = w_0^f(w)$. Second, we have $\underline{w}_0^e(w) = \min\{L_2^e(k^e, k^e) - L_2^e(k_1^e, 0), \overline{w}_0^e(w)\} \geq \underline{w}_0^f(w) = \min\{(L_2^f(k^f, k^f) - L_2^f(k_1^f, 0))^+, \overline{w}_0^f(w)\}$ because (1) we have $[L_2^e(k^e, k^e) - L_2^e(k_1^e, 0)] - [L_2^f(k^f, k^f) - L_2^f(k_1^f, 0)]^+ \geq 0$ because

$L_2^e(k^e,k^e) \geq L_2^e(k_1^e,0)$ (see the proof of Proposition III.2), $L_2^e(k^e,k^e) = L_2^f(k^f,k^f)$, and $L_2^e(k_1^e,0) - L_2^f(k_1^f,0) = V_2^e(k_1^e,0) - V_2^f(k_1^f,0) \leq 0$ by Lemma B.4; (2) we have $\overline{w}_0^e(w) = L_1^e(k_1^e,0) - L_1^e(0,0)$ and $\overline{w}_0^f(w) = L_1^f(k_1^f,0) - L_1^f(0,0)$, so $\overline{w}_0^e(w) - \overline{w}_0^f(w) = L_1^e(k_1^e,0) - L_1^f(k_1^f,0) = V_1^e(k_1^e,0) + wk_1^e - V_1^f(k_1^f,0) - wk_1^f \geq 0$ by Lemma B.3 and B.4.
ii) We first observe by equation (3.3) and (3.5) that $k^e = k^f$. Then by Lemma B.1, we have $k^e \leq k_1^e \leq 2k^e$. By Lemma B.3, we have $k_1^f \leq k_1^e$. Therefore, when $0 \leq w_0 < \underline{w}_0^f(w)$, both firms invest with either capacity type so the total capacity investment $2k^e = 2k^f$; when $\underline{w}_0^f(w) \leq w_0 < \underline{w}_0^e(w)$, both firms invest with exclusive capacity but only one firm invests with first-priority capacity, so the total capacity investment $2k^e \geq k_1^f$; when $\underline{w}_0^e(w) \leq w_0 < \overline{w}_0^f(w)$, only one firm invests with either capacity type, so the total capacity investment $k_1^e \geq k_1^f$; when $\overline{w}_0^f(w) \leq w_0 < \overline{w}_0^e(w)$, one firm invests with exclusive capacity while neither firm invests with first-priority capacity, so the total capacity investment $k_1^e \geq 0$; when $w_0 \geq \overline{w}_0^e(w)$, neither firm invests with either type of capacity. Therefore, the total capacity investment is higher with exclusive capacity. □

Proof of Proposition III.6. i) When $0 \leq w_0 < \underline{w}_0^f(w)$, both firms invest in the supplier with either capacity type, so we have $V_i^e(k^e,k^e) = V_i^f(k^f,k^f)$ and firms are indifferent between exclusive and first-priority capacity.
ii) When $\underline{w}_0^f(w) \leq w_0 < \underline{w}_0^e(w)$, only one firm invests with first-priority capacity and both firms invest with exclusive capacity. We have $V_1^f(k_1^f,0) \geq V_1^f(k^f,0) \geq V_1^f(k^f,k^f) = V_1^e(k^e,k^e)$ and $V_2^f(k_1^f,0) \geq V_2^f(k^f,k^f) = V_2^e(k^e,k^e)$. Therefore, both firms prefer the first-priority capacity.
iii) When $\underline{w}_0^e(w) \leq w_0 < \overline{w}_0^f(w)$, only one firm invests with either type of capacity. We have $V_1^e(k_1^e,0) \geq V_1^f(k_1^f,0)$ and $V_2^e(k_1^e,0) \leq V_2^f(k_1^f,0)$ by Lemma B.4. When $\overline{w}_0^f(w) \leq w_0 < \overline{w}_0^e(w)$, one firm invests with exclusive capacity while neither firm invests with first-priority capacity, $V_1^e(k_1^e,0) \geq V_1^e(0,0) = V_1^f(0,0)$ and $V_2^e(k_1^e,0) \leq V_2^e(0,0) = V_2^f(0,0)$ by equation (B.25).

iv) When $w_0 \geq \overline{w}_0^e(w)$, neither firm invests with either type capacity. Therefore, the firms are indifferent between the exclusive and first-priority capacity, i.e., $V_i^e(0,0) = V_i^f(0,0)$. □

Proof of Proposition III.7. i) By equation (3.7) and (3.8), if $k_1 = k_2$, we have $V_s^e(k_1,k_2) = V_s^f(k_1,k_2)$. Therefore, when $0 \leq w_0 < \underline{w}_0^f(w)$ or $w_0 \geq \overline{w}_0^e(w)$, we have $V_s^e(k^e,k^e) = V_s^f(k^f,k^f)$ and $V_s^e(0,0) = V_s^f(0,0)$ respectively.
ii) When $\underline{w}_0^f(w) \leq w_0 < \underline{w}_0^e(w)$, both firms invest with exclusive capacity and only one firm invests with first-priority capacity, so by equation (3.7) and (3.8), we have

$$V_s^e(k^e,k^e) = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+3k^e)}} c(k_0+2k^e)\xi f(\xi)d\xi + \int_{\frac{a-c}{b(\frac{3k_0}{2}+3k^e)}}^{1} \frac{2c(a-c)}{3b} f(\xi)d\xi \tag{B.26}$$

$$V_s^f(k_1^f,0) = \int_0^{\frac{2(a-c)}{3b(k_0+k_1^f)}} c(k_0+k_1^f)\xi f(\xi)d\xi + \int_{\frac{2(a-c)}{3b(k_0+k_1^f)}}^{1} \frac{2c(a-c)}{3b} f(\xi)d\xi \tag{B.27}$$

We note that $2k^e = 2k^f \geq k_1^e \geq k_1^f$ (see Lemma B.1 and B.3.) Therefore, it follows that $V_s^e(k^e,k^e) \geq V_s^f(k_1^f,0)$.

Similarly, when $\overline{w}_0^f(w) \leq w_0 < \overline{w}_0^e(w)$, one firm invests with exclusive capacity and neither firm invests with first-priority capacity, so we have $V_s^e(k_1^e,0) \geq V_s^f(0,0)$. The details are omitted for space. □

Lemma B.5 (First-best order quantity and ex post profit)**.**

Let the capacity investment sizes be k, and we obtain the following:

realized yield ξ	*order quantity* q^m	*ex post profit* π^m
$0 \leq \xi \leq \frac{a}{2b(k_0+k)}$	$(k_0+k)\xi$	$(a-b(k_0+k)\xi)(k_0+k)\xi$
$\frac{a}{2b(k_0+k)} < \xi \leq 1$	$\frac{a}{2b}$	$\frac{a^2}{4b}$

Proof of Lemma B.5. The proof is similar to the proof of Lemma III.1 with only one firm in the market. The details are omitted for space. □

Proof of Proposition III.8. Given the capacity investment level of k, the firm's expected profit $V^m(k)$ is as follows:

$$V^m(k) = \int_0^{\frac{a}{2b(k_0+k)}} [a - b(k_0 + k)\xi](k_0 + k)\xi f(\xi)d\xi + \int_{\frac{a}{2b(k_0+k)}}^{1} \frac{a^2}{4b} f(\xi)d\xi - w_0 \mathbf{1}_{\{k>0\}} - wk \tag{B.28}$$

Therefore, if the firm decides to invest, the optimal capacity investment level k^m should satisfy the following first order condition:

$$\int_0^{\frac{a}{2b(k_0+k^m)}} [a - 2b(k_0 + k^m)\xi]\xi f(\xi)d\xi - w = 0 \tag{B.29}$$

It follows that

$$\frac{\partial k^m}{w} = -\frac{1}{2b\int_0^{\frac{a}{2b(k_0+k^m)}} \xi^2 f(\xi)d\xi} \leq 0 \tag{B.30}$$

so we have k^m decreases in w.

We then define $L^m(k; w) \triangleq E[\pi^m(k, \boldsymbol{\xi})] - wk$, and $w_0^m(w) \triangleq L^m(k^m) - L^m(0)$. Then it follows that if $w_0 \leq w_0^m(w)$, the monopoly will invest k^m; if $w_0 > w_0^m(w)$, the monopoly will not invest. For the monotonicity of $w_0^m(w)$, by envelope theorem, we have

$$\frac{\partial w_0^m(w)}{w} = -k^m \leq 0 \tag{B.31}$$

Therefore, we have proved that $w_0^m(w)$ decreases in w. □

Proof of Proposition III.9. i) We know from equation (3.3) and (3.5) that $k^e = k^f$. Then in equation (3.3), if we substitute $2k^e$ with k^m, the first order condition is as follows:

$$\int_0^{\frac{2a}{3b(k_0+k^m)}} \left[a - \frac{3b\,(k_0+k^m)}{2}\xi\right] \xi f(\xi) d\xi - w$$
$$\geq \int_0^{\frac{a}{2b(k_0+k^m)}} [a - 2b(k_0+k^m)\xi]\xi f(\xi) d\xi - w = 0 \tag{B.32}$$

Therefore, we have $2k^e = 2k^f \geq k^m$. That is, the total investment in the competing case exceeds the total investment in the first-best case.

We next show that $k_1^e \geq k^m$. In equation (3.4), if we substitute k_1^e with k^m, then we have

$$\int_0^{\frac{a}{b(\frac{3k_0}{2}+2k^m)}} \left[a - b\left(\frac{3k_0}{2} + 2k^m\right)\xi\right] \xi f(\xi) d\xi - w \geq \int_0^{\frac{a}{2b(k_0+k^m)}} [a - 2b\,(k_0+k^m)\,\xi]\,\xi f(\xi) d\xi - w = 0 \tag{B.33}$$

Therefore, we have that $k_1^e \geq k^m$.

Finally, by definition we have

$$w_0^m(w) = \int_0^{\frac{a}{2b(k_0+k^m)}} b(k^m)^2\xi^2 f(\xi) d\xi + \int_{\frac{a}{2b(k_0+k^m)}}^{\frac{a}{2bk_0}} \left[\frac{a^2}{4b} - (a - bk_0\xi)k_0\xi\right] f(\xi) d\xi \tag{B.34}$$

$$\overline{w}_0^e(w) = \int_0^{\frac{a}{b(\frac{3k_0}{2}+2k_1^e)}} b(k_1^e)^2\xi^2 f(\xi) d\xi + \int_{\frac{a}{b(\frac{3k_0}{2}+2k_1^e)}}^{\frac{2a}{3bk_0}} \left[\frac{(a - \frac{bk_0\xi}{2})^2}{4b} - (a - bk_0\xi)\frac{k_0}{2}\xi\right] f(\xi) d\xi \tag{B.35}$$

As the integrand of $\overline{w}_0^e(w)$ is greater than the integrand of $w_0^m(w)$ and both integrands are greater than 0, we have $\overline{w}_0^e(w) \geq w_0^m(w)$.

ii) See Proposition III.5 (ii). $\square$

Proof of Proposition III.10. We first show that k^e and k_1^e decreases in c. Implicitly differentiate equation (3.3) with respect to c, we obtain

$$\frac{\partial k^e}{\partial c} = -\frac{\int_0^{\frac{a-c}{3b(\frac{k_0}{2}+k)}} \xi f(\xi) d\xi}{3b \int_0^{\frac{a-c}{3b(\frac{k_0}{2}+k)}} \xi^2 f(\xi) d\xi} \leq 0 \tag{B.36}$$

Recall that $3k^e = 2k_1^e$ (see the proof of Lemma B.1.) Therefore, we also have $\frac{\partial k_1^e}{\partial c} \leq 0$.

We next show that $\overline{w}^e(w)$ and $\underline{w}^e(w)$ decrease in c. Taking a partial derivative with respect to c and applying envelope theorem and $3k^e = 2k_1^e$, we have

$$\frac{\partial \overline{w}^e(w)}{\partial c} = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^e)}} -k_1^e \xi f(\xi) d\xi + \int_{\frac{a-c}{b(\frac{3k_0}{2}+2k_1^e)}}^{\frac{2(a-c)}{3bk_0}} -\frac{a-c-\frac{3bk_0\xi}{2}}{2b} \xi f(\xi) d\xi \leq 0 \tag{B.37}$$

$$\frac{\partial \underline{w}^e(w)}{\partial c} = \int_0^{\frac{a-c}{b(\frac{3k_0}{2}+3k^e)}} -\left(\frac{k_0}{12} + \frac{2}{3}k^e\right) \xi f(\xi) d\xi + \int_{\frac{a-c}{b(\frac{3k_0}{2}+3k^e)}}^{\frac{2(a-c)}{3bk_0}} \left(\frac{k_0}{4}\xi - \frac{2(a-c)}{9b}\right) f(\xi) d\xi \leq 0 \tag{B.38}$$

Therefore we have that both $\overline{w}^e(w)$ and $\underline{w}^e(w)$ decrease in c. $\square$

APPENDIX C

Proofs and Technical Details for Chapter IV: To Share or Not to Share? Capacity Investments in a Shared Supplier

In the supporting document, we provide detailed proofs and technical details for all results in the independent market case. For the Cournot market and the supplier's profit sections, we sketch the outline of proofs, provide additional technical details and detailed proofs for important results, and illustrate most proofs from one firm's perspective, as proofs for the other case are similar and hence omitted for space. To simplify the notations, given capacity types (κ_i, κ_j), we define $V_i^{\kappa_i\kappa_j} \triangleq \Pi_i^{\kappa_i\kappa_j}(K_i, K_j)$, firm i's subgame perfect expected profit as $\Pi_i^{\kappa_i\kappa_j}(w_{\kappa_i}, w_{\kappa_j}) \triangleq \Pi_i^{\kappa_i\kappa_j}(K_i^{\kappa_i\kappa_j}, K_j^{\kappa_i\kappa_j}, w_{\kappa_i}, w_{\kappa_j})$, and for given capacity costs, we define $\Pi_i^{\kappa_i\kappa_j} \triangleq \Pi_i^{\kappa_i\kappa_j}(K_i^{\kappa_i\kappa_j}, K_j^{\kappa_i\kappa_j})$. Finally, we simplify integrals using $\mathbf{d}F_i$ for $\mathbf{d}F_i(a_i)$ in the independent market, and $\mathbf{d}F$ for $\mathbf{d}F(a)$ in the Cournot market.

Proof of Lemma IV.1. We consider two cases, depending on firm j's capacity type choice:

<u>Case $(\kappa_i e)$</u>: If firm j chooses the exclusive capacity $(\kappa_j = e)$, firm i solves $\max_{q_i} (a_i - b_i q_i - c)q_i$ s.t. $0 \leq q_i \leq K_i$. Following the concavity of the profit, we have firm i's

optimal order decision $q_i^* = \min\left\{\frac{(a_i-c)^+}{2b_i},\ K_i\right\}$.

<u>Case $(\kappa_i f)$</u>: If firm j chooses the first-priority capacity ($\kappa_j = e$), firm i solves $\max_{q_i}(a_i - b_i q_i - c)q_i$ s.t. $0 \le q_i \le K_i + (K_j - q_j)^+$. Hence, firm i's best response function is $q_i^* = \min\left\{\frac{(a_i-c)^+}{2b_i},\ K_i + (K_j - q_j)^+\right\}$. Notice firm j's best response function is $q_j^* = \min\left\{\frac{(a_j-c)^+}{2b_j},\ K_j\right\}$ if $\kappa_i = e$, or $q_j^* = \min\left\{\frac{(a_j-c)^+}{2b_j},\ K_j + (K_i - q_i)^+\right\}$ if $\kappa_i = f$. In either case, solving the system of best response functions yields the order quantities. □

Firms' equilibrium capacity investment We characterize firm i's equilibrium capacity by analyzing derivatives of firms' profits with respect to firms' capacity level decisions. In addition, when $\kappa_j = f$, we show the monotonicity of firm i's profit with respect to K_j, and the monotonicity of the best response function, which will be used to prove other results. In what follows, we focus on the non-trivial case where $K_i^{\kappa_i\kappa_j} > 0$.

<u>Case $(\kappa_i e)$</u>: We take the first order derivative of $V_i^{\kappa_i e} = \Pi_i^{\kappa_i e}(K_i, K_j)$ with respect to K_i, and obtain

$$\frac{\partial V_i^{\kappa_i e}}{\partial K_i} = \int_{2b_iK_i+c}^{\infty} (a_i - c - 2b_iK_i)\,\mathbf{d}F_i - w_{\kappa_i} \tag{C.1}$$

Then the second order derivative is as follows.

$$\frac{\partial^2 V_i^{\kappa_i e}}{\partial K_i^2} = \int_{2b_iK_i+c}^{\infty} -2b_i\mathbf{d}F_i \le 0$$

Therefore, observing that $V_i^{\kappa_i e}$ does not change with respect to K_j, we obtain the optimal capacity to build as $K_i^{\kappa_i e} = \left\{K :\ \frac{\partial V_i^{\kappa_i e}}{\partial K_i}(K, K_j) = 0\right\}$.

<u>Case $(\kappa_i f)$</u>: We take the first order derivative of $V_i^{\kappa_i f} = \Pi_i^{\kappa_i f}(K_i, K_j)$ with respect to

K_i as follows:

$$
\begin{aligned}
\frac{\partial V_i^{\kappa_i f}}{\partial K_i} = & \int_0^c \int_{2b_i(K_i+K_j)+c}^{\infty} [a_i - c - 2b_i(K_i + K_j)]\, \mathbf{d}F_i \mathbf{d}F_j \\
& + \int_c^{2b_jK_j+c} \int_{2b_i(K_i+K_j)+c-\frac{b_i(a_j-c)}{b_j}}^{\infty} \left[a_i - c + \frac{b_i(a_j-c)}{b_j} - 2b_i(K_i+K_j)\right] \mathbf{d}F_i \mathbf{d}F_j \\
& + \int_{2b_jK_j+c}^{\infty} \int_{2b_iK_i+c}^{\infty} (a_i - c - 2b_iK_i)\, \mathbf{d}F_i \mathbf{d}F_j - w_{\kappa_i}
\end{aligned} \tag{C.2}
$$

Therefore, the second order derivative is as follows.

$$
\begin{aligned}
\frac{\partial^2 V_i^{\kappa_i f}}{\partial K_i^2} = & -2b_i \int_0^c \int_{2b_i(K_i+K_j)+c}^{\infty} \mathbf{d}F_i \mathbf{d}F_j - 2b_i \int_{2b_jK_j+c}^{\infty} \int_{2b_iK_i+c}^{\infty} \mathbf{d}F_i \mathbf{d}F_j \\
& -2b_i \int_c^{2b_jK_j+c} \int_{2b_i(K_i+K_j)+c-\frac{b_i(a_j-c)}{b_j}}^{\infty} \mathbf{d}F_i \mathbf{d}F_j \leq 0
\end{aligned}
$$

Hence, for a fixed K_j, define $\hat{K}_i(K_j) = \left\{K : \frac{\partial V_i^{\kappa_i f}}{\partial K_i}(K, K_j) = 0\right\}$ and the best response function is $K_i^*(K_j) = \max\left\{\hat{K}_i(K_j),\ 0\right\}$. Finally, we have that firms' equilibrium capacity levels $(K_i^{\kappa_i f}, K_j^{f\kappa_i})$ are $\left\{(K_i, K_j) : \frac{\partial V_i^{\kappa_i f}}{\partial K_i}(K_i, K_j) = 0 \text{ and } \frac{\partial V_j^{f\kappa_i}}{\partial K_j}(K_j, K_i) = 0\right\}$.

We next characterize the monotonicity of firm i's profit with respect to K_j when $\kappa_j = f$.

Lemma C.1. $\frac{\partial V_i^{\kappa_i f}}{\partial K_j} \geq 0$.

Proof of Lemma C.1. Taking derivative of $V_i^{\kappa_i f}$ with respect to K_j yields

$$\frac{\partial V_i^{\kappa_i f}}{\partial K_j} = \int_0^c \int_{2b_i(K_i+K_j)+c}^{\infty} \left[a_i - c - 2b_i(K_i + K_j)\right] \mathbf{d}F_i \mathbf{d}F_j$$
$$+ \int_c^{2b_jK_j+c} \int_{2b_i(K_i+K_j)+c-\frac{b_i(a_j-c)}{b_j}}^{\infty} \left[a_i - c + \frac{b_i(a_j - c)}{b_j} - 2b_i(K_i + K_j)\right] \mathbf{d}F_i \mathbf{d}F_j \geq 0$$

□

Intuitively, if firm j's capacity is higher, it is more likely that firm i can leverage a part of it. We finally characterize the monotonicity of the best response function when $\kappa_j = f$ as follows.

Lemma C.2. $-1 \leq \frac{\mathbf{d}K_i^*(K_j)}{\mathbf{d}K_j} \leq 0$ *almost everywhere.*

Proof of Lemma C.2. Implicitly differentiating $\frac{\partial V_i^{\kappa_i f}}{\partial K_i} = 0$ with respect to K_j, we obtain that

$$\frac{\mathbf{d}\hat{K}_i(K_j)}{\mathbf{d}K_j}$$
$$= -\frac{\int_c^{2bK_j+c} \int_{2b(K_i+K_j)+2c-a_j}^{\infty} \mathbf{d}F_i \mathbf{d}F_j + \int_0^c \int_{2b(K_i+K_j)+c}^{\infty} \mathbf{d}F_i \mathbf{d}F_j}{\int_{2bK_j+c}^{\infty} \int_{2bK_i+c}^{\infty} \mathbf{d}F_i \mathbf{d}F_j + \int_c^{2bK_j+c} \int_{2b(K_i+K_j)+2c-a_j}^{\infty} \mathbf{d}F_i \mathbf{d}F_j + \int_0^c \int_{2b(K_i+K_j)+c}^{\infty} \mathbf{d}F_i \mathbf{d}F_j}$$

It follows that $-1 \leq \frac{\mathbf{d}\hat{K}_i(K_j)}{\mathbf{d}K_j} \leq 0$. Recall $K_i^*(K_j) = \max\left\{\hat{K}_i(K_j),\ 0\right\}$, and we have $-1 \leq \frac{\mathbf{d}K_i^*(K_j)}{\mathbf{d}K_j} \leq 0$ a.e. □

Intuitively, the higher firm j's capacity level is, the lower that firm i needs to build for herself because of the higher chance to access firm j's leftover capacity.

Proof of Lemma IV.2. Part (i): When $w_e > w_f$, we have $K_i^{fe} \geq K_i^{ee}$ because $\frac{\partial V_i^{fe}}{\partial K_i} = \frac{\partial V_i^{ee}}{\partial K_i} + (w_e - w_f) \geq \frac{\partial V_i^{ee}}{\partial K_i}$, and $\frac{\partial V_i^{fe}}{\partial K_j} = \frac{\partial V_i^{ee}}{\partial K_j} = 0$. We also have $K_i^{fe} \geq K_i^{ff}$ because

$\frac{\partial V_i^{fe}}{\partial K_j} = 0$ and

$$\begin{aligned}
&\frac{\partial V_i^{fe}}{\partial K_i} - \frac{\partial V_i^{ff}}{\partial K_i} \\
= &\int\limits_{2b_iK_i+c}^{\infty} (a_i - c - 2b_iK_i)\,\mathbf{d}F_i - \int\limits_{2b_jK_j+c}^{\infty}\int\limits_{2b_iK_i+c}^{\infty} (a_i - c - 2b_iK_i)\,\mathbf{d}F_i\mathbf{d}F_j \\
&- \int\limits_{c}^{2b_jK_j+c}\int\limits_{2b_i(K_i+K_j)+c-\frac{b_i(a_j-c)}{b_j}}^{\infty} \left[a_i - c + \frac{b_i(a_j-c)}{b_j} - 2b_i(K_i+K_j)\right]\mathbf{d}F_i\mathbf{d}F_j \\
&- \int\limits_{0}^{c}\int\limits_{2b_i(K_i+K_j)+c}^{\infty} [a_i - c - 2b_i(K_i+K_j)]\,\mathbf{d}F_i\mathbf{d}F_j \\
&\geq \int\limits_{2b_iK_i+c}^{\infty} (a_i - c - 2b_iK_i)\,\mathbf{d}F_i - \int\limits_{0}^{\infty}\int\limits_{2b_iK_i+c}^{\infty} (a_i - c - 2b_iK_i)\,\mathbf{d}F_i\mathbf{d}F_j = 0
\end{aligned}$$

Similarly, we have $K_i^{ee} \geq K_i^{ef}$ because $\frac{\partial V_i^{ee}}{\partial K_j} = 0$ and $\frac{\partial V_i^{ee}}{\partial K_i} \geq \frac{\partial V_i^{ef}}{\partial K_i}$. Finally, we have $K_i^{ff} \geq K_i^{ef}$ as follows: $\frac{\partial V_i^{ff}}{\partial K_i} = \frac{\partial V_i^{ef}}{\partial K_i} + (w_e - w_f) \geq \frac{\partial V_i^{ef}}{\partial K_i}$, so define $\tilde{K}_i^{ff} \triangleq \arg\max_{K_i} \Pi_i^{ff}(K_i,\ K_j^{fe})$, and we have $\tilde{K}_i^{ff} \geq K_i^{ef}$; we also have $K_i^{ff} \geq \tilde{K}_i^{ff}$ by Lemma C.2 and the fact that $K_j^{fe} \geq K_j^{ff}$.

Part (ii): The analysis for $w_e \leq w_f$ is similar and therefore omitted for space. □

Proof of Theorem IV.3. Part (i) $w_e \leq w_f$. If $\kappa_j = e$, focusing on the case where $K_i^{ee} > 0$ and $K_i^{fe} > 0$, we have $\frac{\partial V_i^{\kappa_i e}}{\partial K_i}(K_i^{\kappa_i e}, K_j^{e\kappa_i}) = \int_{2b_iK_i^{\kappa_i e}+c}^{\infty} a_i - c - 2b_iK_i^{\kappa_i e}\mathbf{d}F_i - w_{\kappa_i} = 0$ from equation (C.1), and $K_i^{ee} \geq K_i^{fe}$ from Lemma IV.2. Therefore, recall equation (4.6), and we obtain that

$$\Pi_i^{\kappa_i e} = \int\limits_{c}^{2b_iK_i^{\kappa_i e}+c} \frac{(a_i-c)^2}{4b_i}\mathbf{d}F_i + \int\limits_{2b_iK_i^{\kappa_i e}+c}^{\infty} b_i(K_i^{\kappa_i e})^2\mathbf{d}F_i$$

It follows that the profit difference is

$$\Pi_i^{ee} - \Pi_i^{fe} = \int\limits_{2b_iK_i^{fe}+c}^{2b_iK_i^{ee}+c} \left[\frac{(a_i-c)^2}{4b_i} - b_i(K_i^{fe})^2\right] \mathbf{d}F_i + \int\limits_{2b_iK_i^{ee}+c}^{\infty} \left[b_i(K_i^{ee})^2 - b_i(K_i^{fe})^2\right] \mathbf{d}F_i$$
$$\geq 0$$

Therefore, we have $\kappa_i = e$ given that $\kappa_j = e$.

Otherwise, if $\kappa_j = f$, we have

$$\Pi_i^{ef} \geq \Pi_i^{ef}\left(K_i^{ff},\ K_j^{fe}\right) \geq \Pi_i^{ff}\left(K_i^{ff},\ K_j^{fe}\right) \geq \Pi_i^{ff}$$

The first inequality follows that $K_i^{ef} = \text{argmax}_{K_i \geq 0}\left\{\Pi_i^{ef}(K_i, K_j^{fe})\right\}$. The second inequality follows the fact that $\Pi_i^{ef}(K_i, K_j) = \Pi_i^{ff}(K_i, K_j) + (w_f - w_e)K_i \geq \Pi_i^{ff}(K_i, K_j)$. The third inequality follows Lemma C.1. Therefore, we have $\kappa_i = e$ given that $\kappa_j = f$. Combining the two cases, it is dominant that $\kappa_i = e$ when $w_e \leq w_f$.

Part (ii) $w_e > w_f$. If $\kappa_j = e$, focusing on the non-trivial case of $K_i^{fe} > 0$ and $K_i^{ee} > 0$, we have

$$\Pi_i^{ee} - \Pi_i^{fe} = \int\limits_{2b_iK_i^{ee}+c}^{2b_iK_i^{fe}+c} \left[b_i(K_i^{ee})^2 - \frac{(a_i-c)^2}{4b_i}\right] \mathbf{d}F_i + \int\limits_{2b_iK_i^{fe}+c}^{\infty} \left[b_i(K_i^{ee})^2 - b_i(K_i^{fe})^2\right] \mathbf{d}F_i$$
$$\leq 0$$

Therefore, $\kappa_i = f$ given $\kappa_j = e$.

If $\kappa_j = f$, for given w_f, we first show that $\Pi_i^{ef}(w_e, w_f)$ decreases in w_e. Notice K_j^{fe} and K_j^{ff} do not change with respect to w_e, so we apply envelope theorem and obtain that

$$\frac{\partial \Pi_i^{ef}(w_e, w_f)}{\partial w_e} = -K_i^{ef} \leq 0$$

Observing that $\Pi_i^{ff}(w_f, w_f)$ does not change with respect to w_e, we define

$$\bar{w}_e{}^I(w_f) \triangleq \max\left\{\inf\left\{w : \Pi_i^{ef}(w, w_f) \leq \Pi_i^{ff}(w_f, w_f)\right\}, w_f\right\}$$

Combining with the analysis of the previous case, we conclude when $w_f < w_e \leq \bar{w}_e{}^I(w_f)$, we have $(\kappa_i, \kappa_j) = (e, f)$; when $w_e > \bar{w}_e{}^I(w_f)$, we have $(\kappa_i, \kappa_j) = (f, f)$. $\square$

Proof of Corollary IV.4. Part (i) $w_e \leq w_f$. First, when $w_e = w_f = w$, we have $\Pi_i^{ee} \leq \Pi_i^{ff}$, because

$$\Pi_i^{ee} = \Pi_i^{ee}\left(K_i^{fe}, K_j^{ff}\right) \leq \Pi_i^{ff}\left(K_i^{fe}, K_j^{ff}\right) \leq \Pi_i^{ff}$$

The first equality follows (1) $K_i^{ee} = K_i^{fe}$ because $\frac{\partial V_i^{fe}}{\partial K_i} = \frac{\partial V_i^{ee}}{\partial K_i}$ when $w_e = w_f$; (2)

$\frac{\partial V_i^{\kappa_i e}}{\partial K_j} = 0$. The next inequality follows that when $w_e = w_f$, we have

$$
\begin{aligned}
&\Pi_i^{ff}(K_i, K_j) - \Pi_i^{ee}(K_i, K_j) \\
&= \int_0^c \Bigg[\int_{2b_i(K_i+K_j)+c}^{\infty} [[a_i - c - b_i(K_i + K_j)](K_i + K_j) - (a_i - c - b_iK_i)K_i] \,\mathbf{d}F_i \\
&+ \int_{2b_iK_i+c}^{2b_i(K_i+K_j)+c} \left[\frac{(a_i - c)^2}{4b_i} - (a_i - c - b_iK_i)K_i \right] \mathbf{d}F_i \Bigg] \,\mathbf{d}F_j \\
&+ \int_c^{2b_jK_j+c} \Bigg[\int_{2b_i(K_i+K_j)+c-\frac{b_i(a_j-c)}{b_j}}^{\infty} \Bigg[[a_i - c - b_i(K_i + K_j - \frac{a_j - c}{2b_j})](K_i + K_j - \frac{a_j - c}{2b_j}) \\
&\qquad - (a_i - c - b_iK_i)K_i \Bigg] \mathbf{d}F_i \\
&\qquad + \int_{2b_iK_i+c}^{2b_i(K_i+K_j)+c-\frac{b_i(a_j-c)}{b_j}} \left[\frac{(a_i - c)^2}{4b_i} - (a_i - c - b_iK_i)K_i \right] \mathbf{d}F_i \Bigg] \,\mathbf{d}F_j \geq 0
\end{aligned}
$$

The last inequality follows $K_i^{ff} = \operatorname{argmax}_{K_i \geq 0} \left\{ \Pi_i^{ff}(K_i, K_j^{ff}) \right\}$.

Second, using envelope theorem, we have $\frac{\partial \Pi_i^{ee}(w_e, w_e)}{\partial w_e} = -K_i^{ee} \leq 0$ and $\frac{\partial \Pi_i^{ff}(w_f, w_f)}{\partial w_e} = 0$. Therefore, define

$$
\underline{w_e}^I(w_f) \triangleq \max \left\{ \inf \left\{ w : \Pi_i^{ee}(w, w) \leq \Pi_i^{ff}(w_f, w_f) \right\}, 0 \right\}
$$

So we conclude a prisoner's dilemma occurs, in which $(\kappa_i, \kappa_j) = (e, e)$ by Theorem IV.3 but $\Pi_i^{ee} \leq \Pi_i^{ff}$, when $\underline{w_e}^I(w_f) \leq w_e \leq w_f$.

Part(ii) follows Theorem IV.3 and the definition of a free-rider equilibrium. $\square$

While the analysis for the Cournot market follows the same general structure as the independent market, it is significantly more complex because of an explosion in the number of cases to consider: e.g. in the third stage, depending on market demand, each firm can order its unconstrained quantity, or quantity constrained only by its own capacity, or quantity constrained by its capacity and available leftover capacity from the other firm. To help navigate this, we first define some thresholds on the market signal a, which we will use for all the proofs in this section. Let $\lambda_1^1 \triangleq 3bK_1 + 2c_1 - c_2$, $\lambda_2^1 \triangleq bK_1 + 2bK_2 + c_2$, $\lambda_1^2 \triangleq 3bK_2 + 2c_2 - c_1$, $\lambda_2^2 \triangleq 2bK_1 + bK_2 + c_1$, $\lambda_3^2 \triangleq 2bK_2 + c_2$, $\lambda_4^2 \triangleq bK_2 + c_1$, $\lambda_5^1 \triangleq b(K_1 + K_2) + c_1$, $\lambda_5^2 \triangleq 2b(K_1 + K_2) + c_2$, $\lambda_6^1 = \frac{3b(K_1+K_2)+c_1+c_2}{2}$, and $K_T \triangleq K_1 + K_2$. The lemma as stated in the body of the chapter showed equilibrium order quantities for the case $(\kappa_1, \kappa_2) = (e, f)$; we first provide the corresponding expressions for other types of capacity investments.

Lemma C.3 (Lemma IV.5 continued)**.** *Suppose that firm 1 invests in K_1 units of exclusive capacity and firm 2 in K_2 units of exclusive capacity. The equilibrium order quantities for given demand signal $\theta = a$ are:*

Conditions	$(q_1^{ee},\ q_2^{ee})$
$\max\{\lambda_2^1, \lambda_2^2\} \le a$	(K_1, K_2)
$\max\{\lambda_1^2, \lambda_4^2\} \le a < \lambda_2^2$	$(\frac{a-c_1-bK_2}{2b}, K_2)$
$\lambda_3^2 \le a < \lambda_4^2$	$(0, K_2)$
$\lambda_1^1 \le a < \lambda_2^1$	$(K_1,\ \frac{a-c_2-bK_1}{2b})$
$2c_1 - c_2 \le a < \min\{\lambda_1^2, \lambda_1^1\}$	$(\frac{a-2c_1+c_2}{3b}, \frac{a-2c_2+c_1}{3b})$
$a < \min\{\lambda_3^2, 2c_1 - c_2\}$	$(0,\ \frac{a-c_2}{2b})$

Suppose that firm 1 invests in K_1 units of first-priority capacity and firm 2 in K_2 units of exclusive capacity. The equilibrium order quantities for given demand signal $\theta = a$ are:

Conditions	(q_1^{fe}, q_2^{ef})
$\max\{\lambda_2^1, \lambda_2^2\} \leq a$	(K_1, K_2)
$\max\{\lambda_6^1, \lambda_5^1\} \leq a < \lambda_2^2$	$(\frac{a-c_1}{b} - K_T, 2K_T - \frac{a-c_1}{b})$
$\lambda_5^2 \leq a < \lambda_5^1$	$(0, K_T)$
$\lambda_1^1 \leq a < \lambda_2^1$	$(K_1, \frac{a-c_2-bK_1}{2b})$
$2c_1 - c_2 \leq a < \min\{\lambda_6^1, \lambda_1^1\}$	$(\frac{a-2c_1+c_2}{3b}, \frac{a-2c_2+c_1}{3b})$
$a < \min\{2c_1 - c_2, \lambda_5^2\}$	$(0, \frac{a-c_2}{2b})$

Suppose that firm 1 invests in K_1 units of first-priority capacity and firm 2 in K_2 units of first-priority capacity. The equilibrium order quantities for given demand signal $\theta = a$ are:

Conditions	(q_1^{ff}, q_2^{ff})
$\max\{\lambda_2^1, \lambda_2^2\} \leq a$	(K_1, K_2)
$\max\{\lambda_6^1, \lambda_5^1\} \leq a < \lambda_2^2$	$(\frac{a-c_1}{b} - K_T,\ 2K_T - \frac{a-c_1}{b})$
$\lambda_5^2 \leq a < \lambda_5^1$	$(0, K_T)$
$\lambda_6^1 \leq a < \lambda_2^1$	$(2K_T - \frac{a-c_2}{b},\ \frac{a-c_2}{b} - K_T)$
$2c_1 - c_2 \leq a < \lambda_6^1$	$(\frac{a-2c_1+c_2}{3b},\ \frac{a-2c_2+c_1}{3b})$
$a < \min\{2c_1 - c_2, \lambda_5^2\}$	$(0,\ \frac{a-c_2}{2b})$

The expressions can be derived fairly straightforwardly from the definition of the game.

Firms' expected profits The firms' expected profits are a function of capacity type choices (κ_i, κ_j) and capacity levels (K_i, K_j) from Lemma IV.5. Depending on the realized demand signal, firms may have different equilibrium order quantities, and how the equilibrium quantities shift with respect to the realized demand signal depends on the capacity levels K_1 and K_2. Therefore, we classify firms' expected profits based on the value of the two capacity levels. For the interest of space, we only show expressions for $(\kappa_1, \kappa_2) = (e, e)$.

If $K_1 \le K_2 + \frac{c_2}{b} - \frac{c_1}{b}$, we have $\lambda_1^1 \le \lambda_2^1$, and

$$V_1^{ee} = \int\limits_{2c_1-c_2}^{\lambda_1^1} \frac{(a-2c_1+c_2)^2}{9b}\mathbf{d}F + \int\limits_{\lambda_1^1}^{\lambda_2^1} \frac{(a-2c_1+c_2-bK_1)K_1}{2}\mathbf{d}F$$
$$+ \int\limits_{\lambda_2^1}^{\infty}(a-c_1-bK_T)K_1\mathbf{d}F - w_eK_1$$

$$V_2^{ee} = \int\limits_{c_2}^{2c_1-c_2} \frac{(a-c_2)^2}{4b}\mathbf{d}F + \int\limits_{2c_1-c_2}^{\lambda_1^1} \frac{(a-2c_2+c_1)^2}{9b}\mathbf{d}F + \int\limits_{\lambda_1^1}^{\lambda_2^1} \frac{(a-c_2-bK_1)^2}{4b}\mathbf{d}F$$
$$+ \int\limits_{\lambda_2^1}^{\infty}(a-c_2-bK_T)K_2\mathbf{d}F - w_eK_2$$

Two other cases are possible. If $K_1 > K_2 + \frac{c_2}{b} - \frac{c_1}{b} > 0$, we have $\lambda_1^2 \le \lambda_2^2$ and (V_1^{ee}, V_2^{ee}) can be found by changing the integrands and limits above appropriately, and likewise for the remaining case where $K_2 + \frac{c_2}{b} - \frac{c_1}{b} \le 0$ (where we have $\lambda_3^2 \le \lambda_4^2 \le \lambda_2^2$).

Observe that these profit functions are continuous and differentiable with respect to (K_1, K_2). We omit the detailed expressions for the derivatives here for space.

Proof of Lemma IV.6. Part (i) $w_e = w_f = w$. We first establish that in an equilibrium, it cannot be the case that $K_1^{\kappa_1\kappa_2} \neq (K_2^{\kappa_2\kappa_1} + \frac{c_2}{b} - \frac{c_1}{b})^+$ in Step 1 and then show a strategy satisfying the conditions in the lemma is indeed an equilibrium in Step 2. For brevity we only show a complete analysis for the case $(\kappa_1, \kappa_2) = (e, e)$.

Step 1: We first consider the case where $K_1 < K_2 + \frac{c_2}{b} - \frac{c_1}{b}$ and $K_2 + \frac{c_2}{b} - \frac{c_1}{b} > 0$. Let the support of the demand signal θ be $[a_L,\ a_H]$. If $a_H < \lambda_1^1$, it is trivial that both firms have incentive to decrease the capacity, so the (K_1, K_2) cannot be an equilibrium. If $\lambda_1^1 < a_H$, we have

$$\frac{\partial V_1^{ee}}{\partial K_1} - \frac{\partial V_2^{ee}}{\partial K_2} = \int_{\lambda_1^1}^{\lambda_2^1} \frac{(a - 2c_1 + c_2 - 2bK_1)}{2} \mathbf{d}F + \int_{\lambda_2^1}^{\infty} (bK_2 + c_2 - bK_1 - c_1) \mathbf{d}F > 0$$

In this case, we have $K_2 \geq K_1 + \frac{c_1}{b} - \frac{c_2}{b} > 0$, so it cannot be the case where $K_2 = 0$, $\frac{\partial V_1^{ee}}{\partial K_1} = 0$, and $\frac{\partial V_2^{ee}}{\partial K_2} \leq 0$. Therefore at least one firm has incentive to deviate. Following a similar argument, one can show that when $K_1 \geq K_2 + \frac{c_2}{b} - \frac{c_1}{b}$ and $K_2 + \frac{c_2}{b} - \frac{c_1}{b} > 0$, or $K_2 + \frac{c_2}{b} - \frac{c_1}{b} \leq 0$, the $(K_1,\ K_2)$ where $K_1 \neq (K_2 + \frac{c_2}{b} - \frac{c_1}{b})^+$ cannot be an equilibrium.

Step 2: We prove this by showing that at (K_1, K_2) where $K_1 = (K_2 + \frac{c_2}{b} - \frac{c_1}{b})^+$, neither firm has incentive to deviate. We focus on the non-trivial case where $K_1 > 0$, and present the analysis for firm 1. The analysis for firm 2 is similar. If firm 1 deviates to $K_1 - \epsilon$ with $\epsilon > 0$, then

$$\begin{aligned}
&\frac{\partial V_1^{ee}}{\partial K_1}(K_1 - \epsilon, K_2) - \frac{\partial V_1^{ee}}{\partial K_1}(K_1, K_2) \\
&= \int_{\lambda_1^1 - 3b\epsilon}^{\lambda_2^1 - b\epsilon} \frac{a - 2c_1 + c_2 - 2bK_1 + 2b\epsilon}{2} \mathbf{d}F \\
&\quad + \int_{\lambda_2^1 - b\epsilon}^{\lambda_2^1} [a - c_1 - b(2K_1 - 2\epsilon + K_2)] \mathbf{d}F + \int_{\lambda_2^1}^{\infty} 2b\epsilon \mathbf{d}F \geq 0
\end{aligned}$$

So firm 1 has incentive to increase capacity. On the other hand, if firm 1 deviates to $K_1 + \epsilon$, we have

$$\begin{aligned}
&\frac{\partial V_1^{ee}}{\partial K_1}(K_1 + \epsilon,\ K_2) - \frac{\partial V_1^{ee}}{\partial K_1}(K_1,\ K_2) \\
&= -\int_{\lambda_2^2}^{\lambda_2^2 + 2b\epsilon} [a - c_1 - b(2K_1 + K_2)] \mathbf{d}F + \int_{\lambda_2^2 + 2b\epsilon}^{\infty} -2b\epsilon \mathbf{d}F \leq 0
\end{aligned}$$

So firm 1 has incentive to decrease capacity.

Following a similar argument, we show the results hold for (f, f), (e, f) and (f, e) (omitted for space).

Part (ii) $w_e > w_f$. We first analyze the case when $\kappa_1 = \kappa_2$. Referring to part (i), we have

$$K_2^{ff} = \left\{ K : \int\limits_{2bK+c_2+(bK+c_2-c_1)^+}^{\infty} \left[a - c_2 - 2bK - (bK + c_2 - c_1)^+ \right] \mathbf{d}F - w_f = 0 \right\}$$

and

$$K_2^{ee} = \left\{ K : \int\limits_{2bK+c_2+(bK+c_2-c_1)^+}^{\infty} \left[a - c_2 - 2bK - (bK + c_2 - c_1)^+ \right] \mathbf{d}F - w_e = 0 \right\}$$

Because $\int_{2bK+c_2+(bK+c_2-c_1)^+}^{\infty} [a - c_2 - 2bK - (bK + c_2 - c_1)^+]\mathbf{d}F$ decreases in K, and $K_1^{\kappa_1\kappa_2} = (K_2^{\kappa_2\kappa_1} + \frac{c_2}{b} - \frac{c_1}{b})^+$, it follows that if $w_e > w_f$, $K_i^{ee} \leq K_i^{ff}$.

We next prove the case of $(\kappa_1, \kappa_2) = (e, f)$ by contradiction, and the proof for the other case is similar. Assume $K_1 > (K_2 + \frac{c_2}{b} - \frac{c_1}{b})^+$, when $K_2 + \frac{c_2}{b} - \frac{c_1}{b} \geq 0$, we have

$$\begin{aligned} &\frac{\partial V_1^{ef}}{\partial K_1} - \frac{\partial V_2^{fe}}{\partial K_2} \\ &= -\int\limits_{\lambda_1^2}^{\lambda_2^2} \frac{a - 2c_2 + c_1 - 2bK_2}{2} \mathbf{d}F + \int\limits_{\lambda_2^2}^{\infty} (c_2 + bK_2 - c_1 - bK_1) \mathbf{d}F - w_e + w_f \\ &\leq -w_e + w_f < 0 \end{aligned}$$

When $K_2 + \frac{c_2}{b} - \frac{c_1}{b} < 0$, we have

$$\frac{\partial V_1^{ef}}{\partial K_1} - \frac{\partial V_2^{fe}}{\partial K_2} = -\int_{\lambda_3^2}^{\lambda_4^2} (a - c_2 - 2bK_2)\mathbf{d}F - \int_{\lambda_4^2}^{\lambda_2^2} \frac{a - 2c_2 + c_1 - 2bK_2}{2}\mathbf{d}F$$
$$+ \int_{\lambda_2^2}^{\infty} (c_2 + bK_2 - c_1 - bK_1)\mathbf{d}F - w_e + w_f \leq -w_e + w_f < 0$$

Therefore, at least one of the two firms has incentive to deviate, and in equilibrium $K_1^{ef} \leq (K_2^{fe} + \frac{c_2}{b} - \frac{c_1}{b})^+$. Similarly, it can be shown that if $w_e > w_f$, in equilibrium $K_1^{fe} \geq (K_2^{ef} + \frac{c_2}{b} - \frac{c_1}{b})^+$.

Part (iii) $w_e < w_f$. The proof is similar to the proof of part (ii), and is omitted for space. □

Proof of Theorem IV.7. Part (i) $w_e > w_f$. We show the results in three steps: for any $(\bar{K}_i, \bar{K}_j)$ such that $\frac{\partial V_i^{ef}}{\partial K_i}(\bar{K_i}^{ef}, \bar{K_j}^{fe}) = 0$ and $\frac{\partial V_j^{fe}}{\partial K_j}(\bar{K_j}^{fe}, \bar{K_i}^{ef}) = 0$, (1)$K_i^{ff} \geq \bar{K_i}^{ef}$ and $\bar{K_i^{fe}} \geq K_i^{ee}$; (2) $\Pi_i^{ff} \geq \Pi_i^{ef}(\bar{K_i}^{ef}, \bar{K_j}^{fe})$; (3) $\Pi_i^{fe}(\bar{K_i}^{fe}, \bar{K_j}^{ef}) \geq \Pi_i^{ee}$. Notice that following the same proof of Lemma IV.6(ii), we can show that $\bar{K_1}^{ef} \leq (\bar{K_2}^{fe} + \frac{c_2}{b} - \frac{c_1}{b})^+$, and $\bar{K_1}^{fe} \geq (\bar{K_2}^{ef} + \frac{c_2}{b} - \frac{c_1}{b})^+$. For the interest of space, we focus on the non-trivial case where $K_1^{\kappa_1\kappa_2} > 0$, $\bar{K_1}^{\kappa_1\kappa_2} > 0$, $K_2{}^{\kappa_2\kappa_1} + \frac{c_2}{b} - \frac{c_1}{b} > 0$ and $\bar{K_2}^{\kappa_2\kappa_1} + \frac{c_2}{b} - \frac{c_1}{b} > 0$.

From Lemma IV.6, we have

$$K_2^{ff} = \left\{ K : \int_{3bK+2c_2-c_1}^{\infty} (a - 3bK - 2c_2 + c_1)\mathbf{d}F - w_f = 0 \right\}$$

By the fact that $\frac{\partial V_2^{fe}}{\partial K_2}(\bar{K_2}^{fe}, \bar{K_1}^{ef}) = 0$, we get

$$\frac{\partial V_2^{fe}}{\partial K_2}(\bar{K_2}^{fe}, \bar{K_1}^{ef}) = \int\limits_{\frac{3b(\bar{K_2}^{fe}+\bar{K_1}^{ef})+c_1+c_2}{2}}^{b\bar{K_1}^{ef}+2b\bar{K_2}^{fe}+c_2} \left[2b(\bar{K_2}^{fe} + \bar{K_1}^{ef}) - 2(a - c_2)\right] \mathbf{d}F$$

$$+ \int\limits_{b\bar{K_1}^{ef}+2b\bar{K_2}^{fe}+c_2}^{\infty} \left[a - c - b(\bar{K_1}^{ef} + 2\bar{K_2}^{fe})\right] \mathbf{d}F - w_f = 0$$

Notice if $K_1 + 2K_2 > 3K_2^{ff} + \frac{c_2}{b} - \frac{c_1}{b}$, we have

$$\begin{aligned}
\frac{\partial V_2^{fe}}{\partial K_2} &= \int\limits_{\frac{3b(K_1+K_2)+c_1+c_2}{2}}^{bK_1+2bK_2+c_2} [2b(K_1 + K_2) - 2(a - c_2)] \mathbf{d}F \\
&\quad + \int\limits_{bK_1+2bK_2+c_2}^{\infty} [a - c_2 - b(K_1 + 2K_2)] \mathbf{d}F - w_f \\
&\leq \int\limits_{\frac{3b(K_1+K_2)+c_1+c_2}{2}}^{bK_1+2bK_2+c_2} [2b(K_1 + K_2) - 2(a - c_2)] \mathbf{d}F \\
&\quad + \int\limits_{3bK_2^{ff}+2c_2-c_1}^{\infty} \left(a - 3bK_2^{ff} - 2c_2 + c_1\right) \mathbf{d}F - w_f \\
&= \int\limits_{\frac{3b(K_1+K_2)+c_1+c_2}{2}}^{bK_1+2bK_2+c_2} [2b(K_1 + K_2) - 2(a - c_2)] \mathbf{d}F \leq 0
\end{aligned}$$

Then firm 2 has incentive to decrease the capacity level. Therefore, $\bar{K_1}^{ef} + 2\bar{K_2}^{fe} \leq 3K_2^{ff} + \frac{c_2}{b} - \frac{c_1}{b}$, and we have $\bar{K_1}^{ef} + 2(\bar{K_1}^{ef} - \frac{c_2}{b} + \frac{c_1}{b}) \leq 3K_2^{ff} + \frac{c_2}{b} - \frac{c_1}{b}$, so we obtain $3\bar{K_1}^{ef} \leq 3(K_2^{ff} + \frac{c_2}{b} - \frac{c_1}{b}) = 3K_1^{ff}$. Therefore, we have shown that $K_1^{ff} \geq \bar{K_1}^{ef}$. Following a similar analysis we show that $K_2^{ff} \geq K_2^{ef}$ (omitted for space.)

Similarly, by the fact that $\frac{\partial V_1^{ef}}{\partial K_1}(\bar{K_1}^{ef}, \bar{K_2}^{fe}) = 0$ and Lemma IV.6(i), we have $b\bar{K_1}^{ef} + 2b\bar{K_2}^{fe} \geq bK_1^{ee} + 2bK_2^{ee}$. Together with the fact that $\bar{K_1}^{ef} \leq \bar{K_2}^{fe} + \frac{c_2}{b} - \frac{c_1}{b}$, we have $\bar{K_2}^{fe} \geq K_2^{ee}$. Again, we show that $\bar{K_1}^{fe} \geq K_1^{ee}$ following a similar analysis.

Next we show that $\Pi_1^{ff} \geq \Pi_1^{ef}(\bar{K_1}^{ef}, \bar{K_2}^{fe})$, and the analysis for firm 2 is similar. Focusing on the non-trivial case where $K_1^{ff} > 0$ and $\bar{K_1}^{ef} > 0$, after simplifying the expressions, we have the profit difference

$$
\begin{aligned}
&\Pi_1^{ff} - \Pi_1^{ef}(\bar{K_1}^{ef}, \bar{K_2}^{fe}) \\
&= \int\limits_{\frac{3b(\bar{K_1}^{ef}+\bar{K_2}^{fe})+c_1+c_2}{2}}^{b\bar{K_1}^{ef}+2b\bar{K_2}^{fe}+c_2} \left[\frac{(a-2c_1+c_2)^2}{9b} - 2b(\bar{K_1}^{ef})^2 + 2b(\bar{K_2}^{fe})^2 \right. \\
&\qquad\qquad \left. -(3a-c_2-2c_1)\bar{K_2}^{fe} + \frac{(a-c_1)(a-c_2)}{b}\right] \mathbf{d}F \\
&+ \int\limits_{b\bar{K_1}^{ef}+2b\bar{K_2}^{fe}+c_2}^{bK_1^{ff}+2bK_2^{ff}+c_2} \left[\frac{(a-2c_1+c_2)^2}{9b} - b(\bar{K_1}^{ef})^2\right] \mathbf{d}F \\
&+ \int\limits_{bK_1^{ff}+2bK_2^{ff}+c_2}^{\infty} \left[b(K_1^{ff})^2 - b(\bar{K_1}^{ef})^2\right] \mathbf{d}F \geq 0
\end{aligned}
$$

That is, $\Pi_1^{ff} \geq \Pi_1^{ef}(\bar{K_1}^{ef}, \bar{K_2}^{fe})$. Similarly, we show that $\Pi_2^{ff} \geq \Pi_2^{ef}(\bar{K_2}^{ef}, \bar{K_1}^{fe})$ (omitted for space.)

Finally we show that $\Pi_2^{fe}(\bar{K_2}^{fe}, \bar{K_1}^{ef}) \geq \Pi_2^{ee}$, and the analysis for firm 1 is similar.

Again, leveraging that $\frac{\partial V_2^{fe}}{K_2}(\bar{K_2}^{fe}, \bar{K_1}^{ef}) = 0$ and $\frac{\partial V_2^{ee}}{K_2}(K_2^{ee}, K_1^{ee}) = 0$, we have

$$
\begin{aligned}
&\Pi_2^{fe}(\bar{K_2}^{fe}, \bar{K_1}^{ef}) - \Pi_2^{ee} \\
&= \left[\int_{c_2}^{2c_1 - c_2} \frac{(a-c_2)^2}{4b} \mathbf{d}F + \int_{2c_1-c_2}^{\frac{3b(\bar{K_1}^{ef} + \bar{K_2}^{fe}) + c_1 + c_2}{2}} \frac{(a - 2c_2 + c_1)^2}{9b} \mathbf{d}F \right. \\
&\left. + \int_{b\bar{K_1}^{ef} + 2b\bar{K_2}^{fe} + c_2}^{\infty} b(\bar{K_2}^{fe})^2 \mathbf{d}F + \int_{\frac{3b(\bar{K_1}^{ef} + \bar{K_2}^{fe}) + c_1 + c_2}{2}}^{b\bar{K_1}^{ef} + 2b\bar{K_2}^{fe} + c_2} \frac{(a - c_2 - b\bar{K_1}^{ef})^2 - (b\bar{K_2}^{fe})^2}{b} \mathbf{d}F \right] \\
&- \left[\int_{c_2}^{2c_1-c_2} \frac{(a-c_2)^2}{4b} \mathbf{d}F + \int_{2c_1-c_2}^{3bK_1^{ee} + 2c_1 - c_2} \frac{(a - 2c_2 + c_1)^2}{9b} \mathbf{d}F + \int_{3bK_1^{ee} + 2c_1 - c_2}^{\infty} b(K_2^{ee})^2 \right] \\
&\triangleq \left[\int_{c_2}^{\infty} l_2^{fe}(a) \mathbf{d}F \right] - \left[\int_{c_2}^{\infty} l_2^{ee}(a) \mathbf{d}F \right]
\end{aligned}
$$

Following that $\bar{K_2}^{fe} \geq K_2^{ee}$ and Lemma IV.6, we have $l_2^{fe}(a) \geq l_2^{ee}(a)$ for any $a \geq 0$. It follows that $\Pi_2^{fe}(\bar{K_2}^{fe}, \bar{K_1}^{ef}) \geq \Pi_2^{ee}$. Similarly, we show $\Pi_1^{fe}(\bar{K_1}^{fe}, \bar{K_2}^{ef}) \geq \Pi_1^{ee}$ (omitted for space).

Therefore, when $w_e > w_f$, it is dominant that $\kappa_i = f$.

Part (ii) $w_w = w_f$. The results follow from Lemma IV.6, and the fact that when $K_1^{\kappa_1\kappa_2} = (K_2^{\kappa_2\kappa_1} + \frac{c_2}{b} - \frac{c_1}{b})^+$, firm i will not access firm j's leftover for any realization of θ.

Part (iii) $w_e < w_f$. The proof is similar to part (i), and the details are omitted for space. □

Proof of Corollary IV.8. This directly follows Theorem IV.7, and the fact that if $w_e \leq w_f$, we have $\Pi_i^{ee} \geq \Pi_i^{ff}$; otherwise, $\Pi_i^{ee} \leq \Pi_i^{ff}$. □

Proof of Theorem IV.9. We first derive the supplier's production profit Π_{sp}^{ee}. By equation (C.1), we have that the equilibrium capacity $K_i^{\kappa_i\kappa_j}$ satisfies

$$\int\limits_{2b_iK_i^{ee}+c}^{\infty} (a_i - c - 2b_iK_i^{ee})\mathbf{d}F_i - w_e = 0$$

Deriving from this equation, we obtain that

$$\int\limits_{2b_iK_i^{ee}+c}^{\infty} K_i^{ee}\mathbf{d}F_i = \int\limits_{2b_iK_i^{ee}+c}^{\infty} \frac{a_i - c}{2b_i}\mathbf{d}F_i - \frac{w_{\kappa_i}}{2b_i} \tag{C.3}$$

By Lemma IV.1, equation (4.9) and equation (C.3), we have the supplier's production profit as

$$\begin{aligned}
\Pi_{sp}^{ee} =& c\sum_{i=1}^{2}\left\{\int\limits_{c}^{2bK_i^{ee}+c} \frac{a_i - c}{2b_i}\mathbf{d}F_i + \int\limits_{2b_iK_i^{ee}+c}^{\infty} K_i^{ee}\mathbf{d}F_i\right\} \\
=& c\sum_{i=1}^{2}\left\{\int\limits_{c}^{2b_iK_i^{ee}+c} \frac{a_i - c}{2b_i}\mathbf{d}F_i + \int\limits_{2b_iK_i^{ee}+c}^{\infty} \frac{a_i - c}{2b_i}\mathbf{d}F_i + \frac{w_{\kappa_i}}{2b_i}\right\} \\
=& c\sum_{i=1}^{2}\left\{\int\limits_{c}^{\infty} \frac{a_i - c}{2b_i}\mathbf{d}F_i + \frac{w_{\kappa_i}}{2b_i}\right\}
\end{aligned}$$

The expressions for the (e, f), (f, f), and (f, f) case can be obtained similarly.

The expressions for supplier's profit from capacity investment follow from the definition. □

Proof of Corollary IV.10. The proof directly follows the expression of $\Pi_{sp}^{\kappa_i\kappa_j}$ as shown in Theorem IV.9. □

Proof of Theorem IV.11. Part (i) (Independent market). Following Theorem IV.9, we have the profit from production $\Pi_{sp}^{\kappa_i\kappa_j}$ remains the same for any (κ_i, κ_j), when $w_e = w_f = w$. However, following Lemma IV.2, we have $K_i^{ee} = K_i^{fe} \geq K_i^{ff}$ and also $K_i^{ee} \geq K_i^{ef}$. Therefore, the profit from capacity investment $\Pi_{sc}^{\kappa_1\kappa_2} = \gamma(wK_1^{\kappa_1\kappa_2} +$

$wK_2^{\kappa_2\kappa_1}$) is the highest when $(\kappa_i, \kappa_j) = (e, e)$. Therefore, the supplier always prefers to offer the exclusive capacity.

Part (ii) (Cournot market). This directly follows Theorem IV.7, and the observation that under the equilibrium capacity specified in Lemma IV.6(i), firm i will never access firm j's capacity (also see Lemma IV.5). $\square$

BIBLIOGRAPHY

BIBLIOGRAPHY

Arya, A., B. Mittendorf, D. E. M. Sappington. 2008. The make-or-buy decision in the presence of a rival: Strategic outsourcing to a common supplier. *Management Science* **54**(10) 1747–1758.

Aston, A. 2009. Wal-Mart: Making its suppliers go green. *Bloomberg Businessweek* http://www.businessweek.com/magazine/content/09_21/b4132044814736.htm. Retrieved October 31, 2011.

Aviv, Y., A. Pazgal. 2005. A partially observed markov decision process for dynamic pricing. *Management Science* **51**(9) 1400–1416.

Azoury, K. S. 1985. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science* **31**(9) 1150–1160.

Babich, V. 2010. Independence of capacity ordering and financial subsidies to risky suppliers. *Manufacturing & Service Operations Management* **12**(4) 583–607.

Besanko, D., U. Doraszelski, L. X. Lu, M. Satterthwaite. 2010. Lumpy capacity investment and disinvestment dynamics. *Operations Research* **58**(4-Part-2) 1178–1193.

Besbes, O., A. Muharremoglu. 2013. On implications of demand censoring in the newsvendor problem. *Management Sci.* **59**(6) 1407–1424.

Besbes, O., A. Zeevi. 2009. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research* **57**(6) 1407–1420.

Boyacı, T., Ö. Özer. 2010. Information acquisition for capacity planning via pricing and advance selling: When to stop and act? *Operations Research* **58**(5) 1328–1349.

Burnetas, A., S. Gilbert. 2001. Future capacity procurements under unknown demand and increasing costs. *Management Science* **47**(7) 979–992.

Burnetas, A. N., C. E. Smith. 2000. Adaptive ordering and pricing for perishable products. *Operations Research* **48**(3) 436–443.

Cachon, G. P., P. T. Harker. 2002. Competition and outsourcing with scale economies. *Management Science* **48**(10) 1314–1333.

Cachon, G. P., M. A. Lariviere. 1999. Capacity choice and allocation: Strategic behavior and supply chain performance. *Management Science* **45**(8) 1091–1108.

Cachon, G. P., M. A. Lariviere. 2005. Supply chain coordination with revenue-sharing contracts: Strengths and limitations. *Management Science* **51**(1) 30–44.

Chao, X., H. Chen, S. Zheng. 2009. Dynamic capacity expansion for a service firm with capacity deterioration and supply uncertainty. *Operations Research* **57**(1) 82–93.

Chen, L., E. L. Plambeck. 2008. Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing & Service Operations Management* **10**(2) 236–256.

Davis, J., B. Crothers. 1999. Apple invests $100 million in Samsung. *CNET News* Http://news.cnet.com/Apple-invests-100-million-in-Samsung/2100-1040_3-229154.html, Retrieved Oct. 31, 2011.

Davis, M. H. A., M. A. H. Dempster, S. P. Sethi, D. Vermes. 1987. Optimal capacity expansion under uncertainty. *Advances in Applied Probability* **19**(1) 156–176.

Dignan, L. 2012. Supply chain wars: Hon Hai's Sharp investment helps Apple vs. Samsung. *ZDNet* http://www.zdnet.com/blog/btl/supply-chain-wars-hon-hais-sharp-investment-helps-apple-vs-samsung/72577. Retrieved May 20, 2012.

Dixit, A. 1980. The role of investment in entry-deterrence. *The Economic Journal* **90**(357) 95–106.

Dixit, A. K., R. S. Pindyck. 1994. *Investment under uncertainty*. Princeton University Press, Princeton, NJ.

Eberly, J. C., J. A. Van Mieghem. 1997. Multi-factor dynamic investment under uncertainty. *Journal of Economics Theory* **75**(2) 345–387.

Economist, *The.* 1996. Asian semiconductors: Saving chips from market dips. (January 20) 63.

Eppen, G. D., A. V. Iyer. 1997. Improved fashion buying with bayesian updates. *Operations Research* **45**(6) 805–819.

Federgruen, A., N. Yang. 2009. Competition under generalized attraction models: Applications to quality competition under yield uncertainty. *Management Science* **55**(12) 2028–2043.

Feng, Q., L. X. Lu. 2012. The strategic perils of low cost outsourcing. *Management Science* **58**(6) 1196–1210.

Ferrari, B. 2011. Apple's secret sauce revisited: Strategically investing in supply chain capabilities. *Supply Chain Matters* Http://www.theferrarigroup.com/supply-chain-matters/2011/04/27/apples-secret-sauce-revisited-strategically-investing-in-supply-chain-capabilities/, Retrieved Oct. 31, 2011.

Ford Motor Company. 2012. Profitable growth for all: Ford Motor Company 2012 annual report http://corporate.ford.com/doc/ar2012-2012%20Annual%20Report.pdf. Retrieved June 1, 2013.

Freidenfelds, J. 1981. *Capacity expansion: analysis of simple models with applications*. Elsevier North Holland, New York.

Gallego, G. 1992. A minmax distribution free procedure for the (Q,R) inventory model. *Operations Research Letter* **11**(1) 55–60.

Goyal, M., S. Netessine. 2007. Strategic technology choice and capacity investment under demand uncertainty. *Management Science* **53**(2) 192–207.

Handfield, R. B., D. R. Krause, T. V. Scannell, R. M. Monczka. 2000. Avoid the pitfalls in supplier development. *Sloan Management Review* **41**(2) 37–49.

Haywood, K. 2009. GM helps axle maker avoid Chapter 11 . *The Wall Street J.* http://online.wsj.com/article/NA_WSJ_PUB:SB125322972929221255.html. Retrieved October 31, 2011.

Huh, W. T., R. Levi, P. Rusmevichientong, J. B. Orlin. 2011. Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Oper. Res.* **59**(4) 929–941.

Huh, W. T., P. Rusmevichientong. 2009. A nonparametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research* **34**(1) 103–123.

Iyer, A. V., L. B. Schwarz, S. A. Zenios. 2005. A principal-agent model for product specification and production. *Management Science* **51**(1) 106–119.

Kaminsky, P. M, M. K. Yuen. 2011. Production capacity investment with data updates. Working paper. University of California, Berkeley, CA.

Krause, D. R. 1997. Supplier development: Current practices and outcomes. *Journal of Supply Chain Management* **33**(2) 12–19.

Krause, D. R., R. B. Handfield, T. V. Scannell. 1998. An empirical investigation of supplier development: Reactive and strategic processes. *Journal of Operations Management* **17**(1) 39–58.

Krause, D. R., R. B. Handfield, B. B. Tyler. 2007. The relationships between supplier development, commitment, social capital accumulation and performance improvement. *Journal of Operations Management* **25**(2) 528–545.

Kwon, H. D., S. A. Lippman. 2011. Acquisition of project-specific assets with Bayesian updating. *Operations Research* **59**(5) 1119–1130.

LaPedus, M. 1995. Two in Taiwan lay new plans. *Electronic Buyers' News* (August 28) 10.

LaPedus, M. 2009. Intel invests in ASMI. *EE Times* http://www.eetimes.com/electronics-news/4082860/Intel-invests-in-ASMI. Retrieved October 31, 2011.

Lariviere, M. A., E. L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science* **45**(3) 346–363.

Li, C. 2013. Sourcing for supplier effort and competition: Design of the supply base and pricing mechanism. *Management Science* **59**(6) 1389–1406.

Li, C., L. G. Debo. 2009. Strategic dynamic sourcing from competing suppliers with transferable capacity investment. *Naval Research Logistics* **56**(6) 540–562.

Li, L., I. Duenyas, S. M. R. Iravani. 2011. Capacity option transfer rights: Do they benefit suppliers? Working paper. Northwestern University, Evanston, IL, and University of Michigan, Ann Arbor, MI.

Lovejoy, W. S. 1991. A survey of algorithmic methods for partially observed markov decision processes. *Annals of Operations Research* **28**(1) 47–65.

Lovejoy, W. S. 1993. Suboptimal policies, with bounds, for parameter adaptive decision processes. *Operations Research* **41**(3) 583–599.

Luss, H. 1982. Operations research and capacity expansion problems: A survey. *Operations Research* **30**(5) 907–947.

Manne, A. S. 1967. *Investments for capacity expansion, size, location, and time-phasing*. MIT Press, Cambridge, MA.

Mattioli, D., L. Burkitt. 2013. Companies seek to avoid China new year hangover. *The Wall Street Journal* http://online.wsj.com/article/SB10001424127887323864304578315872427597756.html. Retrieved April 7, 2013.

Monahan, G. E. 1982. State of the arta survey of partially observable markov decision processes: Theory, models, and algorithms. *Management Science* **28**(1) 1–16.

Osawa, J., M.-J. Lee. 2013. Samsung to invest $110 million in Sharp. *The Wall Street Journal* http://online.wsj.com/article/SB10001424127887323628804578343912771185952.html. Retrieved March 6, 2013.

Philip, L. 2013. Mahindra & Mahindra raises capacity of Quanto, XUV500 and Rexton. *Economic Times* http://articles.economictimes.indiatimes.com/2013-01-14/news/36332058_1_xuv500-ssangyong-rexton-mahindra-s-quanto. Retrieved March 8, 2014.

Plambeck, E. L., T. A. Taylor. 2005. Sell the plant? The impact of contract manufacturing on innovation, capacity, and profitability. *Management Science* **51**(1) 133–150.

Plambeck, E. L., T. A. Taylor. 2007. Implications of renegotiation for optimal contract flexibility and investment. *Management Science* **53**(12) 1872–1886.

Scarf, H. 1959. Bayes solutions of the statistical inventory problem. *The Annals of Mathematical Statistics* **30**(2) 490–508.

Sprumont, Y. 1991. The division problem with single-peaked preferences: A characterization of the uniform allocation rule. *Econometrica* **59**(2) 509–519.

Team, T. 2012. Is Apple the invisible force behind the Sharp-Foxconn deal? *Forbes* http://www.forbes.com/sites/greatspeculations/2012/04/03/is-apple-the-invisible-force-behind-the-sharp-foxconn-deal/. Retrieved May 20, 2012.

Thakkar, K. 2012. M&M to reopen XUV 500 booking nationally from 8 June; ramps up capacity to 4,000 units a month. *Economic Times* http://economictimes.indiatimes.com /news/news-by-industry/auto/automobiles/ mm-to-reopen-xuv-5oo-booking-nationally-from-8-june-ramps-up-capacity-to-4000-units-a-month/articleshow/13594257.cms. Retrieved April 11, 2013.

Ülkü, S., L. B. Toktay, E. Yücesan. 2005. The impact of outsourced manufacturing on timing of entry in uncertain markets. *Production Operations Management* **14**(3) 301–314.

Ülkü, S., L. B. Toktay, E. Yücesan. 2007. Risk ownership in contract manufacturing. *Manufacturing & Service Operations Management* **9**(3) 225–241.

Van Mieghem, J. A. 2003. Capacity management, investment, and hedging: Review and recent developments. *Manufacturing & Service Operations Management* **5**(4) 269–302.

Wadecki, A. A., V. Babich, O. Q. Wu. 2011. Manufacturers' competition and subsidies to suppliers. H. Gurnani, A. Mehrotra, S. Ray, eds., *Supply Chain Disruptions: Theory and Practice of Managing Risk*. Springer-Verlag London Ltd, forthcoming.

Wang, W., M. Ferguson, S. Hu, G. C. Souza. 2013. Dynamic capacity investment with two competing technologies. *Manufacturing & Service Operations Management* Forthcoming.

Wang, Y., W. Gilland, B. Tomlin. 2010. Mitigating supply risk: Dual sourcing or process improvement? *Manufacturing & Service Operations Management* **12**(3) 489–510.

Wang, Y., Y. Xiao, N. Yang. 2014. Improving reliability of a shared supplier with competition and spillovers. *European Journal of Operations Research* Forthcoming.

Yang, Z., A. Göker, B. Volodymyr, B. Damian. 2009. Supply disruptions, asymmetric information, and a backup production option. *Management Science* **55**(2) 192–209.

Ye, Q., I. Duenyas. 2007. Optimal capacity investment decisions with two-sided fixed-capacity adjustment costs. *Operations Research.* **55**(2) 272–283.

Zhu, K., R. Q. Zhang, F. Tsung. 2007. Pushing quality improvement along supply chains. *Management Science* **53**(3) 421–436.