ESSAYS ON SOCIALLY RESPONSIBLE OPERATIONS WITH A FOCUS ON AGRICULTURAL AND HEALTH CARE INDUSTRIES

MING ZHAO

Ph.D

The Hong Kong Polytechnic University

2015

The Hong Kong Polytechnic University

Department of Logistics and Maritime Studies

Essays On Socially Responsible Operations With A Focus On Agricultural And Health Care Industries

Ming ZHAO

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

December 2014

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________(Signed)

Ming ZHAO (Name of student)

Abstract

This dissertation studies the socially responsible operations in agricultural and health care industries. In the first essay, we examine whether the government or non-governmental organizations (NGOs) can improve farmers' welfare by offering agricultural advice and market information. Towards this end, we consider a situation where farmers decide whether to use market information to improve their production plans or adopt agricultural advice to improve their operations when they engage in Cournot competition under both uncertain market demand and uncertain process yield. We show that both farmers will use the market information to improve their profits in equilibrium. Hence, relative to the base case in which market information is not available, the provision of market information can improve the farmers' total welfare (i.e., total profit for both farmers). Moreover, when the underlying process yield is highly uncertain or when the products are highly heterogeneous, the provision of market information is welfare-maximizing in the sense that the maximum total welfare of farmers is attained when both farmers utilize market information in equilibrium. Furthermore, in equilibrium, whether a farmer adopts the agricultural advice depends on the size of the requisite upfront investment. More importantly, we show that agricultural advice is not always welfare improving unless the upfront investment is sufficiently low. This result implies that to improve farmers' welfare, governments should consider offering farmer subsidies.

In the second essay, we study the performance measurements of health care systems. Many governments use waiting time per admission (i.e., the waiting time that a patient spends per admission) as a measurement to evaluate the congestion level and the performance of the health care system. Adopting this criterion may force doctors to speed up their service and spend less time on each patient, resulting in a decline in service quality and an increase in readmission rate. To characterize both system congestion and service quality, we propose a new performance measurement, total waiting time, which is the total amount of time a patient spends in the system before being cured. We then consider the optimal design and control issues of the health care system based on this new measurement. We model the health care system as an M/M/1 queue with Bernoulli feedback, where the service rate is a decision variable of the health care provider (HCP) and the readmission rate (i.e., the feedback) is increasing in the service rate. We study the decision problems for the three parties in the system: patients, the HCP and the social planner (government). We demonstrate that a naive adoption of the waiting time reduction target could even worsen the system performance, leading to a higher congestion level, lower accessibility for new patients and a busier HCP. We find that the social optimality cannot be achieved via the single pricing mechanism. Instead, the social planner needs to adopt a regulation mechanism with multi-dimensional control variables.

Publications arising from this thesis

- Tang, C. S., Y. Wang, M. Zhao. 2014. The implications of utilizing market information and adopting agricultural advice for farmers in developing economies. Forthcoming in *Production and Operations Management*.

- Guo, P., C. S. Tang, Y. Wang, M. Zhao. 2014. Why Minimizing Waiting Time in a Health Care System Could be Bad? *Working Paper*.

The author of this thesis also conducted another project during his PhD study, which results in a paper as follows:

- Zhao, M., Y. Wang, X. Gan. 2014. Signalling effect of daily deal promotion for a startup service provider. Forthcoming in *Journal of the Operational Research Society*.

However, because this paper is by no means related to the topic of socially responsible operations, we do not include it in this thesis.

Acknowledgements

During my Ph.D study, I was very fortunate to learn from many famous scholars in different fields. Before the submission of the thesis, I wish to express my deepest gratitude to them.

First of all, I wish to give my sincere thanks to my supervisors Dr. Yulan Wang, Dr. Xianghua Gan and Prof. Houmin Yan for their professional guidance and generous support. Dr. Xianghua Gan gave me the basic training in game theory and modelling skills that enable me to handle many challenging problems. Dr. Yulan Wang did not only give me the professional training in marketing operations interface, academic writing and how to derive managerial insights, but also help me to build up my network. I cannot image what my academic career would be without her help.

I would also like to thank Prof. Christopher Tang for providing me a wonderful research topic that leads to the first essay of this thesis. Prof. Tang leads me to enter the field of socially responsible operations. The thesis would not have been possible without his help.

Much credit also goes to Dr. Pengfei Guo, who leads me to enter the field of health care operations and help me to finish the second essay of this thesis. Dr. Guo instructed me in stochastic process and queueing models. He was always more than willing to share with me his wealth of knowledge and expertise.

During my Ph.D study, I have learned consumer behavior from Prof. Jianmin Jia and learned inventory models from Prof. Youhua Chen. I sincerely appreciate their guidance that equipped me with necessary knowledge and experience to conduct research in marketing and inventory management.

I also wish to thank Prof. Mingming Leng and Dr. Xubin Zhang for their discussion on other two marketing projects, which enriched my knowledge on channel marketing and consumer perception.

I owe a tremendous debt of gratitude to my parents. Without their constant support and unconditional love, I could not have done this.

Finally, the thesis is dedicated to my girlfriend, Dr. Yunpeng Yang. Every time when I faced difficulty in my study, she would always be there to support and encourage me. Her love always keeps me going.

Contents

Chapter 1

Introduction

Over the past few years, there has been increasing attentions on socially responsible operations. Being socially responsible requires organizations to act for the benefit of society instead of their own good. According to Paul Polman, CEO of Unilever,“A company’s contribution to society is absolutely critical in today’s environment. $\cdots$ It is very clear that this world faces some considerable challenges: poverty, water, global warming and climate change. Businesses like Unilever have a responsibility here and thus a major role to play. $\cdots$ if we do the right thing and leverage this enormous scale, we have a tremendous opportunity to make a major impact on society and the environment.”[1] Therefore, beyond profit maximization, a socially responsible organization should incorporate social welfare, economic and environmental effects into its objectives. Such objectives raise issues fundamentally different from those examined in the existing operations management literature, especially in agricultural and health care industries. This dissertation studies the socially responsible operations in agricultural and health care industries. More specifically, in Chapter 2, we study whether the government or non-governmental organizations (NGOs) can improve farmers’ welfare by offering agricultural advice and market information. In Chapter 3, we examine the performance measurements of health care systems. Below we overview each chapter and summarize the major implications.

Chapter 2 is motivated by the recent effort to alleviate farmers’ poverty in developing economies. Agriculture is a dominant section in most of the developing countries. For example, agriculture accounts for 70% of the employment, 33% of the total GDP and 40% of the total export earnings in Sub-Saharan Africa (John-

[1] See http://www.criticaleye.net/archive.cfm?id=299

son and Hazell 2002). One of the biggest challenge faced by most of the developing countries is how to alleviate farmers' poverty. For example, nearly 20% of farmers in India live below the poverty line (Sharma 2013), and around 99 million small-scale farmers in China have an annual income lower than the national poverty alleviation standard of 2,300 yuan (Pierson 2013). According to the World Bank, 70% of the world's poor people live in rural areas of developing countries and are engaged in agricultural activities for their livelihood.[2]

To alleviate poverty in developing countries, governments and NGOs disseminate two types of information: (1) agricultural advice to enable farmers to improve their operations (cost reduction, quality improvement, and process yield increase); and (2) market information about future price/demand to enable farmers to make better production planning decisions. This information is usually disseminated free of charge. While farmers can use the market information to improve their production plans without incurring any (significant) cost, adopting agricultural advice to improve operations requires upfront investment, for example, equipment, fertilizers, pesticides, and higher quality seeds. We examine whether farmers should use market information to improve their production plans or adopt agricultural advice to improve their operations when they engage in Cournot competition under both uncertain market demand and uncertain process yield.

Our analysis indicates that both farmers will use the market information to improve their profits in equilibrium. Hence, relative to the base case in which market information is not available, the provision of market information can improve the farmers' total welfare (i.e., total profit for both farmers). Moreover, when the underlying process yield is highly uncertain or when the products are highly heterogeneous, the provision of market information is welfare-maximizing in the sense that the maximum total welfare of farmers is attained when both farmers utilize

[2] See http://data.worldbank.org/about/world-development-indicators-data/agriculture-and-rural-development

market information in equilibrium. Furthermore, in equilibrium, whether a farmer adopts the agricultural advice depends on the size of the requisite upfront investment. More importantly, we show that agricultural advice is not always welfare improving unless the upfront investment is sufficiently low. This result implies that to improve farmers' welfare, governments should consider offering farmer subsidies.

In chapter 3, we investigate how to evaluate a health care system's performance. Recently, due to greater data availability, many governments and organizations are developing measurements to evaluate a health care system's performance. Those measurements can help people better understand where and what changes need to be made in order to improve performance and the quality of care delivered. Therein, as an important measurement to evaluate a health care system's congestion levels and performance, waiting time per admission (i.e., the waiting time that a patient spends per admission) is widely adopted in many countries such as UK and New Zealand. However, adopting this criterion may force doctors to speed up their service and spend less time on each patient, resulting in a decline in service quality and an increase in readmission rate. This motivate us to study the performance indicators in health care industries by taking the tradeoff between service speed and service quality in terms of readmission rate into account. To this end, we model the health care system as an M/M/1 queue with Bernoulli feedback, where the service rate is a decision variable of the HCP and the readmission rate (i.e., the feedback) is increasing in the service rate. The existence of readmission risk enable us to define three new performance measurements: (i) *effective service rate*, which is the mean number of patients cured by the health care provider (HCP) per unit time; (ii) *total waiting time*, which is the total amount of time a patient spends in the system before being cured; (iii) *the utilization rate for new patients*, which measures how much time the HCP spends on treating new patients.

We first conduct the sensitive analysis of those measurements with respect to the service rate. Our results show that an increase in service rate may increase the total waiting time but reduce the waiting time per admission. Therefore, reducing waiting time per admission may incentivize the HCP to increase its service speed, which in turn increases the readmission rate and intensifies the system congestion in terms of the total waiting time. Furthermore, a higher total utilization rate may lead to a smaller utilization rate for new patients. This reveals that keeping the HCP busy may cause the HCP to spend less time on treating new patients.

Next, by examining patients' joining decision, we find that increasing total effective arrival rate may reduce the effective arrival rate of new patients, which implies that improving the accessibility for all the patients may reduce the accessibility for new patients. Furthermore, increasing the arrival rate of new patients (total arrival rate) may reduce the utilization rate for new patients (total utilization rate). This implies that reducing HCP's idle time may reduce the accessibility of the health care services. Moreover, by studying the HCP's decision of service rate, we demonstrate that a higher price may mitigate system congestion in terms of total waiting time but intensify system congestion in terms of waiting time per admission. Finally, we also derive the optimal price from the perspective of the HCP and the social planner respectively. We show that if the price is determined by the HCP, the service rate is socially optimal, while if the price is controlled by the social planner, the service rate is suboptimal. Therefore, price control hampers the efficiency of the HCP's care delivery.

Chapter 2

The Implications of Utilizing Market Information and Adopting Agricultural Advice for Farmers in Developing Economies

2.1 Introduction

Agriculture plays an important role in emerging economies. For example, the agricultural sector accounts for 50% and 74% of the total workforce in India and Kenya, respectively. However, the farmers in these regions remain poor because they lack opportunities to improve their operations so that they can produce higher yield, more available and better quality crops at lower cost. The lack of market information and agricultural advice often results in market inefficiency, poor yields, and huge crop wastage, all of which damage farmers' earnings and livelihoods. Without agricultural advisory information, farmers may not be able to make proper planning decisions in areas such as pest control and soil depletion, which often lead to low and uncertain yields. Moreover, without information about future market price trends, farmers cannot make effective production quantity decisions, which in turn can affect their realized profit. These factors further aggravate the difficult situation faced by those at the bottom of the pyramid, as this large population of farmers has few income sources (Jensen 2007).

Recognizing these challenges faced by farmers, many governments have developed agricultural extension services by providing the following two types of information: (1) agricultural advice that can help farmers to improve their operations (i.e., *how to grow?*) by reducing operating costs, improving quality, and increasing process yield;[1] and (2) market (price/demand) forecast information that can help farmers to make better long-term production planning decisions (i.e., *how much to grow?*).[2] For example, the Indian government provides both types of information on its website (www.india.gov.in/topics/agriculture) for free.

Due to a lack of Internet access, many governments and NGOs disseminate agricultural advice and market information through different channels including radio, television, and call centers. For example, in Kenya and Mali, an NGO launched a weekly hour-long radio program called Mali Shambani that discusses farming techniques and market price trends, etc. This free radio program also offers an interactive call-in component for farmers to ask agricultural questions via phone or SMS messaging. In the same vein, the India Ministry of Agriculture launched the Kisan Call Centers in 2004 to deliver extension services to farmers over the phone. This free service enables Indian farmers to use their phones to seek advice and to gain access to information posted on the Internet. In another example, in India the NGO Digital Green (digitalgreen.org) distributes farming advisory information through online videos and DVDs delivered to farmers free of charge.

It is certainly a big step forward for more farmers to gain access to agricultural advice and market information. However, even if the information is free of charge, farmers need to decide whether to use such market information when making their production plans, especially when they compete under uncertain market demand and uncertain process yield. Moreover, because the adoption of agricultural advice often requires upfront investment (tools, seeds, fertilizers, etc.), farmers need to decide whether it is cost-effective to do so.

These observations motivate us to examine a situation in which two risk-neutral

[1] Agricultural advisory information includes: (a) tools and equipment for seedbed preparation, sowing, planting and harvesting; (b) high quality seeds and the safe use of approved pesticides and fertilizers; (c) pest management and locust control; and (d) soil and water conservation.

[2] Some governments also provide information about the current market prices in different markets to help farmers to make better short-term selling decisions (i.e., when/where to sell?). For the sake of tractability, we do not model this type of information in this chapter.

farmers engage in Cournot competition under uncertain market demand and uncertain process yield. (Note that our model can be extended to multiple farmers by considering the "proportion" of the farmers who adopt the advice (or utilize the market information).) We provide three justifications for selecting Cournot competition as our modeling choice. First, Carter and MacLaren (1994) argue that Cournot competition is a good approximation of the real economic decision-making for managing perishable products, especially when the production quantity cannot be changed quickly in advance of sales (e.g., fruits, vegetables) or for managing products with lengthy production processes (e.g., tree crops). Second, Deodhar and Sheldon (1996), Dong et al. (2006), and others provide empirical evidence to support the existence of Cournot competition in various agricultural product markets, such as malting barley and banana. Third, if one interprets a "farmer" in our model as a marketing board (or a marketing cooperative) that represents a group of farmers in countries such as India and South Africa, then our model captures the quantity competition between two marketing boards. In this context, each marketing board sets the aggregate production quantity to control the market price. To do so, each marketing board may impose a quota on each farmer's production quantity (Nieuwoudt 1987).

To alleviate poverty, we consider the case in which the government offers either agricultural advice or market information but not both.[3] The analysis for the case in which the government offers only market information is simpler because no upfront investment is involved. However, for the case in which the government offers only agricultural advice, we need to model the upfront investment associated with the adoption of the agricultural advice *endogenously*.

Our model is intended to examine the following research questions:

1. Should all farmers use market information to plan their production in equilib-

[3]The analysis associated with the case in which the government offers both agricultural and market information is intractable because it involves 16 subgames.

rium?

2. Does market information improve farmers' welfare?

3. Should all farmers adopt agricultural advice to improve their operations in equilibrium when upfront investment is involved?

4. Does agricultural advice improve farmers' welfare?

To analyze the implications of providing market information and/or agricultural advice, we first establish a unified approach that combines market information and agricultural advice without considering the upfront investment. This unified approach enables us to investigate the interactions among various types of operational improvement induced by the two types of information. Specifically, we find that the effect of yield improvement is complementary to quality improvement and market demand forecast accuracy. In addition, when the process yield is highly uncertain, yield improvement and cost reduction are always complementary. However, when the uncertainty of the yield is relatively low, yield improvement and cost reduction are complementary if and only if the unit cost is large. Finally, we show that information accuracy improvement has no effect on quality improvement and cost reduction. By examining the equilibrium outcomes associated with the unified approach, we are able to separately investigate whether farmers should utilize market information and whether farmers should adopt agricultural advice. Our equilibrium analysis enables us to obtain the following results. First, we show that the provision of market information always improves the farmers' total welfare (i.e., the sum of the profits of all farmers) and that farmers should use market information to improve their production planning in equilibrium. However, as all farmers use market information to plan their production, we find that market information can maximize farmers' welfare when the underlying process yield is highly uncertain or when the products are highly heterogeneous. Second, in equilibrium, whether a farmer should adopt the agricultural advice is dependent on the size of the requisite upfront investment. More importantly, we show that agricultural advice is not always welfare improving, unless the upfront investment is sufficiently low. This result implies that to improve farmers' welfare, governments should consider offering farmer subsidies.

As such, our work makes two major contributions to the existing literature on socially responsible operations. First, we examine the value of market information when farmers engage in quantity competition under both uncertain market demand and uncertain process yield. Second, by endogenizing the investment decision, we examine the value of agricultural advice.

The rest of this chapter is organized as follows. In Section 2.2, we review the relevant literature. Section 2.3 presents a unified framework that enables us to analyze two separate settings (market information and agricultural advice) by analyzing a single model. We also analyze the underlying subgames and the meta-game in this section. In Section 2.4, we consider the case in which the government offers only market information and analyze the behavior of each farmer in equilibrium. Section 2.5 deals with the case in which the government offers only agricultural advice and each farmer needs to pay an upfront investment if he chooses to adopt the advice. Concluding remarks are provided in Section 2.6 and all proofs are relegated to the Appendix A.

2.2 Literature review

There is limited modeling literature on making socially responsible operations because this topic is an emerging research area in operations management. Accordingly, most of the relevant articles are recent. For example, Chen et al. (2013a) examine the ITC e-Choupal network and discuss how it substantially changes the information and material flows. They show that the implicit agreement between the contracted farmers and the ITC behaves like a formal contract and it is in the ITC's best interest to provide all of the farmers with its services. Chen et al. (2013b) further study the peer-to-peer information sharing in Avaaj Otalo. They articulate why and when farmers have incentives to share demand and price information with other competing farmers, and why those who provide answers to others may be criticized and under-appreciated. Specifically, they show that the responses of the knowledgeable farmers are always less informative than those of the experts. (See Sodhi and Tang 2012 for a comprehensive survey.) Chen and Tang (2013) examine the value of public and private information for the economic development of agricultural business. They show that farmers are more responsive to the public/private signal when the public/private signal is more accurate. Therefore, when the public signal becomes more accurate, the effect of the private signal on the farmers' welfare decreases. Dawande et al. (2013) study the surface water allocation problem in developing economies. An et al. (2014) show that cooperatives are beneficial to the affiliated farmers only when the size of the cooperatives is relatively small. However, all of the aforementioned studies assume that farmers utilize the information they receive. In contrast, we provide the option for each farmer to decide whether to utilize the market information. Because upfront investment is involved, we allow each farmer to decide whether to adopt certain agricultural advice.

In addition to examining the adoption of agricultural advice, we examine the utilization of market information under competition. Therefore, our work is also related to the literature on information sharing in oligopolies. This line of research mainly examines whether firms have incentives to share their private information with competitors. The private information may concern either uncertain common value such as demand intercepts or uncertain private value such as costs. For example, Gal-Or (1985 and 1986) examine whether competing firms should share common demand intercept or production cost information with each other. Her results suggest that the incentive for information sharing crucially depends on the content of information (demand versus cost) and the nature of competition (quantity versus price). Raith (1996) provides a comprehensive survey of this research stream. Vives (1988) studies a large market in which firms engage in Cournot (quantity) competition and have access to private signals about the uncertain market demand. He shows that information is aggregated inefficiently and there is welfare loss even if the market is asymptotically competitive. The same informational setting is adopted by Li (2002). Using a two-tier supply chain relationship, he examines whether a downstream retailer has an incentive to share information with the upstream supplier, because the supplier may pass this information to other competing retailers. Armantier and Richard (2003) empirically show that sharing cost information in the airline industry can benefit the airlines without hurting the consumers. Zhu (2004) explores a B2B exchange that provides an online platform for information transmission. He shows that whether a firm should join the B2B exchange depends on the cost heterogeneity, product differentiation, and the degree of uncertainty. Hueth and Marcoul (2006) consider the information sharing among agricultural intermediaries. They show that even if information sharing can increase the profit of each firm, firms will conceal information in equilibrium. Jansen (2010) studies the information sharing in R&D competition. He shows that the incentive to disclose information depends on whether the winner firm of an R&D race is capable of appropriating the full revenue of its innovation. In this chapter, we explore an entirely different context. First, we consider both the uncertain common value (e.g., market potential) and the uncertain private value (e.g., the process yield). Second, although some of our results can shed light on the farmers' incentive for information sharing, we do not explicitly examine this issue. Instead, we consider the case in which the government offers market information and agricultural advice to farmers. More importantly, we focus on the issue of whether each farmer should utilize market information and whether each farmer should adopt agricultural advice when both market demand and process yield are uncertain.

2.3 The model

Consider two farmers who produce and sell the same crop in a common market.[4] Each farmer i, $i = 1, 2$, incurs a production cost cq_i, where c is the unit production cost and q_i is the production quantity (a decision variable). In the base case, when farmer i processes q_i units, farmer i's actual output is $z_i q_i$, where z_i is the uncertain process yield such that $E(z_i) = \mu_y$, $\mu_y \leq 1$, and $Var(z_i) = \sigma_y^2$. We assume that z_1 and z_2 are independent random variables. This assumption is reasonable given that the farming skill levels of different farmers are normally independent. For notational convenience, we define S^2 and C_y as the second moment and coefficient of variation (CV) of z_i, respectively, so that $S^2 = E(z_i^2) = \sigma_y^2 + \mu_y^2$ and $C_y = \sigma_y / \mu_y$.

To capture the quantity competition under uncertain market demand, we assume that the uncertain market price

$$P = M - (z_1 q_1 + z_2 q_2),$$

where $M > 0$ corresponds to the uncertain market potential.[5] We also assume that M is independent of z_i, $i = 1, 2$, and normally distributed with a mean of μ_m and a variance of σ_m^2, that is, $M \sim N(\mu_m, \sigma_m^2)$. Consider the case in which neither product is available in the market (i.e., $q_1 = q_2 = 0$) so that the market price equals M. In this case, if we let one farmer i produce an infinitesimal amount (and the other farmer produces nothing), then the gross margin of farmer i is $E(Mz_i - c) = \mu_m \mu_y - c$. For notational convenience, hereafter we denote $g = \mu_m \mu_y - c$, which represents the expected gross margin when only one farmer exists in the market (i.e., without considering quantity competition).

[4]Our analysis can be extended to the case in which there are $n > 2$ farmers.

[5]For ease of exposition, we set the price elasticity to 1. However, our model can be extended to the case of $P = M - b(z_1 q_1 + z_2 q_2)$, where $b > 0$.

We now model the implications of the aforementioned information that is intended to help farmers make better production planning decisions and improve their operations.

Market information. The government offers information I that would enable farmers to improve the accuracy of their forecast of the market potential M. To facilitate our analysis, we assume that (M, I) are bivariate normally distributed so that

$$(M, I) \sim N(\mu_m, \mu_I, \sigma_m^2, \sigma_I^2, \rho),$$

where ρ is the correlation coefficient between the market potential and market information and $\rho \in (-1, 1)$. We also assume that both M and I are independent of z_i, $i = 1, 2$.[6] Each farmer can use information I to "update" her forecast on M. By considering the conditional expectation and conditional variance, we get

$$E(M|I) = \mu_m + \rho \frac{\sigma_m}{\sigma_I}(I - \mu_I) \text{ and } Var(M|I) = \sigma_m^2(1 - \rho^2).$$

By noting that $Var(M|I) \leq Var(M)$, we can conclude that each farmer can use information I to obtain a more accurate forecast of the market potential.

Agricultural Advice. If a farmer adopts this agricultural advice by making an upfront investment K, she can enjoy three benefits that are described as follows.

1. **Cost reduction.** Each farmer reduces her unit production cost from c to βc, where $\beta \leq 1$.

2. **Quality improvement.** Each farmer improves her product quality so that the average market potential is increased from μ_m to $\alpha\mu_m$, where $\alpha \geq 1$. Therefore, the improved market potential, denoted by αM, follows the normal distribution with mean $\alpha\mu_m$ and variance σ_m^2, that is, $\alpha M \sim N(\alpha\mu_m, \sigma_m^2)$.

[6]This assumption is reasonable because market information normally will not impact the farming skills of the farmers.

3. **Process yield improvement.** The process yield of farmer i, $i = 1, 2$ is increased from z_i to $z_i^{'}$, where $E(z_i^{'}) = \gamma\mu_y \geq E(z_i)$ ($\gamma \geq 1$) and $Var(z_i^{'}) = \sigma_y^2$.[7] To ensure that the improved yield is bounded by 1, $\gamma\mu_y \leq 1$ is required. Similar to S^2 and g associated with the regular yield z_i as defined earlier, we let $S^{'2} = \sigma_y^2 + \gamma^2\mu_y^2$ and $g^{'} = \alpha\gamma\mu_m\mu_y - \beta c$.[8]

Table 2.1: List of Notations

Notation	Definition
M	market potential (a random variable)
(μ_m, σ_m^2)	mean and variance of market potential
I	market information (a random variable)
(μ_I, σ_I^2)	mean and variance of market information
$\rho = Corr(M, I)$	correlation coefficient between market information and market potential
c	regular unit production cost
K	upfront investment cost for adopting agricultural advice
α	quality improvement parameter, $\alpha \geq 1$
β	cost reduction parameter, $0 < \beta \leq 1$
γ	yield level improvement parameter, $\gamma \geq 1$
q_i	production quantity of farmer i , $i = 1, 2$ (a decision variable)
z_i	regular process yield of farmer i, $i = 1, 2$ (a random variable)
$(\mu_y, \sigma_y^2, S^2, C_y)$	mean, variance, second moment and coefficient of variation (CV) of regular process yield z_i, $i = 1, 2$, where $S^2 = \sigma_y^2 + \mu_y^2$, $C_y = \sigma_y/\mu_y$
$z_i^{'}$	improved process yield of farmer i, $i = 1, 2$
$S^{'2} = \sigma_y^2 + \gamma^2\mu_y^2$	second moment of $z_i^{'}$, $i = 1, 2$
$g = \mu_m\mu_y - c$	gross margin "before" adopting agricultural advice (without considering quantity competition)
$g^{'} = \alpha\gamma\mu_m\mu_y - \beta c$	gross margin "after" adopting agricultural advice (without considering quantity competition)

Table 3.1 summarizes the notations used in this chapter. According to our model description, market information and agricultural advice affect farmers in two different ways. First, when utilizing market information I, each farmer i can use the updated market uncertainty $(M|I)$ to determine her production quantity q_i. Second, when adopting agricultural advice, each farmer incurs an upfront investment K. However, when deciding on her production quantity q_i, farmer i enjoys three benefits

[7]For ease of exposition, we assume that the quality and yield improvements affect only the mean value of z_i but not the variance. However, the structure of the results remains the same when we relax this assumption. Furthermore, this setting also enables us to examine the impact of yield improvement on the uncertainty of process yield in terms of coefficient of variation.

[8]Because $\alpha \geq 1$, $\gamma \geq 1$ and $\beta \leq 1$, we have $S^{'} \geq S$ and $g^{'} \geq g$. Also, to ensure that both $E(M|I)$ and the equilibrium production quantity in our analysis are non-negative, we assume that after adopting the advice, the "improved" gross margin $g^{'}$ is large enough so that $g^{'} + 2\rho\frac{\sigma_m}{\sigma_I}(I - \mu_I) > 0$.

associated with quality improvement via α, cost reduction via β, and process yield improvement via γ. Although market information and agricultural advice affect farmers in different ways, we now present a unified approach so that we can use one generic model to analyze their implications. Specifically, we introduce our unified approach and analyze the equilibrium outcomes of our generic model in the next section. We then use the equilibrium outcomes of our generic model to separately examine the implications of the provision of market information and agricultural advice in the subsequent sections.

2.4 A Unified Approach

In this section, we introduce a unified approach that combines market information and agricultural advice. Recall that the adoption of agricultural advice involves the upfront investment cost K while the adoption of market information does not. If we suppress the upfront investment K that is associated with the adoption of agricultural advice, we can characterize the meta-game between the two farmers as a two-person game in which each player has to decide whether to utilize market information (respectively, whether to adopt agricultural advice). For tractability, we shall assume that under our unified approach, each farmer either utilizes both market information and agricultural advice (denoted by Y), or utilizes nothing at all (denoted as N).[9] Consequently, there are four pairs of strategies: $(N, N), (Y, N), (N, Y)$, and (Y, Y). For each of these four pairs of strategies, there is a corresponding subgame in which both farmers engage in Cournot (quantity) competition. For each subgame, we need to determine the production quantity and the expected payoff of each farmer in equilibrium. We use superscript to denote the equilibrium outcome

[9] If both market information and agricultural advice are available, each farmer has four options to choose from; i.e., whether to utilize market information or not and whether to adopt agricultural advice or not. Under this setting, these two farmers engage in $4\times4 = 16$ corresponding subgames. The analysis of these 16 subgames and the comparisons among equilibrium outcomes of these 16 subgames would become very tedious.

of each subgame. For example, for subgame (Y, N), let q_i^{YN} be the production quantity and π_i^{YN} be the expected payoff of each farmer i in equilibrium, $i = 1, 2$.

Our unified approach for analyzing the implications of market information and agricultural advice can be described as follows. First, we solve all four subgames by determining the expected payoff of each farmer assuming that both market information and agricultural advice are available. Second, to analyze the implications of market information, we first determine the expected payoff of each farmer for the case in which only market information is available by setting $\alpha = \beta = \gamma = 1$. To determine whether each farmer "utilizes" market information in equilibrium, we solve the 2x2 meta-game by using the expected payoffs determined in those four subgames, as shown in Table 2.2.

Table 2.2: Farmers' Expected Payoffs

Farmer 2 / Farmer 1	N (do not utilize/adopt)	Y (utilize /adopt)
N (do no utilize/adopt)	π_1^{NN}, π_2^{NN}	π_1^{NY}, π_2^{NY}
Y (utilize/adopt)	π_1^{YN}, π_2^{YN}	π_1^{YY}, π_2^{YY}

Next, to analyze the implications of agricultural advice, we first determine the expected payoff of each farmer for the case in which only agricultural advice is available by setting $\rho = 0$. To determine whether each farmer "adopts" agricultural advice in equilibrium, we solve the meta-game by using the expected payoffs by accounting for the upfront investment K as determined in the four subgames, as shown in Table 2.2.

2.4.1 Analysis of Subgames

Using the above mentioned unified approach, we now proceed to analyze the four subgames that correspond to those 4 pairs of strategies: $(N, N), (Y, N), (N, Y)$, and (Y, Y). Due to symmetry, the analysis associated with subgame (Y, N) is identical to that of subgame (N, Y). Therefore, it suffices to analyze subgames $(N, N), (Y, Y)$, and (Y, N). After determining the equilibrium outcomes of all four subgames, we can then solve the meta-game as depicted in Table 2.2.

Under our unified approach, subgame (Y, Y) is the most complex because it involves both the utilization of market information and the adoption of agricultural advice. For any revealed market information I, the expected profit of farmer i needs to take into account all three benefits associated with the adoption of agricultural advice; namely, the effective market price becomes $P' = \alpha M - (z_i' q_i + z_{3-i}' q_{3-i})$, the effective output becomes $z_i' q_i$, and the effective production cost becomes $\beta c q_i$. Therefore, the expected profit of farmer i under strategy (Y, Y) can be expressed as

$$\begin{aligned} \pi_i(q_i|I) &= E\{(\alpha M - z_i' q_i - z_j' q_j) z_i' q_i - \beta c q_i | I\} \\ &= g' q_i + \gamma \rho \mu_y \frac{\sigma_m}{\sigma_I}(I - \mu_I) q_i - S'^2 q_i^2 - \gamma^2 \mu_y^2 q_i q_j | I, \text{ for } i, j = 1, 2, j \neq i. \quad (2.1) \end{aligned}$$

Because the other strategies are special cases of strategy (Y, Y), we can use the above expression to determine the farmer's profit under the other strategies. Given the expected profit as stated in (2.1), we are now ready to solve all of the subgames.

Subgame (N, N)

The case in which neither farmer utilizes market information (nor adopts agriculture advice) under strategy (N, N) corresponds to the case that $\alpha = \beta = \gamma = 1$ and $\rho = 0$ so that $g' = g$ and $S'^2 = S^2$. By utilizing (2.1), the expected profit of farmer i can be written as

$$\pi_i(q_i) = g q_i - S^2 q_i^2 - \mu_y^2 q_i q_j, \quad i = 1, 2, \ j \neq i.$$

By noting that $\pi_i(q_i)$ is concave, we obtain farmer i's best response function as follows:

$$q_i(q_j) = \frac{g - \mu_y^2 q_j}{2S^2}, \quad i = 1, 2, \; j \neq i.$$

Solving the best response functions of both farmers, we obtain the following proposition.

Proposition 1. *The equilibrium outcomes associated with strategy* (N, N) *satisfy*

$$q_i^{NN} = \frac{g}{2S^2 + \mu_y^2}, \quad i = 1, 2, \tag{2.2}$$

$$\pi_i^{NN} = \frac{S^2 g^2}{(2S^2 + \mu_y^2)^2}, \quad i = 1, 2, \tag{2.3}$$

where $g = \mu_m \mu_y - c$ *and* $S^2 = \sigma_y^2 + \mu_y^2$*. Both the expected profit* π_i^{NN} *and production quantity* q_i^{NN} *are increasing in* μ_m *and decreasing in* σ_y*. Moreover,* π_i^{NN} *and the expected output* $\mu_y q_i^{NN}$ *are both increasing in* μ_y*.*

Proposition 1 implies that without utilizing market information (or without adopting agricultural advice), a higher market potential μ_m or a lower yield uncertainty (via higher μ_y or lower σ_y) can enable both farmers to produce more and earn more. However, the production quantity q_i^{NN} is not necessarily increasing in μ_y. On the one hand, a higher expected yield μ_y can certainly enable each farmer to generate more output with the same input. On the other hand, to avoid over supply that drives down the market price, a higher expected yield μ_y can also cause each farmer to produce less in equilibrium. Although the production quantity q_i^{NN} is not monotone in μ_y for farmer i, the expected output $\mu_y q_i^{NN}$ is increasing in μ_y. Therefore, a higher expected yield will always lead to a larger expected output even when a farmer produces less.

Subgame (Y, Y)

When both farmers utilize market information (or adopt agriculture advice) under strategy (Y, Y), the expected profit of each farmer for any given information I is given in (2.1). In this case, it is easy to check that the best response function of farmer i is

$$q_i(q_j)|I = \frac{g^{'}\sigma_I + \gamma\rho\mu_y\sigma_m(I-\mu_I)}{2\sigma_I S^{'2}} - \frac{\gamma^2\mu_y^2 q_j}{2S^{'2}}, \quad i = 1, 2, \ j \neq i.$$

Recall that $g^{'} = \alpha\gamma\mu_m\mu_y - \beta c$ and $S^{'2} = \sigma_y^2 + \gamma^2\mu_y^2$. By considering the best response functions of both farmers, we have the following result.

Lemma 1. *For any given market information I, the ex post equilibrium outcomes associated with strategy $(Y,\ Y)$ satisfy*

$$q_i^{YY}|I \quad = \quad \frac{g^{'}\sigma_I + \gamma\rho\mu_y\sigma_m(I-\mu_I)}{\sigma_I(2S^{'2} + \gamma^2\mu_y^2)}, \ \textit{for } i = 1, 2,$$

$$\pi_i^{YY}|I \quad = \quad \frac{S^{'2}[g^{'}\sigma_I + \gamma\rho\mu_y\sigma_m(I-\mu_I)]^2}{\sigma_I^2(2S^{'2} + \gamma^2\mu_y^2)^2}, \ \textit{for } i = 1, 2.$$

Based on Lemma 1, we can obtain the ex ante equilibrium outcomes as follows.

Proposition 2. *The ex ante equilibrium outcomes associated with strategy $(Y,\ Y)$ satisfy*

$$q_i^{YY} \quad = \quad \frac{g^{'}}{2S^{'2} + \gamma^2\mu_y^2}, \ \textit{for } i = 1, 2, \tag{2.4}$$

$$\pi_i^{YY} \quad = \quad \frac{S^{'2}(g^{'2} + \gamma^2\rho^2\mu_y^2\sigma_m^2)}{(2S^{'2} + \gamma^2\mu_y^2)^2}, \ \textit{for } i = 1, 2. \tag{2.5}$$

Both q_i^{YY} and π_i^{YY} are increasing in α and decreasing in β. Furthermore, π_i^{YY} is *increasing in* ρ^2, γ *and* σ_m.

Proposition 2 reveals that when farmers utilize market information (and adopt agricultural advice), they can earn more when the market information becomes more informative (i.e., as ρ^2 increases) or when the agricultural advice becomes more beneficial (i.e., as α and γ increase, or as β decreases). This result is intuitive.

Recall that $E(M|I) = \mu_m + \rho\frac{\sigma_m}{\sigma_I}(I - \mu_I)$ and $Var(M|I) = \sigma_m^2(1-\rho^2)$. Then, we can use $E_I E(M|I) = \mu_m$ and (2.4) to show that the ex ante production quantity is independent of market information via ρ. Next, observe that market information I enables each farmer to reduce the variance of market potential M from $Var(M)$ to $Var(M|I)$, where $Var(M|I) = \sigma_m^2(1 - \rho^2) = Var(M) - \rho^2\sigma_m^2$. By noting that the term $\rho^2\sigma_m^2$ represents the reduction of variance when a farmer utilizes market information, it is intuitive to see that each farmer's expected profit increases in relation to the amount of variance reduction $\rho^2\sigma_m^2$. Finally, similar to Lemma 1, when the expected yield μ_y increases, each farmer may reduce the production quantity q_i^{YY} in equilibrium to avoid over supply that drives down market price.

Subgame (Y, N)

Under strategy (Y, N), farmer 1 is the only farmer who utilizes market information (and adopts agricultural advice). Thus, farmer 1's market potential becomes αM while farmer 2' market potential remains M. By noting that farmer 2's process yield is z_2, we can derive the expected profit of farmer 1 by replacing z_2' with z_2 in (2.1), getting

$$\begin{aligned} \pi_1(q_1) &= E\{(\alpha M - (z_1' q_1 + z_2 q_2)) z_1' q_1 - \beta c q_1 | I\} \\ &= g' q_1 + \gamma\rho\mu_y \frac{\sigma_m}{\sigma_I}(I - \mu_I) q_1 - S'^2 q_1^2 - \gamma\mu_y^2 q_1 q_2 | I. \end{aligned} \tag{2.6}$$

However, although farmer 2 does not utilize market information or adopt agricultural advice, she knows that farmer 1 observes $I = \mu_I$ (in expectation).[10] Therefore, farmer 2's expected profit is

$$\begin{aligned}\pi_2(q_2) &= E\{(M - (z_1^{'} q_1 | I = \mu_I) - z_2 q_2) z_2 q_2 - c q_2\} \\ &= g q_2 - S^2 q_2^2 - (\gamma \mu_y^2 q_1 | I = \mu_I) q_2. \qquad (2.7)\end{aligned}$$

By deriving the first order conditions of (2.6) and (2.7), we can obtain the following best response functions:

$$q_1(q_2)|I = \frac{g^{'}\sigma_I + \gamma\rho\mu_y\sigma_m(I - \mu_I)}{2\sigma_I S^{'2}} - \frac{\gamma\mu_y^2 q_2}{2S^{'2}}. \qquad (2.8)$$

$$q_2(q_1) = \frac{g - \gamma\mu_y^2(q_1 | I = \mu_I)}{2S^2}. \qquad (2.9)$$

Based on the above best response functions, we can derive the ex post equilibrium outcomes whose expressions hinge upon the value of C_y, the coefficient of variation of the process yield.

Lemma 2. *For any given market information I, the ex post equilibrium outcomes*

[10]From (2.8), we can see that $E_I(q_1|I) = (q_1|I = \mu_I)$. Therefore, by assuming that farmer 2 observes the mean market information, we can also obtain the Bayesian equilibrium.

associated with strategy $(Y,\, N)$ *satisfy*

$$q_1^{YN}|I \;=\; \begin{cases} \frac{2S^2g^{'}-\gamma\mu_y^2 g}{4S^2S^{'2}-\gamma^2\mu_y^4} + \frac{\gamma\rho\mu_y\sigma_m(I-\mu_I)}{2\sigma_I S^{'2}}, & if\ C_y^2 > \frac{\gamma g^{'}}{2g} - \gamma^2, \\ \frac{g^{'}\sigma_I + \gamma\rho\mu_y\sigma_m(I-\mu_I)}{2\sigma_I S^{'2}}, & otherwise, \end{cases} \tag{2.10}$$

$$q_2^{YN} \;=\; \begin{cases} \frac{2S^{'2}g-\gamma\mu_y^2 g^{'}}{4S^2S^{'2}-\gamma^2\mu_y^4}, & if\ C_y^2 > \frac{\gamma g^{'}}{2g} - \gamma^2, \\ 0, & otherwise, \end{cases} \tag{2.11}$$

$$\pi_1^{YN}|I \;=\; \begin{cases} S^{'2}\left[\frac{2S^2g^{'}-\gamma\mu_y^2 g}{4S^2S^{'2}-\gamma^2\mu_y^4} + \frac{\gamma\rho\mu_y\sigma_m(I-\mu_I)}{2\sigma_I S^{'2}}\right]^2, & if\ C_y^2 > \frac{\gamma g^{'}}{2g} - \gamma^2, \\ \frac{[g^{'}\sigma_I + \gamma\rho\mu_y\sigma_m(I-\mu_I)]^2}{4\sigma_I^2 S^{'2}}, & otherwise, \end{cases} \tag{2.12}$$

$$\pi_2^{YN} \;=\; \begin{cases} \frac{S^2(2S^{'2}g-\gamma\mu_y^2 g^{'})^2}{(4S^2S^{'2}-\gamma^2\mu_y^4)^2}, & if\ C_y^2 > \frac{\gamma g^{'}}{2g} - \gamma^2, \\ 0, & otherwise. \end{cases} \tag{2.13}$$

When farmer 1 is the only farmer who benefits from utilizing market information (via ρ) and from adopting agricultural advice (via α, β, and γ), Lemma 2 reveals that farmer 1 can afford to force farmer 2 to exit the market when the process yield uncertainty is sufficiently low; i.e., when $C_y^2 \leq \frac{\gamma g^{'}}{2g} - \gamma^2$.

By using Lemma 2, we get the following proposition.

Proposition 3. *The ex ante equilibrium outcomes of farmer 1 associated with strategy* $(Y,\, N)$ *satisfy*

$$q_1^{YN} \;=\; \begin{cases} \frac{2S^2g^{'}-\gamma\mu_y^2 g}{4S^2S^{'2}-\gamma^2\mu_y^4}, & if\ C_y^2 > \frac{\gamma g^{'}}{2g} - \gamma^2, \\ \frac{g^{'}}{2S^{'2}}, & otherwise, \end{cases} \tag{2.14}$$

$$\pi_1^{YN} \;=\; \begin{cases} \frac{S^{'2}(2S^2g^{'}-\gamma\mu_y^2 g)^2}{(4S^2S^{'2}-\gamma^2\mu_y^4)^2} + \frac{\gamma^2\rho^2\mu_y^2\sigma_m^2}{4S^{'2}}, & if\ C_y^2 > \frac{\gamma g^{'}}{2g} - \gamma^2, \\ \frac{g^{'2}+\gamma^2\rho^2\mu_y^2\sigma_m^2}{4S^{'2}}, & otherwise. \end{cases} \tag{2.15}$$

The ex ante equilibrium outcomes of farmer 2 associated with strategy $(Y,\, N)$ *are* *given in Lemma 2, as stated in* (2.11) *and* (2.13)*.*

Proposition 3 has the same interpretation as Lemma 2: in equilibrium, farmer 1 can afford to force farmer 2 to exit the market when the process yield uncertainty is sufficiently low.

Below we compare the equilibrium outcomes of farmer 1 (who utilizes and adopts) and that of farmer 2 (who does not utilize or adopt).

Corollary 3. *When only farmer 1 chooses to invest, then* $q_1^{YN} \geq q_2^{YN}$ *if* $\gamma = 1$*.*

Corollary 3 reveals that farmer 1 will produce more than farmer 2 when the adoption of agricultural advice will not increase the process yield ($\gamma = 1$). However, it is not always true that farmer 1 will produce more. This is the case especially when farmer 1 is concerned about over supply that will drive down the market price. When $\gamma > 1$, farmer 1 can process less input than farmer 2 and yet generate a higher output than farmer 2. Therefore, farmer 1 may produce less.

Finally, by symmetry, we can use Proposition 3 to obtain the equilibrium outcomes associated with strategy (N, Y), where $q_1^{NY} = q_2^{YN}$, $q_2^{NY} = q_1^{YN}$, $\pi_1^{NY} = \pi_2^{YN}$, and $\pi_2^{NY} = \pi_1^{YN}$.

In summary, we have determined each farmer's expected profit in equilibrium associated with each subgame. By using these expected profits, we can specify the payoffs associated with the 2x2 meta-game depicted in Table 2.2. We now proceed to solve this meta-game to examine the conditions under which a farmer will utilize market information (or adopt agricultural advice).

2.4.2 Meta-game Analysis

By examining each farmer's expected profit in equilibrium associated with subgames (N, N), (Y, Y), (Y, N), and (N, Y) in Propositions 1, 2, and 3, we can establish the following lemma via direct comparison.

Lemma 4. *The farmer's expected profits associated with subgames* (N, N)*,* (Y, Y)*,* (Y, N)*, and* (N, Y) *possess the following properties.*

1. $\pi_1^{YN} (= \pi_2^{NY}) \geq \pi_1^{YY} (= \pi_2^{YY}) \geq \pi_1^{NN} (= \pi_2^{NN}) \geq \pi_1^{NY} (= \pi_2^{YN})$.

2. $\pi_1^{YN} - \pi_1^{NN} (= \pi_2^{NY} - \pi_2^{NN}) \geq \pi_1^{YY} - \pi_1^{NY} (= \pi_2^{YY} - \pi_2^{YN})$.

By noting from Lemma 4 that farmer 1's expected profit satisfies $\pi_1^{YN} \geq \pi_1^{NN}$ and $\pi_1^{YY} \geq \pi_1^{NY}$ (and similarly for farmer 2), we can conclude that a farmer can always increase her expected profit by utilizing market information (or adopting agricultural advice) regardless of the strategy selected by the other farmer. Also, by noting that $\pi_1^{YN} - \pi_1^{NN} \geq \pi_1^{YY} - \pi_1^{NY}$ and $\pi_2^{NY} - \pi_2^{NN} \geq \pi_2^{YY} - \pi_2^{YN}$, we can conclude that by utilizing market information (or adopting agricultural advice), the increase in the expected profit of a farmer is higher when the other farmer chooses not to utilize market information (or not to adopt agricultural advice).

Lemma 4 reveals that without considering the upfront investment K associated with the adoption of agricultural advice, a farmer can always benefit from utilizing market information (or adopting agricultural advice). This observation enables us to compare the expected payoffs shown in Table 2.2 by using the inequalities established in Lemma 4. By doing so, we can solve the meta-game as follows.

Corollary 5. *Without considering the upfront investment* K *associated with the adoption of agricultural advice, strategy* (Y, Y) *is the unique equilibrium; i.e., both farmers utilize market information (or adopt agricultural advice) in equilibrium.*

Knowing that market information (or agricultural advice) is beneficial to each farmer when the farmers engage in Cournot competition under both demand and process yield uncertainty, Corollary 5 is a natural consequence, that is, both farmers will utilize market information (or adopt agricultural advice) in equilibrium.

2.4.3 Welfare Improvement

When both farmers utilize market information (or adopt agricultural advice) in equilibrium, the government (or NGO) can measure the farmers' welfare improvement due to the provision of market information (or agricultural advice) against the base case under strategy (N, N). In this case, we define the farmers' welfare improvement according to the term $(\pi_1^{YY} - \pi_1^{NN}) + (\pi_2^{YY} - \pi_2^{NN})$. Due to symmetry (i.e., $\pi_1^{YY} = \pi_2^{YY}$ and $\pi_1^{NN} = \pi_2^{NN}$), it is sufficient for us to focus our analysis on $(\pi_1^{YY} - \pi_1^{NN})$. Comparing the expressions for the farmers' expected profits stated in Propositions 1 and 2, we get the following corollary.

Corollary 6. *Without considering the upfront investment K associated with the adoption of agricultural advice, market information (or agricultural advice) is welfare improving: $(\pi_i^{YY} - \pi_i^{NN}) > 0$ for $i = 1, 2$. Also, the welfare improvement $(\pi_i^{YY} - \pi_i^{NN})$ is decreasing in β, and increasing in α, γ, and ρ^2.*

Corollary 6 reveals that both farmers benefit from utilizing market information (or adopting agricultural advice) in equilibrium. Hence, market information is welfare improving. Without considering the upfront investment K associated with the adoption of agricultural advice, agricultural advice is also welfare improving. (We shall examine the effect of the upfront investment K on the farmers' welfare in Section 5.) It is intuitive to see that the welfare improvement $(\pi_i^{YY} - \pi_i^{NN})$ increases as the benefits associated with market information (via ρ^2) and agricultural advice (via α, γ, and β) increase.

To further investigate the interaction effects among quality improvement, cost reduction, process yield improvement, and forecast accuracy improvement, we establish the following corollary.

Corollary 7. *Without considering the upfront investment K associated with the adoption of agricultural advice, the welfare improvement $(\pi_i^{YY} - \pi_i^{NN})$, $i = 1, 2$,* *possesses the following properties.*

1. *The welfare improvement* $(\pi_i^{YY} - \pi_i^{NN})$ *is supermodular in* (α, γ) , (γ, σ_m) *and* (γ, ρ^2)*; i.e.,*

$$\frac{\partial^2(\pi_i^{YY} - \pi_i^{NN})}{\partial\alpha\partial\gamma} > 0, \ \frac{\partial^2(\pi_i^{YY} - \pi_i^{NN})}{\partial\gamma\partial\sigma_m} > 0 \ \textit{and} \ \frac{\partial^2(\pi_i^{YY} - \pi_i^{NN})}{\partial\gamma\partial\rho^2} > 0.$$

2. *There exist threshold points* $\bar{C}_y$ *and* $\bar{c}$ *such that when* $C_y/\gamma > \bar{C}_y$*, the welfare improvement* $(\pi_i^{YY} - \pi_i^{NN})$ *is always submodular in* (β, γ)*. But when* $C_y/\gamma \leq \bar{C}_y$*, the welfare improvement is submodular in* (β, γ) *if and only if* $c > \bar{c}$*.*

3. *The welfare improvement* $(\pi_i^{YY} - \pi_i^{NN})$ *is modular in* (α, ρ^2) *and* (β, ρ^2)*; i.e.,*

$$\frac{\partial^2(\pi_i^{YY} - \pi_i^{NN})}{\partial\alpha\partial\rho^2} = 0 \ \textit{and} \ \frac{\partial^2(\pi_i^{YY} - \pi_i^{NN})}{\partial\beta\partial\rho^2} = 0.$$

The first statement of Corollary 7 has the following implications. First, quality improvement (via α) and process yield improvement (via γ) are complementary, that is, they generate a "compounding effect" on the farmer's welfare. In this case, the process yield improvement results in a larger expected output, which intensifies the market competition, while the quality improvement leads to a larger market potential, thereby softening the market competition.

Second, process yield improvement (via γ) is more beneficial in terms of welfare improvement when market uncertainty is higher (i.e., when σ_m is larger) or when the market information is more accurate (i.e., when ρ^2 is larger). By noting that the term $\rho^2\sigma_m^2$ represents the reduction of variance when a farmer utilizes market information, this result implies that when the use of market information is more effective in improving the forecast accuracy (via $\rho^2\sigma_m^2$), the farmers have more incentives to improve the process yield.

Next, we examine the second statement of Corollary 4. Note that $C_y/\gamma \equiv \frac{\sigma_y}{\gamma\mu_y}$ represents the coefficient of variation of the "improved process yield" and that the unit cost reduces as β decreases. The second statement of Corollary 4 shows that cost reduction (via β) and process yield improvement (via γ) are complementary when: (1) the improved process yield is highly uncertain, or (2) the improved process yield is relatively stable but the unit cost c is sufficiently large. These results can be explained as follows. First, according to Proposition 2, the effect of yield improvement (via γ) on the production quantity q_i^{YY} is ambiguous. Positively, the yield improvement can enable each farmer to obtain the same output by reducing her production quantity. This indirect cost reduction causes both farmers to produce more. Negatively, the yield improvement may also intensify the market competition and drive down the market price, causing both farmers to produce less in equilibrium. The cost reduction (via β) is complementary to the positive effect but is substituted by the negative effect of yield improvement. Furthermore, from (4), we can easily show that $\partial^2 q_i^{YY}/\partial\gamma\partial c > 0$, which implies that when the unit cost is larger, q_i^{YY} is more likely to increase in γ. This also indicates that when the unit cost is larger, the positive effect tends to be stronger than the negative effect. Next, observing from the best response of both farmers under strategy (Y, Y), if one farmer produces one extra unit, the other's best response is to reduce her production quantity by $\gamma^2\mu_y^2/2S'^2$. This quantity change can measure the product substitution between the two farmers. We refer to this as a substitution factor. By noting that $\gamma^2\mu_y^2/2S'^2 = 1/2(C_y^2/\gamma^2 + 1)$ is decreasing in C_y/γ (the coefficient of variation of the improved process yield), we know that the more uncertain the process yield, the less fierce the market competition. Thus, when the process yield is highly uncertain (i.e., $C_y/\gamma > \bar{C}_y$) such that the market competition is relatively mild, the positive effect of yield improvement always dominates its negative effect. Therefore, the cost reduction and yield improvement are complementary. However, when the process yield is relatively stable, cost reduction and yield improvement are complementary if and only if the unit cost is large enough such that the positive effect of yield improvement can dominate its negative effect.

Finally, the third statement of Corollary 7 shows that the improvement in forecast accuracy has no effect on quality improvement or cost reduction. Therefore, the government prefers the combination of process yield improvement and demand forecast improvement over the combination of demand forecast improvement and cost reduction or quality improvement.

2.5 The Implications of Market Information and Agricultural Advice

Based on the unified approach and the corresponding analysis presented in Section 4, we now examine the implications of market information and agricultural advice separately.

2.5.1 Utilization of Market Information

In this section, we consider the case in which the government only provides market information I that is intended to help farmers improve their production planning. Without the benefits associated with agricultural advice, we have $\alpha = \beta = \gamma = 1$. In this case, $S' = S$, $g' = g$ and $C_y^2 > 0 > \frac{\gamma g^{'}}{2g} - \gamma^2$. Based on Propositions 1, 2, 3, and Lemma 2, both farmers' expected profit under each subgame can be simplified as:

$$\pi_1^{NN} = \pi_2^{NN} = \frac{S^2 g^2}{(2S^2+\mu_y^2)^2},$$

$$\pi_1^{YY} = \pi_2^{YY} = \frac{S^2(g^2+\rho^2\mu_y^2\sigma_m^2)}{(2S^2+\mu_y^2)^2},$$

$$\pi_1^{YN} = \pi_2^{NY} = \frac{S^2 g^2}{(2S^2+\mu_y^2)^2} + \frac{\rho^2\mu_y^2\sigma_m^2}{4S^2}, \tag{2.16}$$

$$\pi_1^{NY} = \pi_2^{YN} = \frac{S^2 g^2}{(2S^2+\mu_y^2)^2}. \tag{2.17}$$

By direct comparison, we establish the following result.

Corollary 8. *When the government provides market information I only,*

1. *The farmer's expected profits satisfy $\pi_1^{YN}(=\pi_2^{NY}) > \pi_1^{YY}(=\pi_2^{YY}) > \pi_1^{NN} = \pi_1^{NY}(=\pi_2^{NN}=\pi_2^{YN})$. Furthermore, $\pi_1^{YN}-\pi_1^{NN}(=\pi_2^{NY}-\pi_2^{NN}) > \pi_1^{YY}-\pi_1^{NY}(=\pi_2^{YY}-\pi_2^{YN})$.*

2. *Both farmers use the market information in equilibrium to make better production planning decisions.*

Because "market information only" is a special case of the unified model, the above corollary is a sharper statement of Lemma 4 and Corollary 5. Next, the inequality $\pi_1^{YN}(=\pi_2^{NY}) > \pi_1^{YY}(=\pi_2^{YY})$ in Corollary 8 implies that the farmer with the market information has no incentives to share the market information with her competitors, which is consistent with the results obtained by Gal-Or (1985, 1986). As stated in Gal-Or (1985, 1986), firms gain from sharing private value (e.g., unit production cost) but lose from sharing common value (e.g., common demand). However, Gal-Or (1985, 1986) addresses the above issue from the viewpoint of profit maximization. In our context, the market potential is a common value to both farmers and we are interested in examining whether the government has incentives to distribute the market information in terms of welfare maximization. That is, is the market information welfare maximizing in the sense that the maximum total welfare of the farmers is attained when both farmers utilize the market information in equilibrium? Recall from Corollary 6 that market information is welfare improving: $(\pi_i^{YY} - \pi_i^{NN}) > 0$ for $i = 1, 2$. However, Corollary 8 reveals that strategy (Y, Y) is the equilibrium but $\pi_1^{YN} > \pi_1^{YY}$ and $\pi_2^{YN} < \pi_2^{YY}$. Therefore, it remains unclear whether $(\pi_1^{YY} + \pi_2^{YY})$ dominates the farmers' total welfare associated with all other strategies. We now examine this question.

We first observe from Corollary 6 that $(\pi_1^{YY} + \pi_2^{YY}) > (\pi_1^{NN} + \pi_2^{NN})$. Due to the symmetry between strategies (Y, N) and (N, Y), we can conclude that market information is welfare maximizing if $(\pi_1^{YY} + \pi_2^{YY})$ is greater than $(\pi_1^{YN} + \pi_2^{YN})$. Given that $g = \mu_m \mu_y - c$ and $S^2 = \sigma_y^2 + \mu_y^2$, we can establish a simple condition under which market information is welfare maximizing: the maximum total welfare of the farmers is attained when both farmers use the market information in equilibrium under strategy (Y, Y).

Proposition 4. *When the government provides market information I only, the provision of this market information is welfare maximizing if and only if the coefficient of variation of the process yield* $\frac{\sigma_y}{\mu_y} \equiv C_y > \sqrt{\frac{\sqrt{2}-1}{2}}$.

Proposition 4 reveals that the market information is welfare maximizing when the regular process yield is highly uncertain. Therefore, if the uncertainty of the process yield is relatively high, even when the farmer with the market information prefers concealing this information, the government has incentives to distribute market information to both farmers to maximize the farmers' welfare. However, when the uncertainty of the process yield is relatively low, neither the government nor the farmer with the market information wants to reveal this information. A close look of the best responses under the four subgames shows that when the government only provides market information, the substitution factor between the two farmers' products is $\mu_y^2/2S^2$, which can be rewritten as $1/2(C_y^2+1)$. This implies that when the regular process yield is highly uncertain, the quantity competition is somewhat mild. In view of this, Proposition 4 actually reveals that the provision of market information is welfare maximizing if and only if the market competition is relatively mild.

Proposition 4 specifies the condition under which the government's provision of market information is welfare maximizing. However, one may wonder whether this result will hold when the products are heterogeneous (measured in terms of substitutability level). To examine this issue, we consider the following inverse demand function to capture product heterogeneity. Specifically, the price of farmer i's product P_i satisfies:

$$P_i = M - (z_i q_i + t z_j q_j),\ \ i,j = 1,2, i \neq j,$$

where t measures the level of substitutability between products. Without loss of generality, we assume that $t \in [0, 1]$ so that a low (high) value of t corresponds to the case in which the products are less (more) substitutable. (Note that the (homogeneous) products are perfect substitutes when $t = 1$.)

Corollary 9. *Suppose the farmers' products are heterogenous so that the market price becomes $P_i = M - (z_i q_i + t z_j q_j)$. Then:*

1. *The provision of market information is always welfare improving.*
2. *When $t < 2(\sqrt{2}-1)$, the provision of market information is always welfare maximizing.*
3. *When $t \geq 2(\sqrt{2}-1)$, the provision of market information is welfare maximizing if and only if $C_y > \sqrt{\frac{(\sqrt{2}+1)t-2}{2}}$.*

Analogous to Proposition 4 associated with the homogenous product case, the first statement of Corollary 6 reveals that when products are heterogeneous, the provision of market information is still welfare improving. In regard to welfare maximization, the second and third statements of Corollary 6 generalize the result stated in Proposition 4 that corresponds to the case when $t = 1$. When t decreases, the products become less substitutable, competition becomes less fierce, and the strategies chosen by the two farmers are less correlated. Therefore, when the level of substitutability is sufficiently low (i.e., when $t < 2(\sqrt{2} - 1)$), one farmer's strategy has little effect on the profit of the other farmer. Consequently, the provision of market information can be welfare-maximizing. Nevertheless, when the product substitutability is relatively high (i.e., $t \geq 2(\sqrt{2} - 1)$), competition becomes more fierce, and the farmers' strategies are more correlated. In this case, we obtain a similar result to that stated in Proposition 4: market information can maximize the total welfare of the farmers if and only if the process yield is highly uncertain, in which case the market competition is relatively soft.

In summary, when the government provides market information I only, we find that both farmers use the market information under strategy (Y, Y) in equilibrium. Also, we show that the market information is certainly welfare improving. However, the market information is welfare maximizing in the sense that the maximum total welfare of farmers is attained under strategy (Y, Y) if and only if the process yield is highly uncertain. In addition, we generalize our model by considering product heterogeneity. We show that if the level of product substitutability is low, the provision of market information is always welfare maximizing. However, when the level of product substitutability is high, the provision of market information is welfare maximizing if and only if the process yield is highly uncertain.

2.5.2 Adoption of Agricultural Advice

In this section, we examine the case in which the government only offers agricultural advice that is intended to help farmers to improve quality, reduce cost, and improve process yield. Without the benefits associated with market information I, we have $\rho = 0$. Without considering the upfront investment K associated with the adoption of agricultural advice, we can use Propositions 1, 2, 3, and Lemma 2 to show that both farmers' expected profit under each strategy can be simplified as:

$$\pi_1^{NN} = \pi_2^{NN} = \frac{S^2 g^2}{(2S^2 + \mu_y^2)^2},$$

$$\pi_1^{YY} = \pi_2^{YY} = \frac{S'^2 g'^2}{(2S'^2 + \gamma^2 \mu_y^2)^2},$$

$$\pi_1^{YN} = \pi_2^{NY} = \begin{cases} \frac{S'^2 (2S^2 g' - \gamma \mu_y^2 g)^2}{(4S^2 S'^2 - \gamma^2 \mu_y^4)^2}, & if\ C_y^2 > \frac{\gamma g'}{2g} - \gamma^2, \\ \frac{g'^2}{4S'^2}, & otherwise. \end{cases}$$

$$\pi_1^{NY} = \pi_2^{YN} = \begin{cases} \frac{S^2 (2S'^2 g - \gamma \mu_y^2 g')^2}{(4S^2 S'^2 - \gamma^2 \mu_y^4)^2}, & if\ C_y^2 > \frac{\gamma g'}{2g} - \gamma^2, \\ 0, & otherwise. \end{cases}$$

By direct comparison and applying Lemma 4, we obtain the following proposition.

Proposition 5. *When the government offers agricultural advice only,*

1. *Without considering the upfront investment K, the farmers' expected profits satisfy $\pi_1^{YN}(= \pi_2^{NY}) > \pi_1^{YY}(= \pi_2^{YY}) > \pi_1^{NN}(= \pi_2^{NN}) > \pi_1^{NY}(= \pi_2^{YN})$. Furthermore, $\pi_1^{YN} - \pi_1^{NN}(= \pi_2^{NY} - \pi_2^{NN}) > \pi_1^{YY} - \pi_1^{NY}(= \pi_2^{YY} - \pi_2^{YN})$.*

2. *By incorporating the upfront investment K, the equilibrium strategy of the*

meta-game as depicted in Table 2.2 can be characterized as follows.[11]

$$Equilibrium = \begin{cases} (Y,Y), & if\ K < \pi_1^{YY} - \pi_1^{NY}, \\ (Y,N)\ and\ (N,Y), & if\ \pi_1^{YY} - \pi_1^{NY} \le K \le \pi_1^{YN} - \pi_1^{NN}, \\ (N,N), & if\ K > \pi_1^{YN} - \pi_1^{NN}. \end{cases}$$

Proposition 5 shows that the upfront investment K has a direct effect on whether a farmer adopts agricultural advice in equilibrium. Specifically, both farmers adopt agricultural advice when the upfront investment K is low (i.e., when $K < \pi_1^{YY} - \pi_1^{NY}$), and no farmer adopts agricultural advice when the upfront investment K is high (i.e., when $K > \pi_1^{YN} - \pi_1^{NN}$). When K is in a moderate range (i.e., when $\pi_1^{YY} - \pi_1^{NY} \le K \le \pi_1^{YN} - \pi_1^{NN}$), one farmer's best response is Y (N) if the other one's strategy is N (Y). Therefore, the equilibrium outcome is asymmetric and only one farmer will adopt agricultural advice by making the upfront investment K.[12]

Without considering the upfront investment cost K, Corollaries 6 and 7 reveal that agricultural advice is welfare improving when both farmers adopt agricultural advice in equilibrium under strategy (Y, Y). However, when incorporating the upfront investment K, Proposition 5 reveals that the equilibrium strategy hinges upon K. To examine whether agricultural advice is welfare improving when accounting for the upfront investment K, we consider the following scenarios.

First, consider the case in which K is high (i.e., when $K > \pi_1^{YN} - \pi_1^{NN}$), agricultural advice does not improve the farmers' welfare because no farmer will adopt the advice.

Second, when K is moderate (i.e., when $\pi_1^{YY} - \pi_1^{NY} < K < \pi_1^{YN} - \pi_1^{NN}$), we have two pure equilibria (Y, N) and (N, Y). Due to symmetry, it suffices to examine

[11] Due to symmetry, it suffices to state only the conditions that are based on the expected profits of farmer 1.

[12] Here, the game becomes a coordination game and has two pure Nash equilibria and one mixed Nash equilibrium (Fudenberg and Tirole 1991). In the mixed Nash equilibrium, farmer i, $i = 1, 2$, chooses to invest with probability $\frac{\pi_1^{YN} - \pi_1^{NN} - K}{\pi_1^{NY} - \pi_1^{YY} + \pi_1^{YN} - \pi_1^{NN}}$.

whether agricultural advice is welfare improving when only farmer 1 adopts the advice in equilibrium under strategy (Y, N). In other words, will $(\pi_1^{YN} - K) + \pi_2^{YN} > \pi_1^{NN} + \pi_2^{NN}$ when K satisfies $\pi_1^{YY} - \pi_1^{NY} < K < \pi_1^{YN} - \pi_1^{NN}$? Note that $\pi_2^{YN} = \pi_1^{NY}$ and $\pi_2^{NN} = \pi_1^{NN}$. The condition $(\pi_1^{YN} - K) + \pi_2^{YN} > \pi_1^{NN} + \pi_2^{NN}$ can be simplified as $\pi_1^{YN} + \pi_1^{NY} - 2\pi_1^{NN} > K$. Combining this simplified condition and the range within which K lies (i.e., $\pi_1^{YY} - \pi_1^{NY} < K < \pi_1^{YN} - \pi_1^{NN}$) along with $\pi_1^{NN} > \pi_1^{NY} (= \pi_2^{YN})$ (see the first statement of Proposition 5), we can conclude that agricultural advice is welfare improving under strategy (Y, N) if and only if the value of K falls within the range $\pi_1^{YY} - \pi_1^{NY} < K < \pi_1^{YN} + \pi_2^{YN} - 2\pi_1^{NN}$. Note that this range exists only when $\pi_1^{YY} - \pi_1^{NY} < \pi_1^{YN} + \pi_2^{YN} - 2\pi_1^{NN}$, which may not hold in general.[13] Consequently, we can conclude that agricultural advice may not be welfare improving under strategy (Y, N) (and strategy (N, Y)).

Third, consider the case in which K is low (i.e., when $K < \pi_1^{YY} - \pi_1^{NY}$). In this case, Proposition 5 reveals that both farmers adopt agricultural advice under strategy (Y, Y). Hence, agricultural advice is welfare improving if and only if $(\pi_1^{NN} + \pi_2^{NN}) < (\pi_1^{YY} - K) + (\pi_2^{YY} - K)$. By symmetry, this condition can be simplified as $K < \pi_1^{YY} - \pi_1^{NN}$. Also, observe from the first statement of Proposition 5 that $\pi_1^{YY} - \pi_1^{NN} < \pi_1^{YY} - \pi_1^{NY}$. Combining the simplified condition with this observation, we can conclude that agricultural advice is welfare improving if and only if the upfront investment is sufficiently low (i.e., when $K < \pi_1^{YY} - \pi_1^{NN}$).

Based on the implications of the aforementioned three scenarios, we establish the following statement.

Corollary 10. *Depending on the upfront investment K, agricultural advice is not*

[13] To elaborate, given that $\pi_1^{NY} = \pi_2^{YN}$, the condition $\pi_1^{YY} - \pi_1^{NY} < \pi_1^{YN} + \pi_2^{YN} - 2\pi_1^{NN}$ can be simplified as $2(\pi_1^{NN} - \pi_1^{NY}) < \pi_1^{YN} - \pi_1^{YY}$, where the right hand side represents the net gain of farmer 1 by being the only one adopting the agricultural advice instead of both farmers utilizing the agricultural advice, and the left hand side represents the net gain of the farmers when they choose not to adopt the agricultural advice. Therefore, when the benefits associated with the agricultural advice are small (via small γ, α , ρ^2, or large β), this condition may not hold. For example, it can be checked that $\pi_1^{YN} - \pi_1^{YY} - 2(\pi_1^{NN} - \pi_1^{NY})$ is negative when $\mu_m = 20$, $c = 3$, $\alpha = 1.3$, $\beta = 0.7$, $\rho = 0.4$, $\sigma_y = 0.7$, $\sigma_m = 0.2$, $\mu_y = 0.7$, and $\gamma < 1.05$.

necessarily welfare improving. However, to ensure that adopting agricultural advice (Y, Y) *improves farmers' total welfare, the government should consider offering subsidies so that the "effective" upfront investment to be borne by each farmer is below the threshold* $(\pi_1^{YY} - \pi_1^{NN})$.

Corollary 10 reveals that offering agricultural advice alone is not sufficient to ensure that the total welfare of farmers will be improved unless either the upfront investment K is sufficiently low (i.e., $K < \pi_1^{YY} - \pi_1^{NN}$) or the government offers subsidies so that the "effective" upfront investment to be borne by each farmer is below the threshold $\pi_1^{YY} - \pi_1^{NN}$. This result may help justify the farmer subsidies offered in developing countries.[14]

2.6 Conclusion

In this chapter, we presented a unified framework for analyzing the implications when the government offers market information that can help farmers to make better (long-term) production planning decisions (or agricultural advice that can help farmers to improve product quality, reduce production cost, and enhance process yield). By considering the case in which farmers engage in Cournot competition under both demand and process yield uncertainty, we showed that without considering the upfront investment, both farmers would utilize market information (or adopt agricultural advice) in equilibrium. We also showed the complementary effects associated with different benefits (quality improvement, cost reduction, and process yield increase).

We then used the results of our general model to analyze the case in which the

[14] For example, to encourage farmers to purchase various types of agricultural equipment to help reduce their production costs, the Department of Agriculture and Cooperation of India offers subsidies in the form of 50% of the equipment cost. See farmech.gov.in/FarmerGuide/BI/11.htm for details. In another example, to improve quality and process yield, small Kenyan farmers can purchase fertilizers from the government owned National Cereals and Produce Board at subsidized prices (i.e., 32% below market price). See http://partnews.brownbag.me/2013/04/15/kenyan-subsidized-fertilizer-explained/ for details.

government only offers market information. We found that both farmers would utilize the market information in equilibrium and that the market information is welfare improving. Moreover, the market information is welfare maximizing when the process yield is highly uncertain.

For the case in which the government only offers agricultural advice, we showed that each farmer will adopt agricultural advice in equilibrium only when the upfront investment is below a certain threshold. We also found that agricultural advice is not necessarily welfare improving. To ensure that agricultural advice can improve the total welfare of farmers, we showed that farmer subsidies are essential especially when upfront investment is high.

Our work is an initial attempt to examine the implications of market information on agricultural advice. However, our model can be extended to the case in which the government offers both market information and agricultural advice, although the analysis becomes intractable because it involves the comparison of 16 different expected profits. Nevertheless, our approach can enable us to analyze this case numerically. Furthermore, our results can shed some light on the interdependencies between market information and agricultural advice. For example, Corollary 7 states that more accurate market information can enhance the effect of yield improvement but has no influence on the effects of quality improvement and cost reduction. Therefore, the government prefers to subsidize the farmers to improve their process yield rather than improve quality and reduce cost when the market information is accurate.

Besides providing market information about future market prices that can help farmers to make long term production planning decisions, various governments are now offering current market price information that can help farmers to make short-term selling decisions (when and where to sell). It is of interest to analytically examine the implications of utilizing the current market price information for farmers in emerging markets especially because the relevant empirical findings have been mixed. For example, Mittal et al. (2010) find that by utilizing current market price information, farmers enjoy higher incomes. However, Fafchamps and Minten (2011) find no evidence supporting this claim under a different experimental setting.

In this study, we focus on the case in which each farmer either adopts all of the agricultural advice by making a full investment or adopts none of the advice. However, different types of advice may require different amounts of investment. Therefore, in the event that each farmer can choose to adopt a particular subset of advice by making the requisite investment, the analysis quickly becomes tedious because the number of subsets of advice grows exponentially. For this reason, we defer this issue as a topic of future research.

In this study, we examined the benefits of farm subsidies in the context of developing countries, where farmers tend to be poor and have little access to financial services and formal training. This context is drastically different from that of developed countries where farmers are relatively wealthy, powerful, and well-trained (Smith 2012). In this case, the farmers can obtain financial support through different financing channels (see http://smallfarm.about.com/od/otherresources/a/farmgrants.htm), and have some control over the market price. For example, in 2013, farmers in Australia strategically held back on wheat sales to maintain a high market price (Thukral and Packham 2014). Currently, there is an on-going debate over whether the government should offer subsidies to farmers in developed countries (Edwards 2009).[15] Because the contexts are very different, there is a need to develop a different model to investigate the value of farm subsidies in developed economies, and we leave this question for future research.

[15] In the United States, the government offers millions of dollars in subsidies to domestic farmers and agribusinesses to supplement their income and help them manage their production and maintenance costs.

Chapter 3

Why Minimizing Waiting Time in Health Care System Could be Bad?

3.1 Introduction

One of the most significant problems faced by the health care systems is the long waiting time. For example, patients in Canada often wait for 18.2 weeks on average to receive medical treatment (Humphreys 2013). In Sweden, the average waiting time for prostate cancer is almost eight months (Bylund 2014). In 2013, more than 21,000 patients in Hong Kong had waited three years before receiving specialist treatment (Tsang 2014). Undoubtedly, excessive waiting time for health care can cause adverse health effects such as low health care access and patient dissatisfaction, which may generate disutility among the patients and hurt the welfare of the society. For instance, due to the long waiting time, nearly 5 million Canadians do not have access to primary health care services.[1] In China, the long waiting time makes doctors have no enough time to properly examine the patients.[2] And it is frequently reported that patients in Sweden with stroke, heart failure and other serious medical conditions are denied or unable to receive proper treatment (Bylund 2014). Furthermore, according to Sanmartin et al. (2010), long waiting times increase patients' stress, anxiety and pain, which in turn strains patient-doctor relationships and damages public perceptions of the health care system.

Given these negative consequences, the health care waiting time has gained popu-

[1] See "Canada ranked last among OECD countries in health care wait times", reported by CTVNews, http://www.ctvnews.ca/health/canada-ranked-last-among-oecd-countries-in-health-care-wait-times-1.1647061#ixzz3EawhlNRm

[2] See http://www.chinaeconomicreview.com/node/27471.

larity as a measurement of a health care system's congestion level and performance. For instance, New Zealand's government uses the Emergency Department (ED) waiting time as a measure to evaluate hospital performance (Jones et al. 2012). Waiting time for the elective surgery is a key performance indicator for hospitals in the United Kingdom (Dimakou 2013). For those countries, reducing waiting times and improving the accessibility of health care services are the long-term goals being pursued. In fact, several policies have been developed to force health care providers (HCPs) to reduce patients' waiting times below a specific target. For example, since 2009, New Zealand's government has required hospitals to reduce their ED waiting times below 6 hours (Jones et al. 2012). In another instance, the British government has set up an 18-week waiting time target, and hospitals that do not start treatment within 18 weeks will be penalized (Winnett 2011).

Does a waiting-time reduction target really solve congestion problem and improve the accessibility of health care services? Critics believe that such targets merely shift the systemic problems instead of solving them. HCPs may find themselves drawing resources and attention away from other aspects of performance measurements, such as health care quality, to just meet waiting time targets. Thus, improvement in the congestion level and the accessibility of health care services may jeopardize other system performances. Specifically, without capacity expansion, the only way for a HCP to reduce patients' waiting time is to spend less time on each patient. Given that a short length of stay (each visit) normally leads to a high rate of readmission (Kociol et al. 2012), this can intensify the congestion in the health care system in the long run. For instance, Winnett (2011) notes that the waiting time target adopted in the UK actually pressured HCPs into discharging patients early, which led to a rise of 31% in the readmission rate from 2006 to 2011. According to Purdy et al. (2012), 35% of hospital admissions in England were readmissions, at a cost of $11 billion in 2011.

Such empirical evidences cast doubt on the conventional wisdom that reducing waiting time can mitigate health care system congestion, questioning that waiting time may not be a proper performance measurement. This motivates us to investigate the following research question: How should we evaluate the performance of a health care system if we take the tradeoff between service speed and service quality (measured by the readmission rate) into consideration. Without differentiating between the new and readmitted patients, one may be misled by the "beautiful" data on a health care system: reduced waiting time per admission, higher accessibility for patients and a higher utilization rate of the HCP. We shall be cautious on those system performance measurements due to the following issues. First, when the waiting time per admission is reduced, is it associated with a higher readmission rate (of old patients)? Second, when the service accessibility is improved, is it associated with a smaller fraction of new patients being treated by the HCP? Third, when the HCP's utilization rate increases, does the HCP spend the time mainly on treating new patients or on treating readmitted patients? To scrutinize these questions, we model a health care system as an M/M/1 queue with Bernoulli feedback, where the feedback represents the readmission rate. The tradeoff between service speed and service quality is captured by assuming that the readmission rate is a decreasing function of service time; that is, if the HCP spends more time on a patient, the probability of relapse and readmission for that patient shall be lower. Compared with the classic queueing models without feedback (e.g., Hassin and Haviv 2003, Anand et al. 2011), this model exhibits the following unique feature: customers in the classic queueing models only request service once while customers in our setting may demand multiple rounds of service.

We then investigate the performance measures that can capture both system congestion and service quality. One natural candidate for the service quality perhaps is the *effective service rate*, which is the mean number of patients cured by the HCP in a unit of time. A two dimensional measurement vector, (*waiting time, effective service rate*), can then provide the comprehensive information on the system: the former one measures the congestion level and the latter one the service quality. However, they are somehow cumbersome. We find that both the congestion level and service quality can actually be well captured by a single measurement, that is, the *total waiting time*, which is defined as the total amount of time a patient spends in the system before being cured. The social planner only needs to trace this performance measure in regulating the health care system. To avoid confusion, we refer the waiting time that a patient spends per admission as the *waiting time per admission*. We also introduce the following measurement, *utilization rate for new patients* to measures how much time the HCP spends on treating new patients. By examining the relationship between these new performance measures with those ordinary ones, we show that some of our conventional understandings on the health care system shall be carefully examined.

1. Misunderstanding one: setting a waiting time reduction target mitigates system congestion. We show that reducing waiting time per admission may incentivize the HCP to increase their service speed that leads to a higher readmission rate and a longer total waiting time.

2. Misunderstanding two: keeping doctors busier can serve more patients and hence increase the service accessibility. We show that a reduction in the HCP's idle time may be mainly attributed to the increase in the number of readmitted patients. Therefore, reducing the HCP's idle time may reduce the accessibility of the health care services and cause the HCP to spend less time on treating new patients.

The above misunderstandings justify the need of adopting the total waiting time as a new measurement. We then investigate the health care system's optimal design and control issues based on this new measure. We consider the following three types of systems, according to the roles of different parties in making queueing, service rate and price decisions.

- Benchmark system: the social planner takes a direct control over the effective arrival rate and the service rate. This model generates a *first-best* result for the service rate decision and serves as a benchmark case.
- Decentralized system: the HCP makes decisions over both the price and the service rate; patients observe such decisions and make their own queueing decision, i.e., to join or to balk. In short, both the price and service rate are set to maximize the profit of the HCP. We call the resulting price decision as a *profit-maximizing price.*
- Regulated system: the social planner determines the price; the HCP must follow such price and then makes its service rate decision; patients observe the price and service rate and then make their queueing decision. In this case, the health care system is subject to a price regulation. We call the resulting price decision a *welfare-maximizing price.*

We first consider the benchmark case and obtain the first-best solution that can serve as a benchmark to evaluate the performance of the health care system. To obtain the equilibrium outcomes associated with the decentralized and regulated cases, we first study the patients' queueing decision and the HCP's service rate decision for a fixed price. Then we derive the optimal price from the perspectives of the HCP and the social planner, respectively. Our equilibrium analysis enables us to obtain the following interesting results.

1. The first-best service rate coincides with the one maximizing the effective service rate and the one minimizing the total waiting time. Therefore, the

total waiting time indeed serves as a good performance measure for the service quality.

2. A higher price always reduces the total waiting time but it does not hold for the waiting time per admission. Interestingly, a lower price can sometimes reduce the waiting time per admission. This conclusion is insightful for a policy maker: if the congestion is mistakenly measured by the waiting time per admission, one might be misled to lower down the price instead of increasing it, which can make the system more congested and harm the social welfare.

3. By comparing the decisions under the decentralized and regulated cases, we find that under no price regulation, the HCP will choose a first-best service rate (socially optimal); under the price regulation, the HCP will serve with a service rate higher than the socially optimal one. Consequently, the single pricing control fails to achieve the first-best outcome.

4. Finally we show that price regulation plus a penalty mechanism based on the readmission rate is sufficient for the policy maker to regulate the system to achieve the first-best outcome.

The remainder of this chapter is organized as follows. Section 3.2 reviews the relevant literature. Section 3.3 introduces the model setup and three new system performance measurements: the effective service rate, the total waiting time and the utilization rate for new patients. We then solve the first-best solution in Section 3.4. In Section 3.5, we derive the equilibrium arrival rate of new patients, the HCP's optimal decision of service rate and the optimal pricing decision. In Section 3.6, we consider the scenario of small market potential arrival rate. Concluding remarks are provided in Section 3.7 and all proofs are relegated to the Appendix B.

3.2 Literature Review

Our study is related to the queueing literature with an endogenous service rate. Hopp et al. (2007) consider operations systems with discretionary task completion times in which the customer value is linked to the service rate. They find that capacity expansion may actually intensify congestion, whereas task variability can improve system performance. Debo et al. (2008) examine the incentives of an expert to induce service; that is, to provide unnecessary service to customers. They show that while service inducement reduces the total welfare, it also enables the expert to obtain a large share of the total welfare. Whether service inducement is profitable depends on which effect dominates. Anand et al. (2011) investigate the equilibrium joining and pricing strategies in the customer-intensive services in which customers' value is decreasing in the service rate. Kostami and Rajagopalan (2014) analyze the quality-speed tradeoff in a dynamic setting. Tong and Rajagopalan (2014) compare the fixed fee and time-based fee schemes in discretionary service. They explore the conditions under which the two pricing schemes outperform each other. Chan et al. (2014) develop a fluid model to examine the effect of speedup on system performance by considering customer return. The queueing model with endogenous service time is also applied in other domains, such as diagnostic services (Paç and Veeraraghavan 2010, Wang et al. 2010, Alizamir et al. 2013), service quality variability (Xu et al. 2012), call center (de Vericourt and Zhou 2005, Hasija et al. 2009) and health care staffing (Yom-Tov and Mandelbaum 2014). Among those papers, only those of de Vericourt and Zhou (2005), Chan et al. (2014) and Yom-Tov and Mandelbaum (2014) consider returning customers. However, even in those studies, the return probability is independent of the service rate, while in our model, the readmission rate is a function of the service rate.

In recent years, many researchers apply operations research tools to study health care problems. Many of such work focus on scheduling issues, e.g., Hassin and

Mendel (2008), Liu et al. (2010), Luo et al. (2012), Mills et al. (2013), Liu and Ziya (2014), Feldman et al. (2014). In addition, there is a growing body of literature on health care system performance. For example, So and Tang (2000) examine the effects of a reimbursement policy for drug prescriptions and derive the optimal prescription policy. Fuloria and Zenios (2001) show that the delivery of medical services is most efficient when the provider's reimbursement is adjusted based on observed patient outcomes. Lee and Zenios (2012) design an evidence-based payment system with risk adjustment for renal dialysis services. Jiang et al. (2012) study the performance-based contract, by which the HCP is penalized if the patients' waiting time is larger than a target for outpatient services. Guo et al. (2013) investigate the efficiency of the conditional and unconditional subsidy schemes in health care systems. However, these works do not consider the readmission rate in their models.

Our work is also related to the literature that study the readmission rate in health care industry. Most of the works in this stream of literature are empirical such as Friedman and Basu (2004), Chollet et al. (2011), Chan et al. (2012), KC and Terwiesch (2012) and Kim et al. (2013). To our best knowledge, no previous works theoretically investigate the tradeoff between service speed and readmission rate.

3.3 Model Setup and Preliminaries

We list the notations used in the chapter in Table 3.1.

We model the health care system as an M/M/1 queue with Bernoulli feedback. Patients arrive at the HCP according to a Poisson process with a rate of Λ. We assume that patients in our model are homogeneous and incur a waiting cost of θ per unit of time spent in the system. This assumption is reasonable given that in many countries such as France, Germany and the US, patients are classified into

Table 3.1: Summary of Notations

Notation	Definition
R	reward received by the cured patient
θ	waiting cost per unit time
p	price per admission
β	penalty cost incurred from each balking patient
Λ	potential arrival rate
λ	effective arrival rate of new patients
λ_b	the socially optimal effective arrival rate of new patients
$\lambda_T = \frac{\lambda}{1-\delta(\mu)}$	total effective arrival rate (including new and readmitted patients)
μ	service rate
μ^*	the socially optimal service rate
$\delta(\mu)$	readmission rate
$o(\mu) = \mu(1 - \delta(\mu))$	effective service rate
$n(\mu) = \frac{1}{1-\delta(\mu)}$	the average number of HCP visits per patient
$W(\lambda, \mu) = \frac{1-\delta(\mu)}{o(\mu)-\lambda}$	waiting time per admission, including waiting time and service time
$TW(\lambda, \mu) = \frac{1}{o(\mu)-\lambda}$	the expected total waiting time
$\rho_N(\lambda, \mu) = \frac{\lambda}{\mu}$	utilization rate for new patients
$\rho_T(\lambda, \mu) = \frac{\lambda}{o(\mu)}$	total utilization rate
$\mu_{\mathcal{W}}$ and $\mu_{T\mathcal{W}}$	the service rates that minimize $W(\lambda, \mu)$ and $TW(\lambda, \mu)$ respectively (for a fixed λ)
μ_{an}, μ_{at}, μ_{un}, μ_{ut}	the service rates that maximize $\lambda(\mu, p)$, $\lambda_T(\mu, p)$, $\rho_N(\lambda(\mu, p), \mu)$ and $\rho_T(\lambda(\mu, p), \mu)$ respectively (for a fixed p)

different diagnosis-related groups according to their respective symptoms, and the patients in the same group demand similar resources and services (e.g., Street et al. 2011). The HCP serves the patients on a first-come-first-serve (FCFS) basis and each treatment takes an exponentially distributed time with a rate of μ. Note that both the Poisson arrival process and exponential service time have been well-tested in the health care literature. For instance, Kim et al. (1999) empirically verify that the arrival process to a hospital intensive care unit is a Poisson process, and the service time follows an exponential distribution.

After a patient is discharged from the HCP, he/she is either *cured* and leaves the system for good or he/she is *readmitted* to the system in the near future.[3] For tractability, we make the following assumptions about the readmission process:

[3]In practice, a readmission is defined by an event when a patient is readmitted to the HCP that occurs within a defined time window (e.g., 28 days in the UK and 30 days in the US) after discharge.

Assumption 1: The relapse period (i.e., the time between discharge and readmission) follows an exponential distribution and the tail probability of the relapse period being longer than the defined time window (e.g., 28 days in the UK and 30 days in the US) is negligible.

Assumption 2: The readmission rate $\delta(\mu)$ is increasing in the service rate μ, where $\delta(\mu) \in [0, 1]$.

Assumption 3: The cure rate $(1 - \delta(\mu))$ is logconcave in μ; i.e., $\log(1 - \delta(\mu))$ is concave so that $g(\mu) = \delta^{'}(\mu)/(1 - \delta(\mu))$ is increasing in μ.

Assumption 1 is reasonable given that Sibbritt (1995) empirically shows that the number of relapsed patients declines exponentially and tail-off to a background level of noisy after almost 28 days. Similar empirical results have been found by Heggestad and Lilleeng (2003) and Glynn et al. (2011). Assumption 2 captures an empirical fact that the readmission rate is increasing in the service rate μ (Kociol et al. 2012). By noting that the elasticity of the cure probability $(1 - \delta(\mu))$ equals $\mu g(\mu)$, assumption 3 guarantees that the readmission rate is more sensitive to the change in service rate when the service rate is larger.[4] Observe that the logistic function $\delta(\mu) = 1/(1 + e^{-a\mu+b})$ with parameters a and b satisfy assumptions 2 and 3, where the logistic function is a standard approach to measure the relationship between the readmission rate and other variables in the healthcare management literature (e.g., Fethke et al. 1986, Morrow-Howell and Proctor 1993). Based on above assumptions, we can model the health care system as a simple Jackson queueing network (Figure 3.1).

[4]The logconcave function is fairly general as it can be used to capture the shape of many probability distributions including Uniform, Normal, Exponential and Weibull (Bagnoli and Bergstrom 2005).

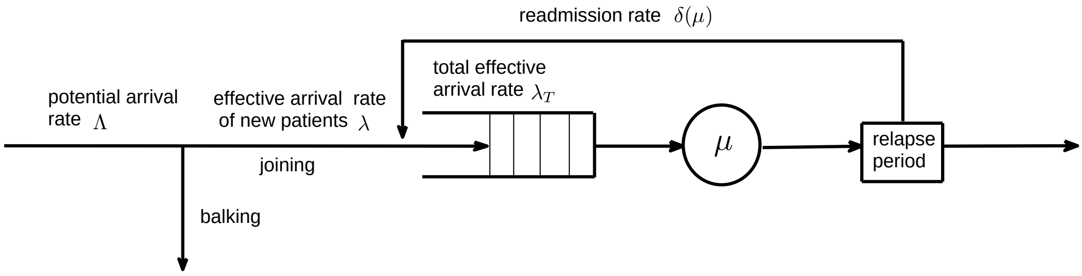


Figure 3.1: A Schematic of The Model

3.3.1 Performance Measurements

The readmission rate enables us to evaluate the service delivery, the congestion levels, the accessibility and the utilization rate of the health care system in two different ways. On one hand, we can measure them for all the patients, as the classic queueing models such as Hassin and Haviv (2003) do. On the other hand, we can also measure them only for the new patients, generating three new performance measurements: the effective service rate, the total waiting time and the utilization rate for new patients.

Service Rate vs. Effective Service Rate

Unlike the queueing system without feedback, due to the readmission risk, the patients that have being treated may not eventually leave the system. Therefore, in a time unit, the number of treated patients is larger than the number of cured patients. In view of this, we define the *effective service rate* as the mean number of patients cured by the HCP in a unit of time. By noting that the probability that a patient is cured equals $1 - \delta(\mu)$ and μ is the average number of patients treated by the HCP per unit time, the corresponding effective service rate can be denoted as

$$o(\mu) = \mu(1 - \delta(\mu)). \tag{3.1}$$

We have the following result regarding $o(\mu)$.

Lemma 11. *$o(\mu)$ is quasi-concave in μ and the optimum μ^* that maximizes $o(\mu)$ uniquely solves*

$$\mu^* g(\mu^*) = 1. \tag{3.2}$$

Furthermore, $o(\mu)$ is concave in μ for $\mu \leq \mu^*$.

Lemma 1 shows that the effective service rate is unimodal in the service rate and the corresponding mode μ^* is the point such that the elasticity of cure probability $1 - \delta(\mu)$ equals to one. This is because an increase in service rate (speed of service) by the HCP has two opposite impacts on the effective service rate. One the one hand, a higher service rate leads to more treated patients per unit of time, but on the other hand, it deteriorates the service quality delivered by the HCP by reducing the likelihood that patient being cured per admission. When the cure probability $1 - \delta(\mu)$ is inelastic in the service rate such that $\mu g(\mu) < 1$, the service quality is less sensitive to the change of service rate. Therefore, the first effect dominates the second effect and an increase in service rate leads to a higher effective service rate. On the contrary, when $\mu g(\mu) > 1$, the cure probability is elastic in the service rate, which implies the service quality is very sensitive to the change in service rate. In this case, the second effect dominates the first effect and therefore a higher service rate results in a smaller effective service rate.

Waiting Time vs. Total Waiting Time

Due to the existing risk of relapse, a patient's waiting time per admission is smaller than the total amount of time he/she spends in the system. In view of this, we define the *total waiting time* as the total amount of time a patient spends in the system before being cured. A patient's total waiting time is determined by (1) the waiting time that a patient spends per admission and (2) the total number of HCP visits that a patient goes through to get cured. To avoid confusion, we call the waiting time that a patient spends per admission as the *waiting time per admission*.

As the number of HCP visits follows a binomial distribution, the probability that a patient visits the HCP exactly i times for him/her to be cured is

$$P(i) = \delta^{i-1}(\mu)(1 - \delta(\mu)).$$

Then the average number of HCP visits per patient can be derived as

$$n(\mu) = \sum_{i=1}^{\infty} iP(i) = \frac{1}{1 - \delta(\mu)}, \tag{3.3}$$

which is increasing in the readmission rate $\delta(\mu)$ and thus increasing in service rate μ.

Let λ and λ_T denote the effective arrival rate of new patients and the system's total effective arrival rate (i.e., the new patients plus the readmitted patients), respectively. For now, we assume that the potential arrival rate Λ is large enough such that some patients will leave without being treated; that is $\lambda < \Lambda$. This assumption is realistic as in practice, most of the health care systems are very congested. Nevertheless, we also discuss the scenario of a small potential arrival rate in section 3.6.

By using the PASTA property of a Jackson network (Jackson 1957), we know that the departure rate of the system equals the total effective arrival rate λ_T. Also, the total effective arrival rate equals to the sum of the arrival rate of new patients λ and the arrival rate associated with the readmissions (which is equal to $\delta(\mu) \cdot \lambda_T$). Therefore,

$$\lambda_T = \lambda + \delta(\mu)\lambda_T \quad \Rightarrow \quad \lambda_T = \frac{\lambda}{1 - \delta(\mu)} = n(\mu)\lambda, \tag{3.4}$$

which implies that the total effective arrival rate for the system equals the effective arrival rate of new patients times the average number of HCP visits a patient goes through to be cured. Then the expected waiting time per admission can be derived as

$$W(\lambda,\,\mu) = \frac{1}{\mu - \lambda_T} = \frac{1-\delta(\mu)}{o(\mu) - \lambda}. \tag{3.5}$$

Utilizing (3.3) and (3.5), we can derive the expected total waiting time of each patient as follows:

$$TW(\lambda,\,\mu) \;=\; n(\mu)W(\lambda,\,\mu) = \frac{1}{o(\mu) - \lambda}. \tag{3.6}$$

Notice that (3.6) is equal to the waiting time of the classic M/M/1 queue with an arrival rate of λ and a service rate of $o(\mu)$. Therefore, the effective service rate $o(\mu)$ in our model plays a similar role as the service rate in the classic M/M/1 queue. Actually, by combining (3.1) and (3.3), we can know that $1/o(\mu) = n(\mu)/\mu$, which represents the total amount of time the HCP spends on treating a patient until he/she is cured. Therefore, $1/o(\mu)$ is the *total service time.*

Given λ, we denote $\mu_{\mathcal{W}}$ and $\mu_{\mathcal{TW}}$ as the service rates that minimize $W(\lambda,\,\mu)$ and $TW(\lambda,\,\mu)$ respectively. Then we can show the following proposition.

Proposition 6. *Given λ, the following properties hold.*

1. *Both $W(\lambda,\,\mu)$ and $TW(\lambda,\,\mu)$ are quasi-convex in μ.*

2. *The service rate $\mu_{\mathcal{TW}}$ that minimizes $TW(\lambda,\,\mu)$ equals μ^*; the service rate $\mu_{\mathcal{W}}$ that minimizes $W(\lambda,\,\mu)$ uniquely solves*

$$\mu_{\mathcal{W}}\, g(\mu_{\mathcal{W}})\rho_T(\lambda,\,\mu_{\mathcal{W}}) = 1. \tag{3.7}$$

Furthermore, $\mu_{\mathcal{W}} > \mu_{\mathcal{TW}} = \mu^*$.

The quasi-convexity of the waiting time per admission and the total waiting time is driven by the tradeoff between service quality and service speed. On one hand, a higher service rate enables the HCP to treat more patients per unit of time, thereby mitigating system congestion. On the other hand, it leads to a poor care quality and a high readmission rate, resulting in a negative (congestion) externality on each admitted patient. Recall that service quality is more sensitive to the service rate when service rate is higher. The first (second) effect dominates the second (first) effect when the service rate is relatively low (high). Therefore, both the waiting time per admission and the total waiting time are first decreasing and then increasing in the service rate. However, because the total waiting time equals the waiting time per admission times the average number of hospital visits, the negative externality has a stronger impact on the total waiting time than that on the waiting time per admission. Therefore, the total waiting time achieves its minimum at a smaller service rate, compared to that for the waiting time per admission. Hence, as depicted in Figure 3.2, when the service rate is very small (i.e., $\mu < \mu^*$), a higher service rate leads to both a shorter waiting time per admission and a shorter total waiting time. Conversely, when the service rate is very large (i.e., $\mu > \mu_{\mathcal{W}}$), a higher service rate results in both a longer waiting time per admission and a longer total waiting time. In contrast to the existing queueing literature (see, e.g., Hassin and Haviv 2003, Anand et al. 2011), here an increase in service speed actually intensifies the system congestion. Interestingly, when the service rate is in a moderate range (i.e., $\mu^* \leq \mu \leq \mu_{\mathcal{W}}$), increasing service rate imposes opposite effects on waiting time per admission and total waiting time: it decreases the waiting time per admission but increases the total waiting time. This reveals that reducing the waiting time per admission may incentivize the HCP to increase their service speed, which, however, may increase the risk of readmission and thus increase the total waiting time. Therefore, an improvement in waiting time per admission may be at the cost of service quality and total waiting time.

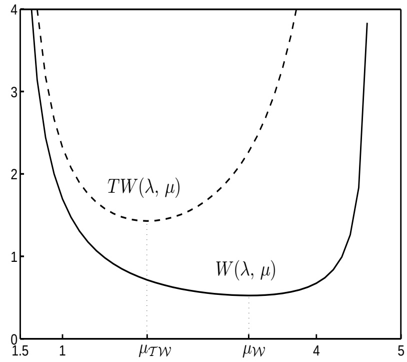


Figure 3.2: Waiting time

Moreover, from (3.6), given λ, the total waiting time is decreasing in the effective service rate $o(\mu)$. According to Lemma 11, it is natural that $TW(\lambda, \mu)$ achieves its minimum at μ^* which maximizes the effective service rate and makes the sensitivity of the cure probability equal one. It is worth noting that the left hand side of (3.7) is equal to the sensitivity of the cure probability times the probability that the HCP is busy. Because the service rate impacts service quality only when the HCP is busy, the left hand side of (3.7) represents the actual sensitivity of the cure probability. In other words, the waiting time per admission is minimized when the actual sensitivity of the cure probability equals to one.

Total Utilization Rate vs. Utilization Rate for New Patients

The relapse risk also enables us to evaluate the utilization rate of the HCP in two different ways: the utilization rate for new patients and the utilization rate for all the patients, which is referred as the total utilization rate. From (3.4), the total utilization rate $\rho_T(\lambda,\, \mu)$ and the utilization rate for new patients $\rho_N(\lambda,\, \mu)$ are respectively

$$\rho_T(\lambda,\, \mu) = \frac{\lambda_T}{\mu} = \frac{\lambda}{o(\mu)} \quad \text{and} \quad \rho_N(\lambda,\, \mu) = \frac{\lambda}{\mu}. \tag{3.8}$$

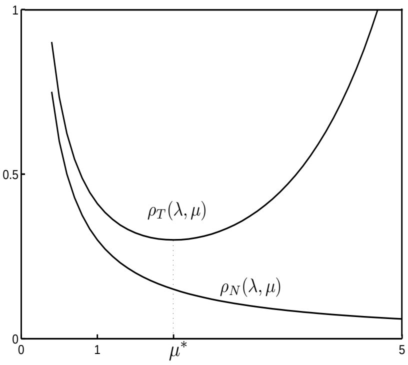


Figure 3.3: Utilization rate

According to (3.8), the total utilization rate in our model is equal to the utilization rate of the classic M/M/1 queue with an arrival rate of λ and a service rate of $o(\mu)$. To ensure the stability of the queue, we need the total utilization rate is smaller than 1; that is, $\lambda < o(\mu)$.

Proposition 7. *Given λ, the total utilization rate $\rho_T(\lambda,\, \mu)$ is quasi-convex in μ and achieves its minimum at μ^*, while the utilization rate for new patients $\rho_N(\lambda,\, \mu)$ is always decreasing in μ.*

By noting that the total utilization rate is decreasing in the effective service rate and utilizing Lemma 11, it is easy to see that the total utilization rate is quasi-convex in μ and achieves its minimum at μ^*. Combining the above facts with Proposition 6, we can conclude that both the patients and the HCP are happy with the service rate μ^*, as both the total waiting time and the total utilization rate are minimized. It is widely believed that reducing systems' congestion levels shall increase the HCP's idle time. Therefore, to improve the performance of the health care system, some balance between patients' waiting time and the HCP's idle time must be achieved (Fetter and Thompson 1996). However, our result shows that this argument is true only when the system congestion is measured in terms of the waiting time per admission. While in terms of the total waiting time, minimizing the system's congestion level and maximizing the HCP's idle time actually coincide. This cautions the social planner that a reduction in the idle time of the HCP may intensify the system congestion and lead to the inefficiency of the health care delivery. Furthermore, as depicted in Figure 3.3, Proposition 7 also tells us that when $\mu \geq \mu^*$, a larger service rate leads to a larger total utilization rate but a smaller utilization rate for new patients. Therefore, a reduction in the HCP's idle time may be mainly attributed to the increase in the number of readmitted patients and instead reduce the amount of time that the HCP spends on treating new patients.

So far, we have established three new performance measurements which are absent in the classic queueing theory. Based on these measurements, we next consider the benchmark case where the social planner can control both the patients and the HCP such that the first-best outcome is achievable.

3.4 First-Best Solution

In order to obtain the first-best outcome, we consider the centralized health care system where the social planner can control both the arrival rate of new patients and the service rate to maximize the social welfare. In the existing queueing literature with identical customers, the social welfare is equal to the profit of the service provider which does not incorporate the loss of balked customers (see, e.g., Naor 1969, Hassin and Haviv 2003). However, in the health care industry, the number of patients who leave without being seen is an important indicator of patient satisfaction (Polevoi et al. 2005). According to Hsia et al. (2011), "Patients who leave without being seen represent the failure of an emergency care delivery system to meet its goals of providing care to those most in need." If a patient leaves without being treated, he/she may get worse in future such that the HCP may need to spend more resources on treating him/her. Therefore, it is necessary to incorporate the social loss of balked patients into the objectives of the social planner. For ease of exposition, we assume that each balked patient induces a β, $\beta > 0$, penalty cost to the health care system. And the social planner determines the effective arrival rate of new patients and service rate to maximize the following social welfare SW:

$$\max_{\lambda,\,\mu} SW(\lambda,\,\mu) \;=\; \underbrace{\lambda[R - \theta TW(\lambda,\,\mu)]}_{(1)} - \underbrace{\beta(\Lambda - \lambda)}_{(2)}. \tag{3.9}$$

where the first term represents the total utilities obtained by the admitted cured patients and the second term represents the total social cost from balked patients. Because the HCP's profit is the internal transfer between the social planner and the HCP, it is canceled out in SW. It is worth noting that similar to Anand et al. (2011), here μ stands for the service speed instead of capacity (i.e., number of doctors), therefore there is no direct cost associated with the change of μ. Furthermore, in reality, people normally use waiting time per admission to measure the system congestion and calculate patients' utilities. If so, from (3.4) and 3.6, the patients' utilities are $\lambda_T[R - \theta W(\lambda,\,\mu)] = \lambda[Rn(\mu) - \theta TW(\lambda,\,\mu)]$. Compared with the first term in (3.9), it calculates the reward R multiple times. This cautions the social planner that based on the waiting time per admission, one may exaggerate the patients' utilities and incorrectly estimate the social welfare.

Let μ_b and λ_b denote the socially desired service rate and the effective arrival rate of new patients under the benchmark (centralized) scenario, respectively.

Proposition 8. *The objective function in* (3.9) *has the following properties.*

1. *Given λ, $SW(\lambda,\,\mu)$ is unimodal in μ and the corresponding mode is equal to μ^*, which is independent of λ. Therefore, $\mu_b = \mu^*$.*

2. *Given μ, $SW(\lambda,\,\mu)$ is concave in λ. Furthermore, the socially optimal effective arrival rate of new patients is $\lambda_b = o(\mu^*) - \sqrt{\theta o(\mu^*)/(R+\beta)}$.*

Observing from the objective function (3.9), given λ, maximizing social welfare is equivalent to minimize the total waiting time. Then according to Proposition 6, the service rate μ^* can simultaneously maximize social welfare and minimize total waiting time. Therefore, to evaluate the efficiency of service delivery of the HCP, the social planner just needs to monitor the total waiting time. Insightfully, by noting from Proposition 7 that the total utilization rate is also minimized at μ^*, we can conclude that given λ, the health care system is socially optimal when the HCP's idle time is maximized. In other words, reducing the HCP's idle time actually hurts the social welfare. In addition, recall from Proposition 6 that the performance of the health care system differs in terms of total waiting time and in terms of waiting time per admission. Now we can conclude that from the perspective of social welfare maximization, the total waiting time is a better performance measurement, as reducing waiting time per admission may increase the total waiting time and harm the social welfare.

3.5 Second-Best Solution

In this section, we study the decision problems for the three parties in the system: the patients, the HCP and the social planner. In reality, government in countries such as UK, Canada and Australia decides the price charged to the patient per admission (e.g., Reinhardt 2006, Klein 2012). We also note that in countries such as the United States and Netherland, the price is either unilaterally set by the government agencies (i.e., the CMS in the U.S. and list A service in Netherland) or negotiated between the HCP and the insurer (Reinhardt 2006, Oostenbrink and Rutten 2006). Naturally, when the bargaining power of the HCP is large enough, the negotiated charging price is similar to the one that the HCP can charge by itself. For example, in 1993, after the merge of two eminent Harvard-affiliated hospitals, the new hospital overwhelms the insurers and could "deny access to the beneficiaries of any insurer who dared not accept whatever they wanted to charge" (Roy 2011).

Therefore, depending on the price is determined by the HCP or by the social planner, we shall analyze two scenarios: the profit-maximizing price and the welfare-maximizing price. Under the profit-maximizing price, the HCP first decides both the price and the service rate. Then patients arrive at the HCP and the new patients decide to join or to balk. Under the welfare-maximizing price, the social planner decides the price first and then the HCP determines its service rate. Finally, the new patients make their joining decision. For the readmitted patients, we assume that they will join certainly.[5] To facilitate our analysis, we first investigate the new patients' joining decision and the HCP's service rate decision by assuming that the price is fixed. Then we study the pricing decisions under the two scenarios.

3.5.1 New Patients' Joining Decision

Classic literature assumes that customers can take waiting time and price into consideration. Here we assume that patients are more strategic than that such that they also know the readmission rate.[6] Therefore, they can calculate their expected utilities by taking the waiting time per admission, readmission rate and the price into account.[7] Then based on (3.3) and (3.5), given μ and p, the expected utility of an admitted cured patient can be written as

$$U = R - [n(\mu)p + \theta TW(\lambda, \mu)] = R - \frac{p}{1-\delta(\mu)} - \frac{\theta}{o(\mu)-\lambda}, \tag{3.10}$$

[5] Insurers normally do not allow the affiliated patients to admit to the hospital which is outside the network of the insurer's health care plan. And in the United States, the insurer intends to maintain a tighter and narrower network. (see http://www.nytimes.com/2013/09/23/health/lower-health-insurance-premiums-to-come-at-cost-of-fewer-choices.html?pagewanted=all&_r=0) Furthermore, admitting to a new hospital within the insurer's network may still cause some inconvenience to the readmitted patients, as they should reestablish a relationship with the doctors in the new hospital.

[6] In many countries such as the United States and Germany, patients can easily obtain those information on the website such as http://www.bkk-klinikfinder.de/.

[7] Stavrunova and Yerokhin (2011) show that health care demand is negatively affected by the waiting time per admission. And Varkevisser et al. (2012) show that a 1% reduction in readmission rate is associated with a 12% increase in hospital demand.

which is the difference between the service reward and the total cost (waiting cost+service fee paid) that the patient incurs. Compared with the customers' utilities under the queueing model without feedback, (3.10) has two differences. First, the waiting time is replaced by the total waiting time. Second, the price is replaced by the total expected payment $p/(1-\delta(\mu))$. These differences reflect the patients' perceptions on the service quality. (3.10) also shows that when the service reward or the maximum effective service rate is so small such that $R < \theta/o(\mu^*)$, the expected utility U is always negative. Then all the patients will leave without being treated. To avoid such uninteresting case, we only consider the scenario that $R \geq \theta/o(\mu^*)$ hereafter.

From (3.10), it is easy to see that the expected utility U is decreasing in λ, which implies that a patient has less incentive to join, as more patients choose to join. Therefore, the patients' best response is to avoid the crowd and there exists only one effective arrival rate of new patients. In equilibrium, the new patients are indifferent between joining and balking. Letting $U = 0$, we can derive the effective arrival rate of new patients as follows:

$$\lambda(\mu, p) = o(\mu) - \frac{\theta(1-\delta(\mu))}{R(1-\delta(\mu)) - p}. \tag{3.11}$$

From (3.11) we can easily see that given the service rate μ, $\lambda(\mu, p)$ is decreasing in price p; that is, the higher the payment charge, the less patients join the HCP. Using (3.4), (3.8) and (3.11), the total effective arrival rate, the total utilization rate and the utilization rate for new patients are respectively

$$\lambda_T(\mu, p) = \mu - \frac{\theta}{R(1-\delta(\mu)) - p}, \tag{3.12}$$

$$\rho_T(\lambda(\mu, p), \mu) = 1 - \frac{\theta}{Ro(\mu) - p\mu}, \tag{3.13}$$

$$\rho_N(\lambda(\mu, p), \mu) = 1 - \delta(\mu) - \frac{\theta(1-\delta(\mu))}{Ro(\mu) - p\mu}. \tag{3.14}$$

In the following, we proceed to investigate the impact of the service rate on $\lambda(\mu, p)$, $\lambda_T(\mu, p)$, $\rho_T(\lambda(\mu, p), \mu)$ and $\rho_N(\lambda(\mu, p), \mu)$. To this end, for a fixed p, we define $\mu_{an}(p)$, $\mu_{at}(p)$, $\mu_{ut}(p)$ and $\mu_{un}(p)$ as the service rates that maximize $\lambda(\mu, p)$, $\lambda_T(\mu, p)$, $\rho_T(\lambda(\mu, p), \mu)$ and $\rho_N(\lambda(\mu, p), \mu)$, respectively. It is worth noting that differing from Proposition 6 and 7, here the effective arrival rate of new patients are endogenous.

Proposition 9. *Given the price p, the following results hold.*

1. *the effective arrival rate of new patients $\lambda(\mu, p)$ is unimodal in μ.*

2. *the total effective arrival rate $\lambda_T(\mu, p)$ is concave in μ. Furthermore, $\mu_{at}(p)$ is the unique solution of*

$$\frac{\theta R\delta'(\mu_{at}(p))}{[R(1-\delta(\mu_{at}(p))) - p]^2} = 1. \tag{3.15}$$

3. *both the total utilization rate $\rho_T(\lambda(\mu, p), \mu)$ and the utilization rate for new patients $\rho_N(\lambda(\mu, p), \mu)$ are unimodal in μ.*

4. *the service rates that maximize arrival rate and utilization rate satisfy: $\mu_{un}(p) < \mu_{ut}(p) < \mu_{an}(p) < \min\{\mu^*, \mu_{at}(p)\}$.*

Proposition 9 shows that the relationship between the effective arrival rate of new patients and the HCP's service rate follows an inverted-U-shape curve. To see the underlying reason, let us examine the impact of the increase in service rate on the waiting cost and the total service fee paid by a patient (i.e., $n(\mu)p$). First, as the service rate increases, the number of HCP visits per patient increases such that the total service fee paid by a patient also increases. This reduces the joining incentives of new patients. Besides, recall from Proposition (6) that the total waiting time first decreases and then increases in the service rate. When the service rate is relatively low, the reduction in waiting cost overwhelms the increases in the total service fee such that more new patients choose to join. When the service rate is relatively high, the reverse is true and less new patients shall join. Obviously, the effective arrival rate of new patients achieves its maximum earlier than the total waiting time achieves its minimum; that is, $\mu_{an}(p) < \mu^*$.

Recall that the total effective arrival rate is equal to the effective arrival rate of new patients plus the arrival rate of the readmitted patients. Because the number of readmitted patients increases in the service rate, the total effective arrival rate is also unimodal in the service rate and achieves its maximum at a larger service rate than the effective arrival rate of new patients; that is, $\mu_{an}(p) < \mu_{at}(p)$. Interestingly, as shown in Figure 3.4(a), when $\mu \in [\mu_{an}(p), \mu_{at}(p)]$, increasing the total number of patients by increasing service speed may actually cause less new patients to receive treatment. This indicates that although a faster service speed services more patients, those patients are mainly readmitted ones due to the deterioration of service quality. Therefore, improving the accessibility for all the patients may reduce the accessibility for new patients.

Moreover, the utilization rate for new patients is also unimodal in μ. This can be explained as follows. When the service rate is low, most of the patients waiting in the system are new patients as the readmission rate is low. In this case, an increase in service rate enables more new patients to receive treatment. However, when the service rate is high, many patients in the queue are those readmitted patients as the

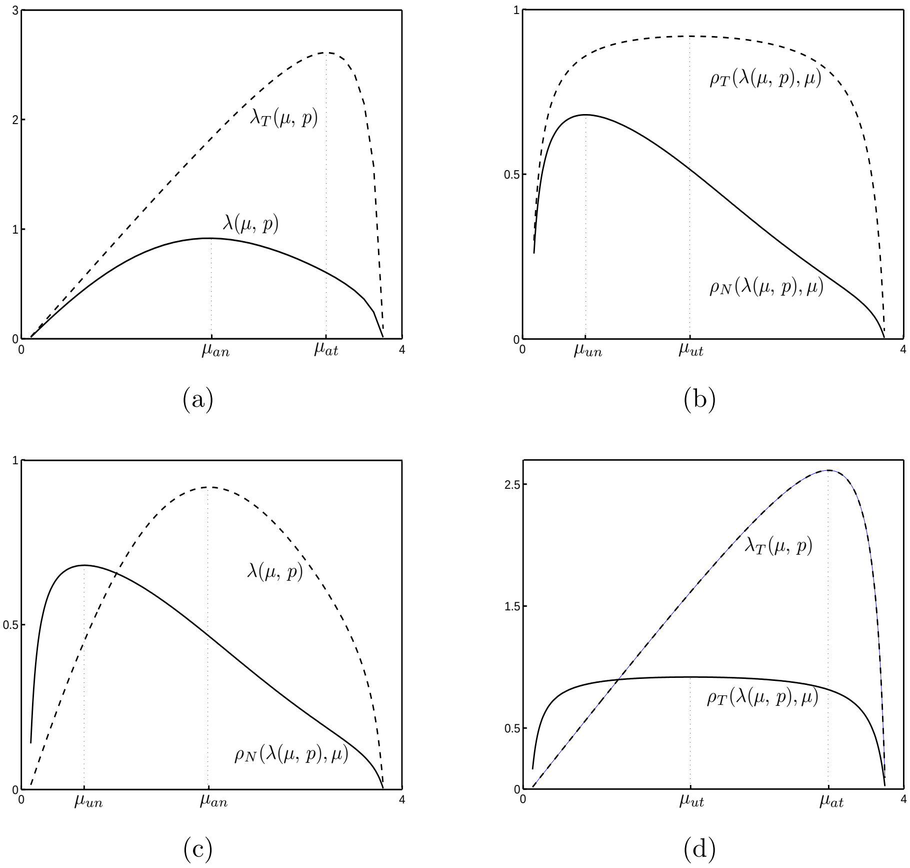


Figure 3.4: Arrival rate and Utilization rate

readmission rate is high. Hence, further increasing service rate reduces the joining incentives of new arrival patients, resulting in a smaller utilization rate for new patients. By noting that the readmission rate increases in the service rate, the total utilization rate is also unimodal in μ and achieves its maximum at a larger service rate compared with the utilization rate for new patients; that is, $\mu_{un}(p) < \mu_{ut}(p)$. Therefore, as shown in Figure 3.4(b), when $\mu \in [\mu_{un}(p), \mu_{ut}(p)]$, a higher service rate leads to a larger total utilization rate but a smaller utilization rate for new patients. Again, this result shows that reducing the HCP's idle time may reduce the amount of time the HCP spends on treating new patients.

Finally, as shown in Figure 3.4(c) (Figure 3.4(d)), because $\mu_{un}(p) < \mu_{ut}(p) < \mu_{an}(p) < \mu_{at}(p)$, we can know that when $\mu \in [\mu_{un}(p), \mu_{an}(p)]$ ($\mu \in [\mu_{ut}(p), \mu_{at}(p)]$), increasing the arrival rate of new patients (total arrival rate) actually reduces the utilization rate for new patients (total utilization rate). In contrast to the conventional wisdom, this conclusion implies that reducing the HCP's idle time may also reduce the accessibility of the health care services.

3.5.2 The HCP's Service Rate Decision

In this section, we are going to examine the HCP's best response by assuming that the price is fixed. The HCP makes the service rate decision μ to maximize its profit Π, which equals the total number of admissions (i.e., the total effective arrival rate $\lambda_T(\mu, p)$) times the payment per admission:

$$\max_{\mu} \Pi(\mu) \quad = \quad \lambda_T(\mu, p)p. \tag{3.16}$$

Observing from (3.16), because the price is fixed, the HCP's objective is actually to maximize the total effective arrival rate. Then according to Proposition 9, we can easily obtain the following result.

Proposition 10. *The HCP's profit* $\Pi(\mu)$ *is concave in the service rate* μ. *The* *corresponding optimal service rate $\mu_{\mathcal{H}}(p)$ is equal to $\mu_{at}(p)$ that maximizes the total effective arrival rate.*

Next, we conduct the sensitivity analyses of the system performance with respect to the price p, which are summarized in following corollary.

Corollary 12. *The sensitivity analysis with respect to the price reveals:*

1. *the optimal service rate $\mu_{\mathcal{H}}(p)$ is decreasing in p. $\mu_{\mathcal{H}}(p) = \mu^*$ if and only if (iff) $p = \bar{p}$, where*

$$\bar{p} = R(1 - \delta(\mu^*)) - \frac{\sqrt{R\theta o(\mu^*)}}{\mu^*}.$$

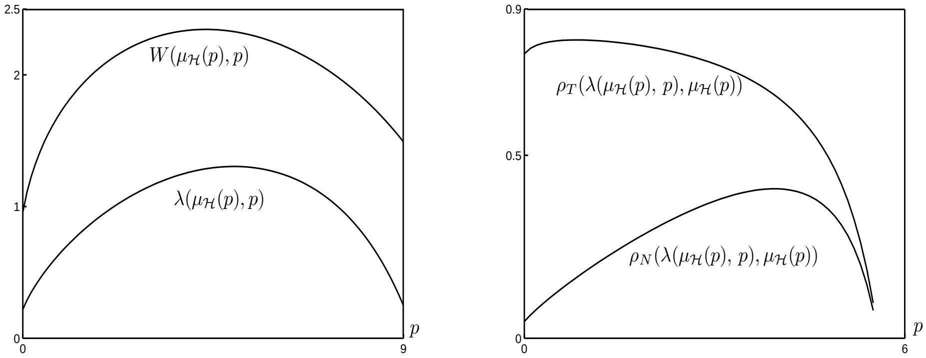
   Correspondingly, $\lambda(\mu^*, \bar{p}) = o(\mu^*) - \sqrt{\theta o(\mu^*)/R}$.

2. *the total effective arrive rate $\lambda_T(\mu_{\mathcal{H}}(p), p)$, the readmission rate $\delta(\mu_{\mathcal{H}}(p))$ and the total waiting time $TW(\lambda(\mu_{\mathcal{H}}(p), p), \mu_{\mathcal{H}}(p))$ are decreasing in p.*

3. *the effective arrival rate of new patients $\lambda(\mu_{\mathcal{H}}(p), p)$ is decreasing in p when $p \geq \bar{p}$.*

The first statement of Corollary 12 shows that when the service price increases such that the revenue generated from each patient is higher, the HCP will spend more time on treating each patient. Therefore, to improve health care quality in terms of readmission rate, the social planner just needs to increase the price.

Besides, the second statement of Corollary 12 implies that an increase in service charge p leads to a reduction in the total effective arrival rate, the readmission rate and the expected total waiting time. This is because as the price increases, the HCP has less incentives to increase its service rate, leading to a reduction in the number of readmitted patients. However, although a higher price can mitigate the system congestion in terms of the total waiting time, it may not hold in terms of the waiting time per admission. See the illustration of Figure 3.5(a) for an example, where both



(a) (b)

Figure 3.5: The impact of price on waiting time and utilization rate

the waiting time per admission and the effective arrival rate of new patients are concave in the service charge p. Interestingly, they both increase in p when p is not too high. This implies that when price is relatively low, an increase in price actually causes patients to wait longer to see a doctor and lead more new patients to join the queue, which are in contrast to the result obtained under the traditional queueing setting (i.e., Hassin and Haviv 2003). Our results can shed important light on the performance measurement of the health care system. It is widely recognized that a higher price can erode the joining incentives of customers and mitigate the system congestion. However, our results show that this is true only when the performance of the health care system is measured by the total arrival rate and the total waiting time. Regarding the waiting time per admission and the arrival rate of new patients, the reverse is actually true when the price is relatively low. Additionally, Figure 3.5(b) shows that a higher price may also have opposite impacts on the utilization rate for new patients and the total utilization rate.

On a separate note, in some circumstances the price of the health care service is determined through the negotiation between the HCP and the insurer. Intuitively, a powerful insurer can negotiate a lower service charge. Corollary 12 then implies that paying low service charge by joining a powerful insurance company may actually hurt the well-being of patients, as the low price incentivizes the HCP to undertreat the patients, resulting in a high readmission rate and poor care quality. For example, in 1998, as a result of Aetna's acquisition of Prudential's health insurance assets, Aetna was believed to become a monopoly insurance company in Dallas and Houston and unduly negotiated down doctors' bill payment. Complaints were received that physicians spent less time with each Aetna patient, resulting in a reduction in care quality (Schwartz 1999). In 1999, the US Justice Department forced the company to sell assets in both cities (Wilke 2004).

3.5.3 Price Decision

Based on the best response function of the HCP established in §3.5.2, we now derive the profit-maximizing price and the welfare-maximizing price, respectively.

Profit-Maximizing Price

If the price is determined by the HCP, it will choose the price to maximize its expected profit $\Pi(\mu_{\mathcal{H}}(p))$. According to Corollary 12, we can obtain the following result.

Proposition 11. *Both the effective service rate $o(\mu_{\mathcal{H}}(p))$ and the HCP's profit $\Pi(\mu_{\mathcal{H}}(p))$ are quasi-concave in p and achieve the maximum at $p = \bar{p}$. Therefore, if the price is determined by the HCP, it will choose $\bar{p}$ and μ^*. Furthermore, $\lambda(\mu^*, \bar{p}) < \lambda_b$.*

Proposition 11 shows that if the price is determined by the HCP, it shall serve in the socially desirable way but less patients can receive treatment compared with the social optimality (the first-best outcome). Recall that $\lambda(\mu, p)$ is decreasing in p, we can conclude that the inefficiency of the health care system under the profit-maximizing price originates from the high price charged by the HCP instead of the suboptimal service rate.

Proposition 11 also sheds light on the effect of merge among HCPs on the health care system performance. According to Creswell and Abelson (2013), for the sake of increasing collective bargaining power, HCPs are merging faster and in greater numbers than they had before. The number of merge transactions between HCPs jumped from 50 in 2009 to 105 in 2012. Concerns about the monopoly issue have been raised over the consolidation among HCPs, which were accused of driving up price and reducing service quality (e.g., Roy 2011). However, Proposition 11 reveals that when the HCP has the pricing power, the performance of the health care system in terms of effective service rate and total waiting time are actually improved. Therefore, as the health care market continues to concentrate, patients may suffer from the price increase but benefit from a more efficient treatment (i.e., a higher effective service rate and a smaller total waiting time).

Welfare-Maximizing Price

We now consider the scenario where the price is determined by the social planner. The social planner shall choose a price to maximize the social welfare. Note that patients in equilibrium are indifferent between joining and balking. Their expected utilities are zero. The corresponding social welfare SW then equals the profit of the HCP minus the penalty cost from the balked patients. Given the HCP's best response $\mu_{\mathcal{H}}(p)$, the social planner's optimization problem is

$$\max_{p} SW(p) \;=\; \Pi(\mu_{\mathcal{H}}(p)) - \beta\,(\Lambda - \lambda(\mu_{\mathcal{H}}(p),\, p))\,.$$

We denote $p_{\mathcal{S}}$ as the optimal solution of the above optimization problem and let $\mu_{\mathcal{S}} = \mu_{\mathcal{H}}(p_{\mathcal{S}})$.

Unfortunately, $SW(p)$ is normally not unimodal in p, thus the optimal price is not unique. Nevertheless, by investigating the first order condition of $SW(p)$, we can derive the following result.

Proposition 12. *In equilibrium,* $p_{\mathcal{S}} < \bar{p}$ *and* $\mu_{\mathcal{S}} > \mu^*$. *Furthermore,* $\lambda(\mu^*, \bar{p}) < \lambda(\mu_{\mathcal{S}}, p_{\mathcal{S}}) < \lambda_b$.

Proposition 12 shows that the regulation via pricing cannot make the health care system achieve the social optimality. Actually, if the price is determined by the social planner, both the service rate and the accessibility for new patients are suboptimal. In particular, the HCP will spend less time on each patient compared with that under the socially optimal service rate. Interestingly, although the service speed increases, less new patients can receive treatment as the readmission rate is too high. Furthermore, recall that the service rate is socially optimal under the profit-maximizing price. We can conclude that the welfare-maximizing price reduces the efficiency of the HCP's care delivery. Finally, compared to the scenario of profit-maximizing price, here the welfare-maximizing price is smaller while the effective arrival rate of new patients is larger. In classic queueing models with identical customers such as Hassin and Haviv (2003), a customer does not consider the resulting negative externality on other customers when deciding whether to join. Therefore, in equilibrium, the service provider can extract all of the consumer surplus. In consequence, the profit-maximizing price and the welfare-maximizing price coincide. However, in the health care industry, the social planner also takes the social cost from the balked patients into account. Consequently, the social planner charges a lower price, leading to a larger service speed and a larger accessibility of the health care services.

So far, we have shown that the one-dimensional control over price cannot induce the health care system to be socially optimal. Another question of interest is that whether the multi-dimensional control can solve this puzzle. We discuss it in the next section.

3.5.4 Discussion: Two-Dimensional Control over Price and Readmission Rate

In reality, the social planner may simultaneously control price and readmission rate. For example, the Centers for Medicare and Medicaid Services (CMS), a federal agency within the U.S. Department of Health and Human Services, launched a *Hospital Readmission Reduction Program* in October 2012. Under this program, a HCP will be penalized a certain percentage of its annual reimbursements if its readmission rate exceeds a pre-determined threshold level. The penalty rate was set to be around 1% in 2013 and will grow to 3% in 2015 (Rau 2012). Consequently, after the implementation of this readmission rate based payment scheme (RBP) by the CMS, in 2013, 2,225 hospitals, among which two-thirds are the national hospitals, are found to be eligible for penalization, leading to a total of $227 million in penalty (Rau 2013).

We next examine whether the two-dimensional control over price and readmission rate can make the health care system to be socially optimal. Let $\hat{\delta}$ represent the threshold readmission rate. If its readmission rate is larger than $\hat{\delta}$, the HCP gets penalized a α ($0 \leq \alpha \leq 1$) portion of its total profit. As the readmission rate $\delta(\mu)$ is strictly increasing in service rate μ, the threshold readmission rate $\hat{\delta}$ corresponds to a threshold service rate $\hat{\mu}$, which is the unique root of $\delta(\mu) = \hat{\delta}$. Then, $\delta(\mu) > \hat{\delta}$ iff $\mu > \hat{\mu}$. Let

$$\mathbf{1}_{\mu > \hat{\mu}} = \begin{cases} 1, & \text{if } \mu > \hat{\mu}, \\ 0, & \text{if } \mu \leq \hat{\mu}. \end{cases}$$

Here, the social planner needs to decide (1) the price p, (2) the penalty rate α, $0 \leq \alpha \leq 1$, and (3) the threshold readmission rate $\hat{\delta}$ (equivalent to a threshold service rate $\hat{\mu}$). And the HCP decides a service rate to maximize its expected profit:

$$\max_{\mu} \Pi_{\mathcal{R}}(\mu) \quad = \quad (1 - \alpha \mathbf{1}_{\mu > \hat{\mu}}) \lambda_T(\mu, p) p.$$

Then, given the HCP's best response $\mu_{\mathcal{R}}(p,\, \alpha,\, \hat{\mu})$, the social planner maximizes the social welfare SW as follows:

$$\begin{aligned}\max_{p,\,\alpha,\,\hat{\mu}} SW &= \lambda(\mu_{\mathcal{R}}(p,\, \alpha,\, \hat{\mu}),\, p)\,[R - \theta TW(\lambda(\mu_{\mathcal{R}}(p,\, \alpha,\, \hat{\mu}),\, p),\, \mu_{\mathcal{R}}(p,\, \alpha,\, \hat{\mu}))] \\ &\quad -\beta(\Lambda - \lambda(\mu_{\mathcal{R}}(p,\, \alpha,\, \hat{\mu}),\, p)).\end{aligned}$$

Recall from (3.11) that given μ, $\lambda(\mu,\, p)$ is decreasing in p. Therefore, letting the service rate $\mu = \mu_b = \mu^*$ (the benchmark socially desired service rate), correspondingly, there exists a unique p^* solving $\lambda(\mu^*,\, p) = \lambda_b$ (the benchmark socially desired effective arrival rate of new patients), which can be explicitly expressed as

$$p^* = R(1 - \delta(\mu^*)) - \frac{\sqrt{\theta o(\mu^*)(R + \beta)}}{\mu^*}. \tag{3.17}$$

According to Proposition 8, p^* is the price that leads to the socially desired outcome, namely, the *socially optimal price.*

Let

$$\bar{\lambda}_T = n(\mu_{\mathcal{H}}(p^*))\lambda(\mu_{\mathcal{H}}(p^*),\, p^*)$$

be the total effective arrival rate when the socially optimal price p^* is charged. Then we have the following proposition.

Proposition 13. *Under the two-dimensional control over price and readmission rate, in equilibrium, the social planner chooses the socially optimal price* p^**, the penalty percentage* $\alpha > 1 - \lambda_T(\mu^*,\, p^*)/\bar{\lambda}_T$*, and a threshold readmission rate* $\hat{\mu} = \mu^*$*, and the HCP chooses the socially optimal service rate* μ^**. Therefore, the health care system is socially optimal.*

Proposition 13 shows that the two dimensional control over price and readmission rate can force the health care system to achieve social optimality. This result is intuitive as the social planner here can use two means, the threshold admission rate/sevice rate and the penalty charge to align the incentives of the HCP with that of the social welfare.

3.6 The Analysis of the Scenario When the Potential Arrival Rate is Small

In this section, we study the scenario under which the potential patient size is relatively small such that it is possible that no patients will leave without being treated. We first derive the first best outcomes.

Proposition 14. *When the potential market size is relatively small such that it is possible that no patients will leave without being treated, then* $\mu_b = \mu^*$ *and* $\lambda_b = \min\{o(\mu^*) - \sqrt{\theta o(\mu^*)/(R+\beta)},\, \Lambda\}$.

According to Proposition 8, the service rate that maximizes $SW(\lambda, \mu)$ is independent of the arrival rate of new patients. Therefore, it is natural that the socially optimal service rate remains the same. However, differing from the scenario under which the potential arrival rate is large, here, when the health care system is less congested (i.e., $\Lambda \leq o(\mu^*) - \sqrt{\theta o(\mu^*)/(R+\beta)}$), all the patients can receive treatment; otherwise, only a proportion of patients are treated while the remaining patients balk the system.

3.6.1 The Second-Best Solution

We now study the decision problems for the patients, the HCP and the social planner, respectively. As before, depending on whether the price is determined by the HCP or by the social planner, we shall analyze two scenarios: the profit-maximizing price and the welfare-maximizing price.

3.6.2 The Patients' Joining Decision

We begin with the patients' joining decision. To avoid confusion, we use $\lambda_{\mathcal{L}}$ to represent the effective arrival rate of new patients hereafter. Furthermore, we name the scenario where all patients join the queue; that is $\lambda_{\mathcal{L}} = \Lambda$, the *full market coverage* scenario, and the scenario where some patients will balk; that is $\lambda_{\mathcal{L}} < \Lambda$, the *partial market coverage* scenario. Note that the partial market coverage scenario has been analyzed in section 3.5. To be consistent, we still use λ to represent the effective arrival rate of new patients under the partial market coverage scenario. According to (3.4) and (3.10), we have

$$\lambda_{\mathcal{L}}(\mu, p) = \min\{\Lambda, \lambda(\mu, p)\} \text{ and } \lambda_T(\mu, p) = n(\mu) \min\{\Lambda, \lambda(\mu, p)\}, \tag{3.18}$$

where $\lambda(\mu, p)$ represents the effective arrival rate of new patients under the partial market coverage scenario and is given by (3.11).

3.6.3 Profit-Maximizing Price

Under the profit-maximizing price, the HCP decides both the service rate and price to

$$\max_{\mu, p} \Pi(\mu, p) \quad = \quad \lambda_T(\mu, p)p = n(\mu) \min\{\lambda(\mu, p), \Lambda\}p.$$

We use the subscript $\mathcal{P}$ to represent the equilibrium outcome associated with the scenario of profit-maximizing price.

Proposition 15. *Under the profit-maximizing price, the following results hold.*

1. *When* $\Lambda > \lambda(\mu^*, \bar{p})$, $\mu_{\mathcal{P}} = \mu^*$, $p_{\mathcal{P}} = \bar{p}$ *and* $\lambda(\mu_{\mathcal{P}}, p_{\mathcal{P}}) < \lambda_b$.

2. *When* $\Lambda \leq \lambda(\mu^*, \bar{p})$, $\mu_{\mathcal{P}} = \mu^*$, $p_{\mathcal{P}} = R(1 - \delta(\mu^*)) - \theta(1 - \delta(\mu^*))/(o(\mu^*) - \Lambda)$ *and* $\lambda(\mu_{\mathcal{P}}, p_{\mathcal{P}}) = \Lambda$.

3. The health care system can achieve the benchmark social optimality iff $\Lambda \leq \lambda(\mu^*, \bar{p})$.

Proposition 15 shows that when the size of the potential patients is relatively small (i.e., $\Lambda \leq \lambda(\mu^*, \bar{p})$), the full market coverage scenario arises and the health care system can achieve socially optimality; otherwise, the partial market coverage scenario is preferred by the HCP and the health care system is suboptimal. Furthermore, the HCP under the scenario of the profit-maximizing price shall always choose the socially optimal service rate.

3.6.4 Welfare-Maximizing Price

Under the welfare-maximizing price, the social planner first decides the price and then the HCP decides the service rate. With the constraint of the potential arrival rate Λ, the HCP's optimization problem is

$$\max_{\mu} \Pi(\mu) \;\; = \;\; \lambda_T(\mu,\, p)p = n(\mu)\min\{\lambda(\mu,\, p),\, \Lambda\}p.$$

Given the HCP's best response $\mu(p)$, the social planner makes the service pricing decision p to maximize the social welfare SW:

$$\max_{p} SW(p) \;\; = \;\; \lambda(\mu(p),\, p)\left[R - \theta TW(\lambda(\mu(p),\, p),\, \mu(p))\right] - \beta\left(\Lambda - \lambda(\mu(p),\, p)\right).$$

We will solve the above game sequence by backward induction. To do so, we first derive the optimal pricing and service rate decisions under both the partial market coverage scenario and the full market coverage scenario. We then derive the equilibrium outcome by comparing the system performance under the two scenarios. Note that we have solved the game under the partial market coverage scenario in section 3.5. In the following, we just need to consider the scenario of the full market coverage. To be consistent, we still use $\mu_{\mathcal{H}}(p)$ and $p_{\mathcal{S}}$ to represent the HCP's best response service rate and the social planner's optimal price under the partial market coverage scenario. And we use $\mu_{\mathcal{F}}(p)$ and $p_{\mathcal{SF}}$ to represent the HCP's best response service rate and the social planner's optimal price under the full market coverage scenario.

3.6.5 Full Market Coverage

Now we consider the scenario under which the potential patient size is relatively small such that the results under the partial market coverage scenario are infeasible; that is, $\lambda(\mu_{\mathcal{H}}(p), p) \geq \Lambda$. In this case, in equilibrium, all patients can receive treatment and the full market coverage scenario arises. Thus, the HCP's optimization problem becomes

$$\begin{aligned} \max_{\mu} \Pi(\mu) &= n(\mu)\Lambda p = \frac{\Lambda p}{1-\delta(\mu)}, \\ s.t. \quad & \lambda(\mu, p) \geq \Lambda. \end{aligned} \tag{3.19}$$

Since the objective function $\Pi(\mu)$ in (3.19) is increasing in μ as $\delta(\mu)$ increases in μ, the HCP under the full market coverage scenario shall set a largest service rate that satisfies the constraint. Because $\lambda(\mu, p)$ is quasi-concave in μ (see Lemma 9), for a given price p, there exist at most two points, denoted by μ_1 and μ_2, such that the constraint is binding (i.e., $\lambda(\mu, p) = \Lambda$). And $\lambda(\mu, p) \geq \Lambda$ iff μ is located between them. Therefore, the optimal service rate under the full market coverage scenario is the larger root of $\lambda(\mu, p) = \Lambda$. In other words, the HCP chooses the largest service rate that satisfies the full coverage requirement. We summarize it in the following proposition.

Proposition 16. *Under the full market coverage scenario, the HCP's optimal service rate $\mu_{\mathcal{F}}(p)$ is the larger root of $\lambda(\mu, p) = \Lambda$.*

Next, we conduct the sensitivity analyses of the system performance with respect to the price p, which are summarized in the following corollary.

Corollary 13. *Under the full market coverage scenario,*

1. *both the optimal service rate* $\mu_{\mathcal{F}}(p)$ *and the corresponding total effective arrival rate* $\lambda_T(\mu_{\mathcal{F}}(p), p)$ *are decreasing in* p*.*

2. *both the waiting time per admission* $W(\Lambda, \mu_{\mathcal{F}}(p))$ *and the total waiting time* $TW(\Lambda, \mu_{\mathcal{F}}(p))$ *are quasi-convex in* p*. Let* $p_{\mathcal{W}}$ *and* $p_{\mathcal{TW}}$ *be the prices that minimize* $W(\Lambda, \mu_{\mathcal{F}}(p))$ *and* $TW(\Lambda, \mu_{\mathcal{F}}(p))$*, respectively. Then* $p_{\mathcal{TW}} > p_{\mathcal{W}}$*. Moreover,*

   - *when* $p \leq p_{\mathcal{W}}$*, both* $W(\Lambda, \mu_{\mathcal{F}}(p))$ *and* $TW(\Lambda, \mu_{\mathcal{F}}(p))$ *are decreasing in* p*;*
   - *when* $p_{\mathcal{W}} < p < p_{\mathcal{TW}}$*,* $W(\Lambda, \mu_{\mathcal{F}}(p))$ *is increasing in* p *while* $TW(\Lambda, \mu_{\mathcal{F}}(p))$ *is decreasing in* p*;*
   - *when* $p \geq p_{\mathcal{TW}}$*, both* $W(\Lambda, \mu_{\mathcal{F}}(p))$ *and* $TW(\Lambda, \mu_{\mathcal{F}}(p))$ *are increasing in* p*.*

Similar to the partial market coverage scenario, the first statement of Corollary 13 shows that under the full market coverage scenario a higher service charge p leads to a lower service rate and a smaller total effective arrival rate. The reason is that when the market is fully covered, the only way for the HCP to increase profit is to speed up the treatment of patients so as to generate more readmitted patients. And as the price increases, the HCP has stronger incentives to increase service speed. By noting that $\mu_{\mathcal{F}}(p)$ is decreasing in p, the quasi-convexity of waiting time per admission and total waiting time with respect to the price p is a natural consequence of Proposition 6. Moreover, the second statement of Corollary 13 also implies that when p is relatively small, increasing price can mitigate the system congestion in terms of both the waiting time per admission and the total waiting time. However, when p is in a moderate range, a higher price leads to a longer waiting time per admission but a shorter total waiting time. Interestingly, when the price is relatively high, a higher price leads to both a longer total waiting time and a longer waiting time per admission. In contrast to the result under the traditional queueing setting, here, increasing price will intensify the system congestion.

Next, we proceed to derive the social planner's pricing decision under the full market coverage scenario. Recall that the full market coverage scenario will arise only when the results under the partial market coverage are infeasible; that is, $p \in \Theta$, where $\Theta = \{p | \lambda(\mu_{\mathcal{H}}(p), p) \geq \Lambda\}$. Then, given the HCP's best response $\mu_{\mathcal{F}}(p)$, the social planner makes its pricing decision p to

$$\max_{p \in \Theta} SW(p) = \Lambda \left[R - \theta TW(\Lambda, \mu_{\mathcal{F}}(p))\right] = \Lambda \left[R - \frac{\theta}{o(\mu_{\mathcal{F}}(p)) - \Lambda}\right]. \quad (3.20)$$

Since the social welfare SW as stated in (3.20) is increasing in the effective service rate (which achieves the maximum at the service rate $\mu = \mu^*$), the social planner shall choose a price that leads to the HCP's best response service rate as close to μ^* as possible. For notational convenience, we let $\mu_{\mathcal{SF}} = \mu_{\mathcal{F}}(p_{\mathcal{SF}})$.

The following proposition summarizes the equilibrium outcome under the full market coverage scenario.

Proposition 17. *Under the full market coverage scenario, the following results hold.*

1. *When* $\Lambda \leq \lambda(\mu^*, \bar{p})$, $p_{\mathcal{SF}} = R(1 - \delta(\mu^*)) - \theta(1 - \delta(\mu^*))/(o(\mu^*) - \Lambda) \geq \bar{p}$ *and* $\mu_{\mathcal{SF}} = \mu^*$.

2. *When* $\Lambda > \lambda(\mu^*, \bar{p})$, $p_{\mathcal{SF}} = \max\{p | \lambda(\mu_{\mathcal{F}}(p), p) \geq \Lambda\} < \bar{p}$ *and* $\mu_{\mathcal{SF}} > \mu^*$.

Proposition 17 shows that if the market is fully covered, that is, all patients are admitted into the health care system, the HCP actually serves in the socially desired way when the potential patient size is relatively small. However, when the size of the potential patients is large, the HCP under the full market coverage scenario will set a service rate larger than the socially desirable one.

3.6.6 Equilibrium Outcome

So far we have derived the optimal equilibrium outcomes associated with the partial market coverage scenario and the full market coverage scenario, respectively. Below we further investigate under which conditions which scenario appears as the equilibrium outcome. Then we can evaluate the performance of the health care system by comparing with the social optimality.

By noting that the equilibrium outcome under the partial market coverage scenario is not unique in general, we denote

$$\lambda_m = \min\{\lambda(\mu_{\mathcal{S}}, p_{\mathcal{S}})\}$$

as the smallest equilibrium effective arrival rate of new patients under the partial market coverage scenario. Then, according to Proposition 12, $\lambda_m > \lambda(\mu^*, \bar{p})$.

Proposition 18. *The equilibrium outcome is the full market coverage scenario if* $\Lambda \leq \lambda_m$ *and is the partial market coverage scenario if* $\Lambda > \lambda_m$.

Proposition 18 specifies the condition under which the partial/full market coverage scenario will occur. Now we are ready to examine the performance of the health care system.

Proposition 19. *When* $\Lambda \leq \lambda(\mu^*, \bar{p})$*, the health care system achieves the social optimality and in this case, the health care system ends up with the full market coverage. Otherwise, the equilibrium service rate is larger than the socially optimal one.*

Proposition 19 shows that if the potential patient size is very small ($\Lambda \leq \lambda(\mu^*, \bar{p})$), it is both in the HCP's best interest and in the social planner's objective to serve all patients, and the health care system can reach social optimality. However, if the potential patient size is relatively large, there exists a conflict between admitting more new patients and reducing the readmitted patients. Therefore, the performance of the health care system can never achieve the social optimality.

3.7 Conclusion

In this chapter, we adopt a queueing model to investigate how to evaluate a health care system's performance. By considering the tradeoff between service speed and service quality in terms of readmission rate, we first define three new system measurements. The first one is the effective service rate, which measures the efficiency of the HCP's service delivery. The second one is the total waiting time, which measures the congestion level of the system. The last one is the utilization rate for new patients, which measures how much time the HCP spends on treating new patients.

Those measurements enable us to obtain different results from those of classic queueing models without feedback. Firstly, by studying the joining decision of the patients, we have the following four important findings. First, increasing service rate may decrease the waiting time per admission but increase the total waiting time. Therefore, a reduction in waiting time per admission may be at the expense of service quality and total waiting time. Second, increasing total utilization rate may actually reduce the utilization rate for new patients. This implies that reducing the HCP's idle time may cause the HCP to spend less time on treating new patients. Third, a larger total arrival rate may lead to a smaller arrival rate of new patients. Therefore, improving the accessibility for all the patients may reduce the accessibility for new patients. Fourth, increasing the arrival rate of new patients (total arrival rate) may reduce the utilization rate for new patients (total utilization rate). This implies that reducing HCP's idle time may reduce the accessibility of the health care services.

Furthermore, by examining the HCP's service rate decision, we show that increasing price may mitigate the system congestion in terms of total waiting time but intensify system congestion in terms of waiting time per admission. Based on the HCP's best response, we derive the optimal price from the perspectives of the HCP and the social planner, respectively. We show that under the profit-maximizing price, the service rate is socially optimal, while under the welfare-maximizing price, the service rate is larger than the socially optimal one. Therefore, regulation via pricing reduces the efficiency of the HCP's care delivery, and the one-dimensional control over price fails to induce the health care system to achieve the social optimality. Finally, we show that a two-dimensional control over price and readmission rate can make the health care system socially optimal.

Our analysis is an initial attempt to examine the efficiency of the health care system by taking into account the relationship between service quality in terms of readmission rate and service speed. Moving forward, there are several directions for future research. One important area is to consider the competition among different HCPs. It is of interest to investigate the impact of competition on the HCPs' choices of service rate. However, this issue involves the patient transfer policy among different HCPs, whose analysis may be significantly different from ours. Another possible extension is to add the mortality rate into our analysis. An implicit assumption in our model is that the patients can definitely be cured in long run. This assumption is realistic for the non-fatal illness such as hip and knee arthrosis. However, for the fatal illness such as diabetes mellitus, the increase in mortality rate is an inevitable consequence of low quality. Nevertheless, as long as the tradeoff between service speed and readmission rate exists, our main results can still hold.

Chapter 4

Summary and Future Research

In this dissertation, we studied several issues related to social responsible operations in agricultural and health care industries. For the first essay, we examined whether farmers should utilize market information to make better production planning decisions and whether farmers should adopt agricultural advice to improve their operations when both market demand and process yield are uncertain. We show that the provision of market information always improves the farmers' total welfare and that farmers should use market information to improve their production planning. However, whether a farmer should adopt the agricultural advice is dependent on the size of the requisite upfront investment. More importantly, agricultural advice is welfare improving if and only if the upfront investment is sufficiently low. There are several possible extensions for this topic. First, for simplicity, we have assumed that farmers are risk neutral. However, in practice, the small farmers may be risk averse. Therefore, it is of interest to further examine how farmers' risk attitude leads to different results. Seconde, we also assumed that both farmers produce the same product. One may consider a situation where both farmers have options to decide which crop to produce. Such a setting enables us to investigate whether the provision of market information and agricultural advice can coordinate farmers' crop selections.

For the second topic, we focused on the performance indicators in health care industry. By distinguishing between the new patients and the readmitted patients, we introduce three new performance measurements: the effective service rate, the total waiting time and the utilization rate for new patients. We show that contingent on different performance measurements, one may draw opposite conclusions about a health care system's performance. To our best knowledge, our work is the first one that study the performance measurements in health care industry by considering the tradeoff between service speed and service quality in terms of readmission rate. Moving forward, there are a lot of directions for future research. For example, in practice, the HCP's payment is sometimes dependent on how much time it spends on treating the patients. For instance, in Japan, the payment for health-care service is made on a per-day basis (Ikegami and Anderson 2012). In Maryland, the inpatient care is determined by unit of time such as per-minute rate for operating room (Conis 2009). It is of interest to investigate which performance measurement is better to evaluate a health care system's performance under the time-based payment scheme. Another possible direction is to consider multi-step treatments and incorporate overtreatment into our analysis. In this thesis, we only analyzed the situation in which patients receive one-step treatment. However, in practice, the treatment for the diseases such as heart disease normally includes multiple steps. It is widely believed that the HCP has incentives to overtreat the patients that need multi-step treatments by providing unnecessary medical services such as needless tests or scans, resulting in huge waste (Berwick and Hackbarth 2012). To prevent overtreatment, it is important to figure out how to measure a health care system's performance when patients need multi-step treatments.

Appendix A

Proofs of Chapter 2

Proof of Proposition 1. As the best response functions of both farmers have the same structure, the equilibrium is symmetric, that is, $q_1^{NN} = q_2^{NN}$. Therefore,

$$q_i^{NN} = \frac{g - \mu_y^2 q_i^{NN}}{2S^2},$$

which yields

$$q_i^{NN} = \frac{g}{2S^2 + \mu_y^2}.$$

Consequently,

$$\pi_i^{NN} = q_i^{NN}\left[g - (S^2 + \mu_y^2)q_i^{NN}\right] = \frac{S^2 g^2}{(2S^2 + \mu_y^2)^2}.$$

Recall that $S^2 = \sigma_y^2 + \mu_y^2$ and $g = \mu_m \mu_y - c$, it is easy to see that π_i^{NN} and q_i^{NN} are increasing in μ_m and decreasing in σ_y. Because $\frac{dS^2}{d\mu_y} = 2\mu_y$ and $\frac{dg}{d\mu_y} = \mu_m$, we have

$$\begin{aligned}
\frac{d\pi_i^{NN}}{d\mu_y} &= \frac{(\frac{dS^2}{d\mu_y}g^2 + S^2\frac{dg^2}{d\mu_y})(2S^2 + \mu_y^2) - 2S^2 g^2(2\frac{dS^2}{d\mu_y} + 2\mu_y)}{(2S^2 + \mu_y^2)^3} \\
&= \frac{2\sigma_y^2 g(\mu_m \sigma_y^2 + \mu_y c) + 2S^2 g\mu_m \sigma_y^2 + 4S^2 g\mu_y c + 2\mu_y S^2 gc}{(2S^2 + \mu_y^2)^3} \\
&> 0.
\end{aligned}$$

$$
\begin{aligned}
\frac{d(\mu_y q_i^{NN})}{d\mu_y} &= q_i^{NN} + \mu_y \frac{dq_i^{NN}}{d\mu_y} \\
&= \frac{g}{2S^2+\mu_y^2} + \frac{\mu_y \left(\frac{dg}{d\mu_y}(2S^2+\mu_y^2) - g(\frac{dS^2}{d\mu_y} + 2\mu_y) \right)}{(2S^2+\mu_y^2)^2} \\
&= \frac{g}{2S^2+\mu_y^2} + \frac{\mu_y(\mu_m(2S^2+\mu_y^2) - 4g\mu_y)}{(2S^2+\mu_y^2)^2} \\
&= \frac{g(2S^2+\mu_y^2) + \mu_y\mu_m(2S^2+\mu_y^2) - 4g\mu_y^2}{(2S^2+\mu_y^2)^2} \\
&= \frac{2g\sigma_y^2 + 2\mu_y\mu_m S^2 + \mu_m\mu_y^3 - g\mu_y^2}{(2S^2+\mu_y^2)^2} \\
&= \frac{2g\sigma_y^2 + 2\mu_y\mu_m S^2 + c\mu_y^2}{(2S^2+\mu_y^2)^2} > 0.
\end{aligned}
$$

▮

Proof of Lemma 1. Observing the best response functions $q_i(q_j)$, we know that in equilibrium $q_1^{YY}|I = q_2^{YY}|I$. Therefore,

$$
q_i^{YY}|I = \frac{g'\sigma_I + \gamma\rho\mu_y\sigma_m(I-\mu_I)}{2\sigma_I S'^2} - \frac{\gamma^2\mu_y^2 q_i^{YY}|I}{2S'^2},
$$

which yields

$$
q_i^{YY}|I = \frac{g'\sigma_I + \gamma\rho\mu_y\sigma_m(I-\mu_I)}{2\sigma_I S'^2 + \sigma_I\gamma^2\mu_y^2}.
$$

Substituting the above equation into $\pi_i(q)|I$, we obtain

$$\begin{aligned}
\pi_i^{YY}|I &= (q_i^{YY}|I)\left[g^{'} + \gamma\rho\mu_y\frac{\sigma_m}{\sigma_I}(I-\mu_I) - (S^{'2}+\gamma^2\mu_y^2)(q_i^{YY}|I)\right] \\
&= \frac{g^{'}\sigma_I + \gamma\rho\mu_y\sigma_m(I-\mu_I)}{2\sigma_I S^{'2} + \sigma_I\gamma^2\mu_y^2}\left[g^{'} + \gamma\rho\mu_y\frac{\sigma_m}{\sigma_I}(I-\mu_I) - \frac{(S^{'2}+\gamma^2\mu_y^2)(g^{'}\sigma_I + \gamma\rho\mu_y\sigma_m(I-\mu_I))}{2\sigma_I S^{'2} + \sigma_I\gamma^2\mu_y^2}\right] \\
&= \frac{S^{'2}[g^{'}\sigma_I + \gamma\rho\mu_y\sigma_m(I-\mu_I)]^2}{\sigma_I^2(2S^{'2}+\gamma^2\mu_y^2)^2}.
\end{aligned}$$

▮

Proof of Proposition 2. From Lemma 1, it is easy to obtain the ex ante equilibria (2.4) and (2.5). Note that only $g^{'}$ is dependent of α and β, and $g^{'}$ is increasing in α and decreasing in β. Therefore, both q_i^{YY} and π_i^{YY} are increasing in α and decreasing in β. Recall that $S^{'2} = \sigma_y^2 + (\gamma\mu_y)^2$ and $g^{'} = \alpha\gamma\mu_m\mu_y - \beta c$. Thus, $\frac{dS^{'2}}{d\gamma} = 2\gamma\mu_y^2$ and $\frac{dg^{'}}{d\gamma} = \alpha\mu_m\mu_y$. Therefore,

$$\begin{aligned}
\frac{d\pi_i^{YY}}{d\gamma} &= \frac{[\frac{dS^{'2}}{d\gamma}(g^{'2} + \gamma^2\mu_y^2\rho^2\sigma_m^2) + 2S^{'2}(\alpha\mu_m\mu_y g^{'} + \gamma\rho^2\mu_y^2\sigma_m^2)](2S^{'2}+\gamma^2\mu_y^2)}{(2S^{'2}+\gamma^2\mu_m^2)^3} \\
&\quad - \frac{-4S^{'2}(g^{'2} + \gamma^2\mu_y^2\rho^2\sigma_m^2)(\frac{dS^{'2}}{d\gamma} + \gamma\mu_y^2)}{(2S^{'2}+\gamma^2\mu_m^2)^3} \\
&= \frac{(2\gamma^2\mu_y^2\sigma_y^2 + 4\sigma_y^4)\gamma\rho^2\mu_y^2\sigma_m^2 + 2g^{'}(\alpha\gamma^2\mu_m\mu_y^3\sigma_y^2 + 2\alpha\mu_m\mu_y\sigma_y^4 + 3\gamma\mu_y^2\beta c S^{'2} + \gamma\mu_y^2\sigma_y^2\beta c)}{(2S^{'2}+\gamma^2\mu_y^2)^3} \quad \text{(A.1)} \\
&> 0.
\end{aligned}$$

Finally, it is easy to verify that π_i^{YY} is increasing in ρ^2 and σ_m. ▮

Proof of Lemma 2. From (2.8), given q_2,

$$(q_1|I=\mu_I) = \frac{g^{'}}{2S^{'2}} - \frac{\gamma\mu_y^2 q_2}{2S^{'2}}.$$

Substituting it into (2.9),

$$q_2 = \frac{g}{2S^2} - \frac{\gamma\mu_y^2(g^{'} - \gamma\mu_y^2 q_2)}{4S^2 S^{'2}}.$$

As the production quantity must be nonnegative, solving the above equation we obtain

$$q_2^{YN} = \begin{cases} \frac{2S^{'2}g - \gamma\mu_y^2 g^{'}}{4S^2S^{'2} - \gamma^2\mu_y^4}, & if\ 2S^{'2}g - \gamma\mu_y^2 g^{'} > 0, \\ 0, & otherwise. \end{cases}$$

Because $2S^{'2}g - \gamma\mu_y^2 g^{'} = \mu_y^2[2(\gamma^2 + C_y^2)g - \gamma g^{'}]$, $2S^{'2}g - \gamma\mu_y^2 g^{'} > 0$ is equivalent to $C_y^2 > \frac{\gamma g^{'}}{2g} - \gamma^2$. Plugging q_2^{YN} into (2.8), we obtain (2.10). Then

$$(q_1^{YN}|I = \mu_I) = \begin{cases} \frac{2S^2 g^{'} - \gamma\mu_y^2 g}{4S^2S^{'2} - \gamma^2\mu_y^4}, & if\ C_y^2 > \frac{\gamma g^{'}}{2g} - \gamma^2, \\ \frac{g^{'}}{2S^{'2}}, & otherwise. \end{cases}$$

When $C_y^2 > \frac{\gamma g'}{2g} - \gamma^2$, from (2.6) and (2.7), we can get

$$\begin{aligned}
\pi_2^{YN} &= q_2^{YN}\left[g - (S^2 q_2^{YN} + \gamma\mu_y^2(q_1^{YN}|I = \mu_I))\right] \\
&= \frac{2S'^2 g - \gamma\mu_y^2 g'}{4S^2S'^2 - \gamma^2\mu_y^4}\left[g - \frac{2S^2S'^2 g - \gamma\mu_y^2 S^2 g'}{4S^2S'^2 - \gamma^2\mu_y^4} - \frac{2\gamma\mu_y^2 S^2 g' - \gamma^2\mu_y^4 g}{4S^2S'^2 - \gamma^2\mu_y^4}\right] \\
&= \frac{S^2(2S'^2 g - \gamma\mu_y^2 g')^2}{(4S^2S'^2 - \gamma^2\mu_y^4)^2}.
\end{aligned}$$

$$\begin{aligned}
\pi_1^{YN}|I &= (q_1^{YN}|I)[g' + \gamma\rho\mu_y\frac{\sigma_m}{\sigma_I}(I - \mu_I) - ((q_1^{YN}|I)S'^2 + \gamma\mu_y^2 q_2^{YN})] \\
&= \left[\frac{2S^2 g' - \gamma\mu_y^2 g}{4S^2S'^2 - \gamma^2\mu_y^4} + \frac{\gamma\rho\mu_y\sigma_m(I - \mu_I)}{2\sigma_I S'^2}\right]\left[g' + \gamma\rho\mu_y\frac{\sigma_m}{\sigma_I}(I - \mu_I) - \left(\frac{2S^2S'^2 g' - \gamma\mu_y^2 S'^2 g}{4S^2S'^2 - \gamma^2\mu_y^4}\right.\right. \\
&\quad \left.\left. + \frac{\gamma\rho\mu_y\sigma_m(I - \mu_I)}{2\sigma_I} + \frac{2S'^2\gamma\mu_y^2 g - \gamma^2\mu_y^4 g'}{4S^2S'^2 - \gamma^2\mu_y^4}\right)\right] \\
&= S'^2\left[\frac{2S^2 g' - \gamma\mu_y^2 g}{4S^2S'^2 - \gamma^2\mu_y^4} + \frac{\gamma\rho\mu_y\sigma_m(I - \mu_I)}{2\sigma_I S'^2}\right]^2.
\end{aligned}$$

When $C_y^2 \le \frac{\gamma g'}{2g} - \gamma^2$, it can be easily shown that $\pi_2^{YN} = 0$ and

$$\pi_1^{YN}|I = (q_1^{YN}|I)\left[g' + \gamma\rho\mu_y\frac{\sigma_m}{\sigma_I}(I - \mu_I) - (q_1^{YN}|I)S'^2\right] = \frac{[g'\sigma_I + \gamma\rho\mu_y\sigma_m(I - \mu_I)]^2}{4\sigma_I^2 S'^2}.$$

∎

Proof of Corollary 3. Suppose $\gamma = 1$. Then, $S^2 = S'^2$. When $C_y^2 > \frac{\gamma g'}{2g} - \gamma^2$, according to Lemma 2 and Proposition 3,

$$q_1^{YN} - q_2^{YN} = \frac{2S^2 g' - \gamma\mu_y^2 g - 2S'^2 g + \gamma\mu_y^2 g'}{4S^2S'^2 - \gamma^2\mu_y^4} = \frac{(2S^2 + \gamma\mu_y^2)(g' - g)}{4S^2S'^2 - \gamma^2\mu_y^4} \ge 0.$$

When $C_y^2 \le \frac{\gamma g'}{2g} - \gamma^2$, $q_2^{YN} = 0$. Thus, when $\gamma = 1$, $q_1^{YN} \ge q_2^{YN}$. ∎

Proof of Lemma 4. We first consider the scenario $C_y^2 > \frac{\gamma g^{'}}{2g} - \gamma^2$. Let

$$L_1 = \frac{S^{'} g^{'}}{(2S^{'2} + \gamma^2 \mu_y^2)}, \text{ and } L_2 = \frac{S^{'}(2S^2 g^{'} - \gamma \mu_y^2 g)}{(4S^2 S^{'2} - \gamma^2 \mu_y^4)}.$$

Then, from (2.5) and (2.15), we have

$$\pi_1^{YY} = L_1^2 + \frac{S^{'2} \gamma^2 \rho^2 \mu_y^2 \sigma_m^2}{(2S^{'2} + \gamma^2 \mu_y^2)^2}, \quad \text{(A.2)}$$

$$\pi_1^{YN} = L_2^2 + \frac{\gamma^2 \rho^2 \mu_y^2 \sigma_m^2}{4S^{'2}}. \quad \text{(A.3)}$$

Note that $\pi_1^{NY} = \pi_2^{YN}$. From (2.13), we have

$$\begin{aligned} \sqrt{\pi_1^{NN}} - \sqrt{\pi_1^{NY}} &= \frac{Sg}{(2S^2 + \mu_y^2)} - \frac{S(2S^{'2} g - \gamma \mu_y^2 g^{'})}{(4S^2 S^{'2} - \gamma^2 \mu_y^4)} \\ &= \frac{[2\mu_y^2 (S^2 \gamma g^{'} - S^{'2} g) + \gamma \mu_y^4 (g^{'} - \gamma g)] S}{(2S^2 + \mu_y^2)(4S^2 S^{'2} - \gamma^2 \mu_y^4)}, \quad \text{(A.4)} \\ L_2 - L_1 &= \frac{S^{'}(2S^2 g^{'} - \gamma \mu_y^2 g)}{(4S^2 S^{'2} - \gamma^2 \mu_y^4)} - \frac{S^{'} g^{'}}{(2S^{'2} + \gamma^2 \mu_y^2)} \\ &= \frac{[2\gamma \mu_y^2 (S^2 \gamma g^{'} - S^{'2} g) + \gamma^2 \mu_y^4 (g^{'} - \gamma g)] S^{'}}{(4S^2 S^{'2} - \gamma^2 \mu_y^4)(2S^{'2} + \gamma^2 \mu_y^2)}, \quad \text{(A.5)} \end{aligned}$$

As $\alpha \geq 1$, $\beta \leq 1$, and $\gamma \geq 1$,

$$\begin{aligned} g^{'} - \gamma g &= \gamma \mu_m \mu_y (\alpha - 1) - (\beta - \gamma) c \geq 0, \\ S^2 \gamma g^{'} - S^{'2} g &= \gamma(\sigma_y^2 + \mu_y^2)(\alpha \gamma \mu_m \mu_y - \beta c) - (\sigma_y^2 + \gamma^2 \mu_y^2)(\mu_m \mu_y - c) \\ &= \sigma_y^2 [\gamma(\alpha \gamma \mu_m \mu_y - \beta c) - (\mu_m \mu_y - c)] + (\alpha - 1)\gamma^2 \mu_m \mu_y^3 + (\gamma - \beta)\gamma \mu_y^2 c \\ &\geq 0. \end{aligned}$$

Thus, $\pi_1^{NN} \geq \pi_1^{NY}$ and $L_2 \geq L_1$. Because

$$\frac{\gamma^2\rho^2\mu_y^2\sigma_m^2}{4S'^2} - \frac{S'^2\gamma^2\rho^2\mu_y^2\sigma_m^2}{(2S'^2+\gamma^2\mu_y^2)^2} = S'^2\gamma^2\rho^2\mu_y^2\sigma_m^2\left(\frac{1}{2S'^2} - \frac{1}{2S'^2+\gamma^2\mu_y^2}\right)\left(\frac{1}{2S'^2} + \frac{1}{2S'^2+\gamma^2\mu_y^2}\right) \geq 0,$$

from (A.2) and (A.3), we have $\pi_1^{YN} - \pi_1^{YY} \geq L_2^2 - L_1^2 \geq 0$. Note that when $\alpha = \gamma = \beta = 1$ and $\rho = 0$, $\pi_i^{YY} = \pi_i^{NN}$. Then, based on Proposition 2, we can derive that $\pi_1^{YN} \geq \pi_1^{YY} \geq \pi_1^{NN} \geq \pi_1^{NY}$. Naturally, due to symmetry, $\pi_2^{NY} \geq \pi_2^{YY} \geq \pi_2^{NN} \geq \pi_2^{YN}$.

Below, we prove that $\pi_1^{YN} - \pi_1^{NN} \geq \pi_1^{YY} - \pi_1^{NY}$. From (A.4) and (A.5), we get

$$\begin{aligned}
\frac{\pi_1^{YN} - \pi_1^{YY}}{\pi_1^{NN} - \pi_1^{NY}} &\geq \frac{L_2^2 - L_1^2}{\pi_1^{NN} - \pi_1^{NY}} \\
&= \frac{L_2 - L_1}{\sqrt{\pi_1^{NN}} - \sqrt{\pi_1^{NY}}}\frac{L_2 + L_1}{\sqrt{\pi_1^{NN}} + \sqrt{\pi_1^{NY}}} \\
&= \frac{\gamma S'(2S^2+\mu_y^2)}{S(2S'^2+\gamma^2\mu_y^2)}\frac{L_2 + L_1}{\sqrt{\pi_1^{NN}} + \sqrt{\pi_1^{NY}}}.
\end{aligned}$$

We can show that

$$\begin{aligned}
\sqrt{\pi_1^{NN}} + \sqrt{\pi_1^{NY}} &= \frac{Sg}{(2S^2+\mu_y^2)} + \frac{S(2S'^2g - \gamma\mu_y^2 g')}{(4S^2S'^2 - \gamma^2\mu_y^4)} \\
&= \frac{[8S^2S'^2g - 2\mu_y^2(S^2\gamma g' - S'^2 g) - \gamma\mu_y^4(g' + \gamma g)]S}{(2S^2+\mu_y^2)(4S^2S'^2 - \gamma^2\mu_y^4)} \\
L_2 + L_1 &= \frac{S'(2S^2g' + \gamma\mu_y^2 g)}{(4S^2S'^2 - \gamma^2\mu_y^4)} + \frac{S'g'}{(2S'^2+\gamma^2\mu_y^2)} \\
&= \frac{[8S^2S'^2g' + 2\gamma\mu_y^2(S^2\gamma g' - S'^2 g) - \gamma^2\mu_y^4(g' + \gamma g)]S'}{(4S^2S'^2 - \gamma^2\mu_y^4)(2S'^2+\gamma^2\mu_y^2)}.
\end{aligned}$$

Because $g^{'} \geq g\gamma$ and $S^2\gamma g^{'} - S^{'2} g \geq 0$,

$$\frac{8S^2S^{'2}g^{'}}{\gamma} + 2\mu_y^2(S^2\gamma g^{'} - S^{'2}g) - \gamma\mu_y^4(g^{'}+\gamma g) \geq 8S^2S^{'2}g - 2\mu_y^2(S^2\gamma g^{'} - S^{'2}g) - \gamma\mu_y^4(g^{'}+\gamma g),$$

which yields

$$\frac{L_2 + L_1}{\sqrt{\pi_1^{NN}} + \sqrt{\pi_1^{NY}}} \geq \frac{\gamma S^{'}(2S^2 + \mu_y^2)}{S(2S^{'2} + \gamma^2\mu_y^2)}.$$

Moreover, because

$$\begin{aligned}
\gamma^2 S^{'2}(2S^2 + \mu_y^2)^2 - S^2(2S^{'2} + \gamma^2\mu_y^2)^2 &= (4S^2S^{'2} - \gamma^2\mu_y^4)(\gamma^2 S^2 - S^{'2}) \\
&= (4\sigma_y^4 + 4\gamma^2\mu_y^2\sigma_y^2 + 4\mu_y^2\sigma_y^2 + 3\gamma^2\mu_y^4)(\gamma^2 - 1)\sigma_y^2 \\
&> 0,
\end{aligned}$$

$\frac{\gamma^2 S^{'2}(2S^2+\mu_y^2)^2}{S^2(2S^{'2}+\gamma^2\mu_y^2)^2} \geq 1$. Thus, $\pi_1^{YN} - \pi_1^{YY} \geq \pi_1^{NN} - \pi_1^{NY}$. Equivalently, $\pi_1^{YN} - \pi_1^{NN} \geq \pi_1^{YY} - \pi_1^{NY}$. Again, by symmetry, $\pi_2^{NY} - \pi_2^{NN} \geq \pi_2^{YY} - \pi_2^{YN}$.

Now we consider the scenario $C_y^2 \leq \frac{\gamma g^{'}}{2g} - \gamma^2$. As $C_y^2 \geq 0$, $g^{'} \geq 2\gamma g$. Thus,

$$\begin{aligned}
\pi_1^{YN} - \pi_1^{YY} &= \frac{g^{'2} + \gamma^2\rho^2\mu_y^2\sigma_m^2}{4S^{'2}} - \frac{S^{'2}(g^{'2} + \gamma^2\rho^2\mu_y^2\sigma_m^2)}{(2S^{'2} + \gamma^2\mu_y^2)^2} \\
&= \frac{(g^{'2} + \gamma^2\rho^2\mu_y^2\sigma_m^2)\gamma^2\mu_y^2(4S^{'2} + \gamma^2\mu_y^2)}{4S^{'2}(2S^{'2} + \gamma^2\mu_y^2)^2} > 0.
\end{aligned}$$

When $C_y^2 \leq \frac{\gamma g^{'}}{2g} - \gamma^2$, $\pi_1^{NY} = \pi_2^{YN} = 0$. Therefore, $\pi_1^{YN} \geq \pi_1^{YY} \geq \pi_1^{NN} \geq \pi_1^{NY}$. By symmetry, $\pi_2^{NY} \geq \pi_2^{YY} \geq \pi_2^{NN} \geq \pi_2^{YN}$. Next, we can show that

$$\begin{aligned}
\pi_1^{YN} - \pi_1^{YY} - (\pi_1^{NN} - \pi_1^{NY}) &= \pi_1^{YN} - \pi_1^{YY} - \pi_1^{NN} \\
&\geq \frac{g^{'2}\gamma^2\mu_y^2(4S^{'2} + \gamma^2\mu_y^2)}{4S'^2(2S'^2 + \gamma^2\mu_y^2)^2} - \frac{S^2 g^2}{(2S^2 + \mu_y^2)^2} \\
&= \frac{g^{'2}\gamma^2\mu_y^2(4S^{'2} + \gamma^2\mu_y^2)(2S^2 + \mu_y^2)^2 - 4S^{'2}S^2g^2(2S^{'2} + \gamma^2\mu_y^2)^2}{4S'^2(2S^{'2} + \gamma^2\mu_y^2)^2(2S^2 + \mu_y^2)^2}.
\end{aligned}$$

Recall that $C_y^2 \leq \frac{\gamma g'}{2g} - \gamma^2$ is equivalent to $2S^{'2}g - \gamma\mu_y^2 g' \leq 0$, thus,

$$\begin{aligned}
\pi_1^{YN} - \pi_1^{YY} - (\pi_1^{NN} - \pi_1^{NY}) &\geq \frac{2S'^2 gg'\gamma(4S^{'2} + \gamma^2\mu_y^2)(2S^2 + \mu_y^2)^2 - 4S^{'2}S^2g^2(2S^{'2} + \gamma^2\mu_y^2)^2}{4S'^2(2S^{'2} + \gamma^2\mu_y^2)^2(2S^2 + \mu_y^2)^2} \\
&= \frac{2S^{'2}g[g'\gamma(4S^{'2} + \gamma^2\mu_y^2)(2S^2 + \mu_y^2)^2 - 2S^2g(2S^{'2} + \gamma^2\mu_y^2)^2]}{4S'^2(2S^{'2} + \gamma^2\mu_y^2)^2(2S^2 + \mu_y^2)^2}.
\end{aligned}$$

Note that

$$\begin{aligned}
g'\gamma(4S^{'2} + \gamma^2\mu_y^2)(2S^2 + \mu_y^2)^2 &\geq 16g'\gamma S^{'2}(S^2)^2 + 4\gamma^3 g'\mu_y^2(S^2)^2 + 4\gamma^3\mu_y^4 S^2 g', \\
2S^2g(2S^{'2} + \gamma^2\mu_y^2)^2 &= 8(S^{'2})^2S^2g + 8S^{'2}S^2g\gamma^2\mu_y^2 + 2\gamma^4\mu_y^4S^2g.
\end{aligned}$$

Because $S^2\gamma g' - S^{'2}g \geq 0$, $2S^{'2}g - \gamma\mu_y^2 g' \leq 0$ and $g' \geq 2\gamma g$,

$$\begin{aligned}
g'\gamma(4S^{'2} + \gamma^2\mu_y^2)(2S^2 + \mu_y^2)^2 - 2S^2g(2S^{'2} + \gamma^2\mu_y^2)^2 &\geq 8S^{'2}S^2(2g'\gamma S^2 - S^{'2}g) + 4\gamma^2\mu_y^2S^2(\gamma\mu_y^2g' \\
&\quad -2S^{'2}g) + 2\gamma^3\mu_y^2S^2(2g'S^2 - \gamma\mu_y^2 g) \\
&\geq 0.
\end{aligned}$$

Therefore, $\pi_1^{YN} - \pi_1^{YY} \geq \pi_1^{NN} - \pi_1^{NY}$, which is equivalent to $\pi_1^{YN} - \pi_1^{NN} \geq \pi_1^{YY} - \pi_1^{NY}$. By symmetry, $\pi_2^{NY} - \pi_2^{NN} \geq \pi_2^{YY} - \pi_2^{YN}$. ∎

Proof of Corollary 6. By noting that $\rho \neq 0$ and π_i^{NN} is independent of α, β, γ, and ρ, Corollary 6 can be easily derived from Proposition 2 and Lemma 4. ∎

Proof of Corollary 7. Note that π_1^{NN} is independent of those parameters. We only need to focus on π_1^{YY}. From (2.5) and (A.1), we have

$$\frac{\partial^2 \pi_1^{YY}}{\partial\gamma\partial\rho^2} = \frac{\gamma\mu_y^2\sigma_m^2(2\gamma^2\sigma_y^2\mu_y^2+4\sigma_y^4)}{(2S'^2+\gamma^2\mu_y^2)^3} > 0, \quad \frac{\partial^2 \pi_1^{YY}}{\partial\gamma\partial\sigma_m} = \frac{2\gamma\sigma_m\rho^2\mu_y^2(2\gamma^2\sigma_y^2\mu_y^2+4\sigma_y^4)}{(2S'^2+\gamma^2\mu_y^2)^3} > 0,$$

$$\frac{\partial^2 \pi_1^{YY}}{\partial\gamma\partial\alpha} = \frac{2g'(\gamma^2\mu_m\mu_y^3\sigma_y^2+2\mu_m\mu_y\sigma_y^4)+2\mu_m\mu_y\gamma(\alpha\gamma^2\mu_m\mu_y^3\sigma_y^2+2\alpha\mu_m\mu_y\sigma_y^4+3\gamma\mu_y^2\beta cS'^2+\gamma\mu_y^2\sigma_y^2\beta}{(2S'^2+\gamma^2\mu_y^2)^3}$$

$$> 0.$$

$$\frac{\partial^2 \pi_1^{YY}}{\partial\alpha\partial\rho^2} = \frac{\partial^2 \pi_1^{YY}}{\partial\beta\partial\rho^2} = 0.$$

Moreover,

$$\begin{aligned}
\frac{\partial^2 \pi_1^{YY}}{\partial\beta\partial\gamma} &= \frac{-2c(\alpha\gamma^2\mu_m\mu_y^3\sigma_y^2+2\alpha\mu_m\mu_y\sigma_y^4+3\gamma\mu_y^2\beta cS'^2+\gamma\mu_y^2\sigma_y^2\beta c)+6g'\gamma\mu_y^2cS'^2+2g'\gamma\mu_y^2\sigma_y^2c}{(2S'^2+\gamma^2\mu_y^2)^3} \\
&= \frac{-2c\alpha\gamma^2\mu_m\mu_y^3\sigma_y^2-4c\alpha\mu_m\mu_y\sigma_y^4+2c(g'-\beta c)(3\gamma\mu_y^2S'^2+\gamma\mu_y^2\sigma_y^2)}{(2S'^2+\gamma^2\mu_y^2)^3} \\
&= \frac{2c[-\alpha\gamma^2\mu_m\mu_y^3\sigma_y^2-2\alpha\mu_m\mu_y\sigma_y^4+(\alpha\gamma\mu_m\mu_y-2\beta c)(3\gamma\mu_y^2S'^2+\gamma\mu_y^2\sigma_y^2)]}{(2S'^2+\gamma^2\mu_y^2)^3} \\
&= \frac{2c[-2\alpha\mu_m\mu_y\sigma_y^4+3\alpha\gamma^2\mu_m\mu_y^3S'^2-2\beta c(3\gamma\mu_y^2S'^2+\gamma\mu_y^2\sigma_y^2)]}{(2S'^2+\gamma^2\mu_y^2)^3} \\
&= \frac{2c[\alpha\mu_m\mu_y(3\gamma^2\mu_y^2S'^2-2\sigma_y^4)-2\beta c(3\gamma\mu_y^2S'^2+\gamma\mu_y^2\sigma_y^2)]}{(2S'^2+\gamma^2\mu_y^2)^3} \\
&= \frac{2c[\alpha\mu_m\mu_y^5(3\gamma^4+3\gamma^2C_y^2-2C_y^4)-2\beta c(3\gamma\mu_y^2S'^2+\gamma\mu_y^2\sigma_y^2)]}{(2S'^2+\gamma^2\mu_y^2)^3} \\
&= \frac{2c[-2\alpha\mu_m\mu_y^5\gamma^4(\frac{C_y^2}{\gamma^2}+\frac{\sqrt{33}-3}{4})(\frac{C_y^2}{\gamma^2}-\frac{\sqrt{33}+3}{4})-2\beta c(3\gamma\mu_y^2S'^2+\gamma\mu_y^2\sigma_y^2)]}{(2S'^2+\gamma^2\mu_y^2)^3}.
\end{aligned}$$

Let $\bar{C}_y = (\sqrt{\sqrt{33}+3})/2$. Obviously, when $C_y/r > \bar{C}_y$, the above equation is definitely negative. Otherwise, because the term inside the bracket in the numerator is decreasing in c, there exists a $\bar{c}$ such that $\frac{\partial^2 \pi_1^{YY}}{\partial\beta\partial\gamma} < 0$ if and only if $c > \bar{c}$. ∎

Proof of Corollary 8. Because $\alpha = \beta = \gamma = 1$, $g = g'$, $S^2 = S'^2$, and $\gamma g' - 2\gamma^2 g \leq 0$. Based on (2.13), we then have

$$\pi_1^{NY} = \frac{S^2(2S'^2 g - \gamma\mu_y^2 g')^2}{(4S^2S'^2 - \gamma^2\mu_y^4)^2} = \frac{S^2 g^2}{(2S^2 + \mu_y^2)^2} = \pi_1^{NN}.$$

By symmetry, we can obtain that $\pi_2^{NN} = \pi_2^{YN}$. Furthermore, because $\rho \neq 0$, it follows easily from Lemma 4 that $\pi_i^{YY} > \pi_i^{NN}$, $\pi_1^{YN} - \pi_1^{NN} > \pi_1^{YY} - \pi_1^{NY}$, and $\pi_2^{NY} - \pi_2^{NN} > \pi_2^{YY} - \pi_2^{YN}$. Finally, due to $\pi_1^{YY} > \pi_1^{NY}$ and $\pi_1^{YN} > \pi_1^{NN}$, the best response of farmer 1 is to invest. By symmetry, farmer 2's best response is the same as farmer 1. Thus, (Y, Y) is the unique equilibrium. ∎

Proof of Proposition 4. From (2.16) and (2.17), we have

$$\pi_1^{YN} = \frac{S^2 g^2}{(2S^2 + \mu_y^2)^2} + \frac{\rho^2\mu_y^2\sigma_m^2}{4S^2}, \text{ and } \pi_2^{YN} = \frac{S^2 g^2}{(2S^2 + \mu_y^2)^2}.$$

Thus,

$$\begin{aligned}
\pi_1^{YY} + \pi_2^{YY} - (\pi_1^{YN} + \pi_2^{YN}) &= \frac{2S^2(g^2 + \rho^2\mu_y^2\sigma_m^2)}{(2S^2 + \mu_y^2)^2} - \frac{2S^2 g^2}{(2S^2 + \mu_y^2)^2} - \frac{\rho^2\mu_y^2\sigma_m^2}{4S^2} \\
&= \frac{\rho^2\mu_y^2\sigma_m^2(2\sqrt{2}S^2 - 2S^2 - \mu_y^2)(2\sqrt{2}S^2 + 2S^2 + \mu_y^2)}{4S^2(2S^2 + \mu_y^2)^2} \\
&= \frac{\rho^2\sigma_m^2\mu_y^4[2(\sqrt{2}-1)C_y^2 + 2\sqrt{2} - 3](2\sqrt{2}S^2 + 2S^2 + \mu_y^2)}{4S^2(2S^2 + \mu_y^2)^2}.
\end{aligned}$$

Therefore, $\pi_1^{YY} + \pi_2^{YY} > \pi_1^{YN} + \pi_2^{YN}$ if and only if $C_y > \sqrt{\frac{\sqrt{2}-1}{2}}$. ∎

Proof of Corollary 9. By noting that $\alpha = \beta = \gamma = 1$ and $P_i = M - (z_i q_i + t z_j q_j)$, $i, j = 1, 2, i \neq j$, we can obtain the farmers' expected profits associated with the four subgames, which are summarized in following table.

Table A.1: Summary of the results under heterogenous products

	Expected profit
(N,N)	$\pi_i(q_i) = gq_i - S^2 q_i^2 - t\mu_y^2 q_i q_j,\ i,j = 1,2, i \neq j,$
(Y,Y)	$\pi_i(q_i \mid I) = gq_i + \rho\mu_y \frac{\sigma_m}{\sigma_I}(I - \mu_I)q_i - S^2 q_i^2 - t\mu_y^2 q_i q_j \mid I$
(Y,N)	$\pi_1(q_1) = gq_1 + \rho\mu_y \frac{\sigma_m}{\sigma_I}(I - \mu_I)q_1 - S^2 q_1^2 - t\mu_y^2 q_1 q_2 \mid I$
	$\pi_2(q_2) = gq_2 - S^2 q_2^2 - t(\mu_y^2 q_1 \mid I = \mu_I)q_2$
(N,Y)	$\pi_1(q_1) = gq_1 - S^2 q_1^2 - tq_1(\mu_y^2 q_2 \mid I = \mu_I)$
	$\pi_2(q_2) = gq_2 + \rho\mu_y \frac{\sigma_m}{\sigma_I}(I - \mu_I)q_2 - S^2 q_2^2 - t\mu_y^2 q_1 q_2 \mid I$

By considering the first order condition, we can obtain the equilibrium outcomes, which are summarized in Table A.2.

Table A.2: Summary of the results under heterogenous products

	Best response	Equilibrium outcomes (ex post)
(N,N)	$q_i(q_j) = \frac{g - t\mu_y^2 q_j}{2S^2}$	$q_i^{NN} = \frac{g}{2S^2 + t\mu_y^2}$, $\pi_i^{NN} = \frac{S^2 g^2}{(2S^2 + t\mu_y^2)^2}$
(Y,Y)	$q_i(q_j) \mid I = \frac{g\sigma_I + \rho\mu_y\sigma_m(I - \mu_I)}{2\sigma_I S^2} - \frac{t\mu_y^2 q_2}{2S^2}$	$q_i^{YY} = \frac{g}{2S^2 + t\mu_y^2}$, $\pi_i^{YY} = \frac{S^2(g^2 + \rho^2\mu_y^2\sigma_m^2)}{(2S^2 + t\mu_y^2)^2}$
(Y,N)	$q_1(q_2) \mid I = \frac{g\sigma_I + \rho\mu_y\sigma_m(I - \mu_I)}{2\sigma_I S^2} - \frac{t\mu_y^2 q_2}{2S^2}$	$q_1^{YN} = \frac{g}{2S^2 + t\mu_y^2}$, $\pi_1^{YN} = \frac{S^2 g^2}{(2S^2 + t\mu_y^2)^2} + \frac{\rho^2\mu_y^2\sigma_m^2}{4S^4}$
	$q_2(q_1) = \frac{g - t\mu_y^2(q_1 \mid I = \mu_I)}{2S^2}$	$q_2^{YN} = \frac{g}{2S^2 + t\mu_y^2}$, $\pi_2^{YN} = \frac{S^2 g^2}{(2S^2 + t\mu_y^2)^2}$
(N,Y)	$q_1(q_2) = \frac{g - t\mu_y^2(q_2 \mid I = \mu_I)}{2S^2}$	$q_1^{NY} = \frac{g}{2S^2 + t\mu_y^2}$, $\pi_1^{NY} = \frac{S^2 g^2}{(2S^2 + t\mu_y^2)^2}$
	$q_2(q_1) \mid I = \frac{g\sigma_I + \rho\mu_y\sigma_m(I - \mu_I)}{2\sigma_I S^2} - \frac{t\mu_y^2 q_1}{2S^2}$	$q_2^{NY} = \frac{g}{2S^2 + t\mu_y^2}$, $\pi_2^{NY} = \frac{S^2 g^2}{(2S^2 + t\mu_y^2)^2} + \frac{\rho^2\mu_y^2\sigma_m^2}{4S^4}$

It is easy to see that $\pi_i^{YY} > \pi_i^{NN}$ and $\pi_1^{YY} + \pi_2^{YY} > \pi_1^{NN} + \pi_2^{NN}$. Furthermore, due to symmetry, to check whether the provision of market information is welfare maximizing, we only need to compare the total welfare of farmers under strategies

(Y, Y) and (Y, N). By direct comparison,

$$\begin{aligned}
\pi_1^{YY} + \pi_2^{YY} - (\pi_1^{YN} + \pi_2^{YN}) &= \frac{2\rho^2\mu_y^2\sigma_m^2}{(2S^2 + t\mu_y^2)^2} - \frac{\rho^2\mu_y^2\sigma_m^2}{4S^4} \\
&= \frac{\rho^2\mu_y^2\sigma_m^2[8S^4 - (2S^2 + t\mu_y^2)^2]}{4S^4(2S^2 + t\mu_y^2)^2} \\
&= \frac{\rho^2\mu_y^2\sigma_m^2[2\sqrt{2}S^2 - (2S^2 + t\mu_y^2)][2\sqrt{2}S^2 + (2S^2 + t\mu_y^2)]}{4S^4(2S^2 + t\mu_y^2)^2} \\
&= \frac{\rho^2\mu_y^4\sigma_m^2[2(\sqrt{2} - 1)C_y^2 + 2(\sqrt{2} - 1) - t][2\sqrt{2}S^2 + (2S^2 + t\mu_y^2)]}{4S^4(2S^2 + t\mu_y^2)^2}.
\end{aligned}$$

Thus, when $t < 2(\sqrt{2} - 1)$, the above equation is always positive. When $t \geq 2(\sqrt{2} - 1)$, $\pi_1^{YY} + \pi_2^{YY} > \pi_1^{NY} + \pi_2^{NY}$ if and only if $C_y > \sqrt{\frac{t-2(\sqrt{2}-1)}{2(\sqrt{2}-1)}}$, which can be simplified as $C_y > \sqrt{\frac{(\sqrt{2}+1)t-2}{2}}$. ∎

Proof of Proposition 5. The first statement follows easily from the proof of Lemma 4. For the second statement, first consider the case that $K \leq \pi_1^{YY} - \pi_1^{NY}$. According to Lemma 4, when $K \leq \pi_1^{YY} - \pi_1^{NY}$, $\pi_1^{YY} - K > \pi_1^{NY}$, and $\pi_1^{YN} - K > \pi_1^{NN}$, which implies that the best response of farmer 1 is to adopt and invest. By symmetry, farmer 2's best response is also to adopt and invest. Therefore, (Y, Y) is the unique equilibrium. The other two cases can be derived similarly and we omit the details here. ∎

Appendix B

Proofs of Chapter 3

Proof of Lemma 11. Taking the first order condition (FOC) of $o(\mu)$ over μ yields

$$\frac{do(\mu)}{d\mu} = 1 - \delta(\mu) - \mu\delta^{'}(\mu) = 0,$$

which can be rewritten as

$$(1 - \delta(\mu))(1 - \mu g(\mu)) = 0.$$

As $g(\mu)$ is increasing in μ, $do(\mu)/d\mu$ crosses zero only once from above. Therefore, $o(\mu)$ is quasi-concave in μ and the optimal service rate μ^* solves the $\mu^* g(\mu^*) = 1$.

Next, we prove that when $\mu \leq \mu^*$, $o(\mu)$ is concave in μ. Note that as $g(\mu)$ is increasing in μ,

$$\frac{dg(\mu)}{d\mu} = \frac{\delta^{''}(\mu)(1 - \delta(\mu)) + (\delta^{'}(\mu))^2}{(1 - \delta(\mu))^2} > 0, \tag{B.1}$$

which yields $\delta^{''}(\mu)(1 - \delta(\mu)) + (\delta^{'}(\mu))^2 > 0$. And we can show that

$$\begin{aligned}
\frac{d^2 o(\mu)}{d\mu^2} &= -2\delta^{'}(\mu) - \mu\delta^{''}(\mu) \\
&= -2\delta^{'}(\mu) + \frac{\mu(\delta^{'}(\mu))^2}{1 - \delta(\mu)} - \frac{\mu\delta^{''}(\mu)(1 - \delta(\mu)) + \mu(\delta^{'}(\mu))^2}{1 - \delta(\mu)} \\
&= -\delta^{'}(\mu) - \frac{\delta^{'}(\mu)}{1 - \delta(\mu)}\frac{do(\mu)}{d\mu} - \frac{\mu\delta^{''}(\mu)(1 - \delta(\mu)) + \mu(\delta^{'}(\mu))^2}{1 - \delta(\mu)}.
\end{aligned}$$

As $o(\mu)$ is increasing in μ when $\mu \leq \mu^*$, the above equation is negative for all $\mu \leq \mu^*$. Therefore, $o(\mu)$ is concave in μ for $\mu \leq \mu^*$. ∎

Proof of Proposition 6. Using the fact that $(1-\delta(\mu))o^{'}(\mu) = (1-\delta(\mu))^2 - o(\mu)\delta^{'}(\mu)$, we have

$$\frac{dW(\lambda,\,\mu)}{d\mu} = -\frac{\delta^{'}(\mu)}{o(\mu)-\lambda} - \frac{(1-\delta(\mu))o^{'}(\mu)}{(o(\mu)-\lambda)^2} = \frac{\lambda\delta^{'}(\mu)-(1-\delta(\mu))^2}{(o(\mu)-\lambda)^2} \quad \text{(B.2)}$$

$$\left.\frac{d^2W(\lambda,\,\mu)}{d\mu^2}\right|_{\frac{dW(\lambda,\,\mu)}{d\mu}=0} = \frac{\lambda\delta^{''}(\mu)+2(1-\delta(\mu))\delta^{'}(\mu)}{(o(\mu)-\lambda)^2}$$

$$= \frac{\lambda\delta^{''}(\mu)(1-\delta(\mu))+\lambda\delta^{'2}(\mu)}{(1-\delta(\mu))(o(\mu)-\lambda)^2} + \frac{2(1-\delta(\mu))^2\delta^{'}(\mu)-\lambda\delta^{'2}(\mu)}{(1-\delta(\mu))(o(\mu)-\lambda)^2}$$

$$= \frac{\lambda\delta^{''}(\mu)(1-\delta(\mu))+\lambda\delta^{'2}(\mu)}{(1-\delta(\mu))(o(\mu)-\lambda)^2} + \frac{(1-\delta(\mu))^2\delta^{'}(\mu)}{(1-\delta(\mu))(o(\mu)-\lambda)^2} > 0.$$

Therefore, $W(\lambda,\,\mu)$ is quasi-convex in μ. And $\mu_{\mathcal{W}}$ satisfies $\lambda\delta^{'}(\mu_{\mathcal{W}})-(1-\delta(\mu_{\mathcal{W}}))^2 = 0$, which can be rewritten as

$$(1-\delta(\mu_{\mathcal{W}}))^2\left[\frac{\lambda}{o(\mu_{\mathcal{W}})}\frac{\mu_{\mathcal{W}}\delta^{'}(\mu_{\mathcal{W}})}{1-\delta(\mu_{\mathcal{W}})}-1\right] = 0.$$

Equivalently, $\mu_{\mathcal{W}}g(\mu_{\mathcal{W}})\rho_T(\lambda,\mu_{\mathcal{W}}) = 1$.

Besides, from (B.2) we can easily show that

$$\left.\frac{dW(\lambda,\,\mu)}{d\mu}\right|_{\mu=\mu^*} = -\frac{\delta^{'}(\mu^*)}{o(\mu^*)-\lambda} < 0.$$

Thus, the service rate that minimizes $W(\lambda,\,\mu)$ is larger than μ^*. Furthermore,

$$\frac{dTW(\lambda,\,\mu)}{d\mu} = -\frac{o^{'}(\mu)}{(o(\mu)-\lambda)^2},$$

which equals zero if $\mu = \mu^*$. And according to Lemma 11,

$$\left.\frac{d^2TW(\lambda,\, \mu)}{d\mu^2}\right|_{\mu=\mu^*} \;=\; -\frac{o''(\mu^*)}{(o(\mu^*)-\lambda)^2} > 0.$$

Therefore, $TW(\lambda,\, \mu)$ is also quasi-convex in μ and attains its minimum at $\mu = \mu^*$. Finally, according to Lemma 11, it is easy to see that the total utilization rate $\rho_T(\lambda,\, \mu)$ is convex in μ and attains its minimum at μ^*. ∎

Proof of Proposition 8. Substituting (3.6) into $SW(\lambda,\, \mu)$, we have

$$SW(\lambda,\, \mu) = \lambda\left[R - \frac{\theta}{o(\mu)-\lambda}\right] - \beta(\Lambda - \lambda). \tag{B.3}$$

Taking the FOC of $SW(\lambda,\, \mu)$ with respect to μ yields

$$\frac{dSW(\lambda,\, \mu)}{d\mu} = \frac{\lambda\theta o'(\mu)}{[o(\mu)-\lambda]^2} = 0 \;\Rightarrow\; o'(\mu) = 0.$$

Therefore, $\mu_b = \mu^*$. Furthermore, as $o(\mu)$ is concave in μ for $\mu \le \mu^*$,

$$\left.\frac{d^2SW(\lambda,\, \mu)}{d\mu^2}\right|_{\mu=\mu_b} = \frac{\lambda\theta o''(\mu_b)}{[o(\mu_b)-\lambda]^2} - \frac{2\lambda\theta(o'(\mu_b))^2}{[o(\mu_b)-\lambda]^3} < 0,$$

which guarantees that $SW(\lambda,\, \mu)$ is quasi-concave in μ and therefore, $\mu_b = \mu^*$ maximizes $SW(\lambda,\, \mu)$. Furthermore, for a fixed μ, we have

$$\frac{d^2SW(\lambda,\, \mu)}{d\lambda^2} = -\frac{2o(\mu)\theta}{[o(\mu)-\lambda]^3} < 0,$$

which shows that $SW(\lambda,\, \mu)$ is concave in λ. Substituting $\mu_b = \mu^*$ into $SW(\lambda,\, \mu)$ and taking the FOC of $SW(\lambda,\, \mu^*)$ with respect to λ yield

$$\frac{dSW(\lambda,\, \mu^*)}{d\lambda} = R - \frac{\theta}{o(\mu^*)-\lambda} - \frac{\lambda\theta}{[o(\mu^*)-\lambda]^2} + \beta = R - \frac{\theta o(\mu^*)}{[o(\mu^*)-\lambda]^2} + \beta = 0. \tag{B.4}$$

Therefore, $\lambda_b = o(\mu^*) - \sqrt{\theta o(\mu^*)/(R+\beta)}$. ∎

Proof of Proposition 9. With a slight abuse of notion, we interchangeable use $\mu_i(p)$ and μ_i, $i \in \{an, at, un, ut\}$. By noting that $o^{'}(\mu) = 1 - \delta(\mu) - \mu\delta^{'}(\mu)$, we have $o(\mu) - \mu o^{'}(\mu) = \mu^2\delta^{'}(\mu)$. Hence,

$$\begin{aligned}
\frac{\partial\lambda(\mu,\, p)}{\partial\mu} &= o^{'}(\mu) - \frac{\theta o^{'}(\mu)}{Ro(\mu) - \mu p} + \frac{\theta o(\mu)(Ro^{'}(\mu) - p)}{[Ro(\mu) - \mu p]^2} \\
&= o^{'}(\mu) - \frac{\theta\delta^{'}(\mu)p}{[R(1-\delta(\mu)) - p]^2}. \qquad \text{(B.5)}
\end{aligned}$$

Then from Lemma 11, we can show that $\partial\lambda(\mu,\, p)/\partial\mu < 0$ for $\mu \geq \mu^*$. By noting that $\mu_{an}(p)$ is the solution of $\partial\lambda(\mu,\, p)/\partial\mu = 0$, then $\mu_{an}(p) < \mu^*$. By Lemma 11 we know $o^{''}(\mu_{an}) < 0$. Then,

$$\begin{aligned}
\left.\frac{\partial^2\lambda(\mu,\, p)}{\partial\mu^2}\right|_{\mu=\mu_{an}} &= o^{''}(\mu_{an}) - \frac{\theta p\delta^{''}(\mu_{an})}{[R(1-\delta(\mu_{an})) - p]^2} - \frac{2R\theta(\delta^{'}(\mu_{an}))^2 p}{[R(1-\delta(\mu_{an})) - p]^3} \\
&= o^{''}(\mu_{an}) + \frac{\theta p(\delta^{'}(\mu_{an}))^2}{(1-\delta(\mu_{an}))[R(1-\delta(\mu_{an})) - p]^2} - \frac{2R\theta(\delta^{'}(\mu_{an}))^2 p}{[R(1-\delta(\mu_{an})) - p]^3} \\
&\quad - \frac{\theta p[\delta^{''}(\mu_{an})(1-\delta(\mu_{an})) + (\delta^{'}(\mu_{an}))^2]}{(1-\delta(\mu_{an}))[R(1-\delta(\mu_{an})) - p]^2} \\
&= o^{''}(\mu_{an}) - \frac{\theta p[\delta^{''}(\mu_{an})(1-\delta(\mu_{an})) + (\delta^{'}(\mu_{an}))^2]}{(1-\delta(\mu_{an}))[R(1-\delta(\mu_{an})) - p]^2} \\
&\quad - \frac{\theta p(\delta^{'}(\mu_{an}))^2(R(1-\delta(\mu_{an})) + p)}{(1-\delta(\mu_{an}))[R(1-\delta(\mu_{an})) - p]^3} \\
&< 0.
\end{aligned}$$

Thus, $\lambda(\mu,\, p)$ is quasi-concave in μ. Taking derivative of $\lambda_T(\mu,\, p)$ with respect to μ

and using (B.1),

$$\begin{aligned}
\frac{\partial \lambda_T(\mu,p)}{\partial \mu} &= 1-\frac{\theta R\delta^{'}(\mu)}{[R(1-\delta(\mu))-p]^2}, &\text{(B.6)}\\
\frac{\partial^2 \lambda_T(\mu,p)}{\partial \mu^2} &= -\frac{\theta R\delta^{''}(\mu)}{[R(1-\delta(\mu))-p]^2}-\frac{2\theta R^2(\delta^{'}(\mu))^2}{[R(1-\delta(\mu))-p]^3}\\
&= \frac{-R\theta}{(1-\delta(\mu))}\left[\frac{(\delta^{'}(\mu))^2(R(1-\delta(\mu))+p)}{[R(1-\delta(\mu))-p]^3}+\frac{\delta^{''}(\mu)(1-\delta(\mu))+(\delta^{'}(\mu))^2}{[R(1-\delta(\mu))-p]^2}\right]\\
&< 0. &\text{(B.7)}
\end{aligned}$$

Thus $\lambda_T(\mu,p)$ is concave in μ and the corresponding optimal service rate $\mu_{at}(p)$ as stated in (3.15) can be obtained by solving the FOC $d\lambda_T(\mu,p)/d\mu = 0$.

By noting that $o^{'}(\mu) = 1-\delta(\mu)-\mu\delta^{'}(\mu)$, (B.6) can be rewritten as

$$\begin{aligned}
\frac{\partial \lambda_T(\mu,p)}{\partial \mu} &= 1-\frac{\theta}{\mu[R(1-\delta(\mu))-p]}+\frac{\theta(Ro^{'}(\mu)-p)}{\mu[R(1-\delta(\mu))-p]^2}\\
&= \frac{\lambda_T(\mu,p)}{\mu}+\frac{\theta(Ro^{'}(\mu)-p)}{\mu[R(1-\delta(\mu))-p]^2}.
\end{aligned}$$

Obviously, the maximum total effective arrival rate should be positive; that is, $\lambda_T(\mu_{at}(p),p) > 0$. Because $\mu_{at}(p)$ is the solution of $\partial\lambda_T(\mu,p)/\partial\mu = 0$, we have $Ro^{'}(\mu_{at})-p < 0$. Substituting (3.15) into (B.5), we have

$$\left.\frac{\partial \lambda(\mu,\, p)}{\partial \mu}\right|_{\mu=\mu_{at}} = \frac{Ro^{'}(\mu_{at})-p}{R} < 0,$$

which implies that $\mu_{at}(p) > \mu_{an}(p)$.

Moreover, using (3.13),

$$\frac{\partial \rho_T(\lambda(\mu,\, p),\mu)}{\partial \mu} = \frac{\theta(Ro^{'}(\mu)-p)}{(Ro(\mu)-p\mu)^2}. \tag{B.8}$$

Recall from Lemma 11 that $o^{'}(\mu) \leq 0$ for $\mu \geq \mu^*$. Then $\mu_{ut}(p) < \mu^*$. Applying Lemma 11 again, $o^{''}(\mu_{ut}) < 0$. Thus,

$$\left.\frac{\partial \rho_T(\lambda(\mu, p), \mu)}{\partial \mu}\right|_{\mu=\mu_{ut}} = \frac{\theta R o^{''}(\mu_{ut})}{(Ro(\mu_{ut}) - p\mu_{ut})^2} < 0,$$

which shows that $\rho_T(\lambda(\mu, p), \mu)$ is quasi-concave in μ. Recall that $Ro^{'}(\mu_{at}) - p < 0$. Then

$$\left.\frac{\partial \rho_T(\lambda(\mu, p), \mu)}{\partial \mu}\right|_{\mu=\mu_{at}} = \frac{\theta(Ro^{'}(\mu_{at}) - p)}{(Ro(\mu_{at}) - p\mu_{at})^2} < 0.$$

Therefore, $\mu_{ut}(p) < \mu_{at}(p)$. In addition, from (B.8), we have $Ro^{'}(\mu_{ut}) - p = 0$. Substituting it into (B.5),

$$\begin{aligned}
\left.\frac{\partial \lambda(\mu, p)}{\partial \mu}\right|_{\mu=\mu_{ut}} &= \frac{p}{R}\left(1 - \frac{\theta R \delta^{'}(\mu_{ut})}{[R(1 - \delta(\mu_{ut})) - p]^2}\right) \\
&= \frac{p}{R}\left.\frac{\partial \lambda_T(\mu, p)}{\partial \mu}\right|_{\mu=\mu_{ut}} > 0.
\end{aligned}$$

The last inequality is because of $\mu_{ut}(p) < \mu_{at}(p)$. Finally, based on (3.14), we have

$$\begin{aligned}
\frac{\partial \rho_N(\lambda(\mu, p), \mu)}{\partial \mu} &= -\delta^{'}(\mu)\left(1 - \frac{\theta}{Ro(\mu) - p\mu}\right) + \frac{\theta(1 - \delta(\mu))(Ro^{'}(\mu) - p)}{(Ro(\mu) - p\mu)^2} \\
&= \frac{-\delta^{'}(\mu)\rho_N(\lambda(\mu, p), \mu)}{1 - \delta(\mu)} + \frac{\theta(1 - \delta(\mu))(Ro^{'}(\mu) - p)}{(Ro(\mu) - p\mu)^2}.
\end{aligned}$$

In equilibrium, the utilization rate for new patients must be positive. By noting that $\mu_{un}(p)$ is the solution of $\partial \rho_N(\lambda(\mu, p), \mu)/\partial \mu = 0$, we obtain $Ro^{'}(\mu_{un}) - p > 0$. And because $o^{'}(\mu) \leq 0$ for $\mu \geq \mu^*$, we have $\mu_{un} < \mu^*$. Again from Lemma 11, we have $o^{''}(\mu_{un}) < 0$. Furthermore, from $\partial\rho_N(\lambda(\mu_{un}, p), \mu_{un})/\partial\mu = 0$, we have

$$1 - \frac{\theta}{Ro(\mu_{un}) - p\mu_{un}} = \frac{\theta(1-\delta(\mu_{un}))(Ro^{'}(\mu_{un}) - p)}{\delta^{'}(\mu_{un})(Ro(\mu_{un}) - p\mu_{un})^2}.$$

Then,

$$\begin{aligned}
\left.\frac{\partial\rho_N(\lambda(\mu, p), \mu)}{\partial\mu}\right|_{\mu=\mu_{un}} &= -\delta^{''}(\mu_{un})\left(1 - \frac{\theta}{Ro(\mu_{un}) - p\mu_{un}}\right) - \frac{\theta\delta^{'}(\mu_{un})(Ro^{'}(\mu_{un}) - p)}{(Ro(\mu_{un}) - p\mu_{un})^2} \\
&\quad - \frac{\theta\delta^{'}(\mu_{un})(Ro^{'}(\mu_{un}) - p)}{(Ro(\mu_{un}) - p\mu_{un})^2} + \frac{\theta(1-\delta(\mu_{un}))Ro^{''}(\mu_{un})}{(Ro(\mu_{un}) - p\mu_{un})^2} \\
&\quad - \frac{2\theta(1-\delta(\mu_{un}))(Ro^{'}(\mu_{un}) - p)^2}{(Ro(\mu_{un}) - p\mu_{un})^3} \\
&= -\frac{\theta(Ro^{'}(\mu_{un}) - p)}{\delta^{'}(\mu_{un})(Ro(\mu_{un}) - p\mu_{un})^2}[\delta^{''}(\mu_{un})(1-\delta(\mu_{un})) + \delta^{'2}(\mu_{un})] \\
&\quad - \frac{\theta\delta^{'}(\mu_{un})(Ro^{'}(\mu_{un}) - p)}{(Ro(\mu_{un}) - p\mu_{un})^2} + \frac{\theta(1-\delta(\mu_{un}))Ro^{''}(\mu_{un})}{(Ro(\mu_{un}) - p\mu_{un})^2} \\
&\quad - \frac{2\theta(1-\delta(\mu_{un}))(Ro^{'}(\mu_{un}) - p)^2}{(Ro(\mu_{un}) - p\mu_{un})^3} < 0.
\end{aligned}$$

Thus, $\rho_N(\lambda(\mu, p), \mu)$ is quasi-concave in μ. From (B.8), we can know that $Ro^{'}(\mu_{ut}) - p = 0$. Then

$$\left.\frac{\partial\rho_N(\lambda(\mu, p), \mu)}{\partial\mu}\right|_{\mu=\mu_{ut}} = \frac{-\delta^{'}(\mu_{ut})\rho_N(\lambda(\mu_{ut}, p), \mu_{ut})}{1-\delta(\mu_{ut})} < 0.$$

Therefore, $\mu_{un} < \mu_{ut}$. ∎

Proof of Corollary 12. With a slight abuse of notion, we interchangeable use

$\mu_{\mathcal{H}}(p)$ and $\mu_{\mathcal{H}}$. From (3.16), we can easily know that

$$\frac{d\Pi(\mu)}{d\mu} = p\frac{\partial \lambda_T(\mu, p)}{\partial \mu}; \quad \frac{d^2\Pi(\mu)}{d\mu^2} = p\frac{\partial^2 \lambda_T(\mu, p)}{\partial \mu^2}.$$

Then, utilizing (3.15) we have

$$\left.\frac{\partial^2 \Pi(\mu)}{\partial \mu \partial p}\right|_{\mu=\mu_{\mathcal{H}}} = 1 - \frac{\theta R\delta^{'}(\mu_{\mathcal{H}})}{[R(1-\delta(\mu_{\mathcal{H}})) - p]^2} - \frac{2p\theta R\delta^{'}(\mu_{\mathcal{H}})}{[R(1-\delta(\mu_{\mathcal{H}})) - p]^3} = -\frac{2p\theta R\delta^{'}(\mu_{\mathcal{H}})}{[R(1-\delta(\mu_{\mathcal{H}})) - p]^3} < 0.$$

Let $\triangle = \delta^{''}(\mu_{\mathcal{H}})(1-\delta(\mu_{\mathcal{H}})) + (\delta^{'}(\mu_{\mathcal{H}}))^2$. By adopting the implicit function theory, utilizing (B.7) we have

$$\frac{d\mu_{\mathcal{H}}(p)}{dp} = -\left.\frac{\frac{\partial^2\Pi(\mu)}{\partial\mu\partial p}}{\frac{d^2\Pi(\mu)}{d\mu^2}}\right|_{\mu=\mu_{\mathcal{H}}} = \frac{-2\delta^{'}(\mu_{\mathcal{H}})(1-\delta(\mu_{\mathcal{H}}))}{(\delta^{'}(\mu_{\mathcal{H}}))^2(R(1-\delta(\mu_{\mathcal{H}})) + p) + \triangle[R(1-\delta(\mu_{\mathcal{H}})) - p]} < 0. \quad \text{(B.9)}$$

Therefore, $\mu_{\mathcal{H}}(p)$ is decreasing in p.

Next, plugging $\mu = \mu^*$ and $p = \bar{p} = R(1-\delta(\mu^*)) - \frac{\sqrt{R\theta o(\mu^*)}}{\mu^*}$ into (B.6), we have

$$\left.\frac{d\Pi(\mu)}{d\mu}\right|_{\mu=\mu^*} = \bar{p} - \frac{\bar{p}\theta R\delta^{'}(\mu^*)}{[R(1-\delta(\mu^*)) - \bar{p}]^2} = \bar{p}\mu^*\left[\frac{1-\delta(\mu^*) - \delta^{'}(\mu^*)\mu^*}{o(\mu^*)}\right] = 0.$$

Therefore, $\mu_{\mathcal{H}}(\bar{p}) = \mu^*$. Recall that $\mu_{\mathcal{H}}(p)$ is decreasing in p. Hence, $\mu_{\mathcal{H}}(p) > \mu^*$ iff $p < \bar{p}$. Plugging $p = \bar{p}$ and $\mu = \mu^*$ into (3.11), we can easily obtain that $\lambda(\mu^*, \bar{p}) = o(\mu^*) - \sqrt{\theta o(\mu^*)/R}$.

As $\delta(\mu)$ is increasing in μ while $\mu_{\mathcal{H}}(p)$ is decreasing in p, $\delta(\mu_{\mathcal{H}}(p))$ is decreasing in p. Substituting $\mu_{\mathcal{H}}(p)$ into (3.12) and taking the derivative over p yields

$$\begin{aligned}\frac{d\lambda_T(\mu_{\mathcal{H}}(p), p)}{dp} &= \frac{d\mu_{\mathcal{H}}(p)}{dp}\left(1 - \frac{\theta R\delta^{'}(\mu_{\mathcal{H}}(p))}{[R(1-\delta(\mu_{\mathcal{H}}(p))) - p]^2}\right) - \frac{\theta}{[R(1-\delta(\mu_{\mathcal{H}}(p))) - p]^2} \\ &= -\frac{\theta}{[R(1-\delta(\mu_{\mathcal{H}}(p))) - p]^2} < 0.\end{aligned}$$

Also, plugging $\lambda(\mu_{\mathcal{H}},\, p)$ into (3.6), we get

$$TW(\lambda(\mu_{\mathcal{H}},\, p),\, \mu_{\mathcal{H}}) \;=\; \frac{R}{\theta} - \frac{p}{\theta(1-\delta(\mu_{\mathcal{H}}))}.$$

Then,

$$\begin{aligned}
\frac{dTW(\lambda(\mu_{\mathcal{H}},\, p),\, \mu_{\mathcal{H}})}{dp} &= -\frac{p\delta^{'}(\mu_{\mathcal{H}})}{\theta(1-\delta(\mu_{\mathcal{H}}))^2}\frac{d\mu_{\mathcal{H}}(p)}{dp} - \frac{1}{\theta(1-\delta(\mu_{\mathcal{H}}))} \\
&= \frac{1}{\theta(1-\delta(\mu_{\mathcal{H}}))}\left[-\frac{p\delta^{'}(\mu_{\mathcal{H}})}{1-\delta(\mu_{\mathcal{H}})}\frac{d\mu_{\mathcal{H}}(p)}{dp} - 1\right].
\end{aligned}$$

Utilizing (B.9), we can show that

$$-\frac{p\delta^{'}(\mu_{\mathcal{H}})}{1-\delta(\mu_{\mathcal{H}})}\frac{d\mu_{\mathcal{H}}(p)}{dp} - 1 \;=\; -\frac{((\delta^{'}(\mu_{\mathcal{H}}))^2+\triangle)(R(1-\delta(\mu_{\mathcal{H}}))-p)}{(\delta^{'}(\mu_{\mathcal{H}}))^2(R(1-\delta(\mu_{\mathcal{H}}))+p)+\triangle[R(1-\delta(\mu_{\mathcal{H}}))-p]} < 0. \quad \text{(B.10)}$$

Thus, $TW(\lambda(\mu_{\mathcal{H}},\, p),\, \mu_{\mathcal{H}})$ is decreasing in p.

Substituting $\mu_{\mathcal{H}}$ into (3.11) and utilizing (3.15), we have

$$\begin{aligned}
\frac{d\lambda(\mu_{\mathcal{H}},\, p)}{dp} &= \frac{\partial\lambda(\mu_{\mathcal{H}},\, p)}{\partial\mu}\frac{d\mu_{\mathcal{H}}(p)}{dp} + \frac{\partial\lambda(\mu_{\mathcal{H}},\, p)}{\partial p} \\
&= \left[o^{'}(\mu_{\mathcal{H}}) - \frac{\theta p\delta^{'}(\mu_{\mathcal{H}})}{[R(1-\delta(\mu_{\mathcal{H}}))-p]^2}\right]\frac{d\mu_{\mathcal{H}}(p)}{dp} - \frac{\theta(1-\delta(\mu_{\mathcal{H}}))}{[R(1-\delta(\mu_{\mathcal{H}}))-p]^2} \\
&= \left[o^{'}(\mu_{\mathcal{H}}) - \frac{p}{R}\right]\frac{d\mu_{\mathcal{H}}(p)}{dp} - \frac{1-\delta(\mu_{\mathcal{H}})}{R\delta^{'}(\mu_{\mathcal{H}})} \\
&= o^{'}(\mu_{\mathcal{H}})\frac{d\mu_{\mathcal{H}}(p)}{dp} + \frac{(1-\delta(\mu_{\mathcal{H}}))}{R\delta^{'}(\mu_{\mathcal{H}})}\left[-\frac{\delta^{'}(\mu_{\mathcal{H}})p}{1-\delta(\mu_{\mathcal{H}})}\frac{d\mu_{\mathcal{H}}(p)}{dp} - 1\right].
\end{aligned}$$

Recall that $\mu_{\mathcal{H}}(p)$ is decreasing in p and when $p \geq \bar{p}$, $\mu_{\mathcal{H}}(p) \leq \mu^*$. Then, Lemma 11 and (B.10) imply that $d\lambda(\mu_{\mathcal{H}},\, p)/dp < 0$ for $p \geq \bar{p}$. ∎

Proof of Proposition 11. From (3.16) and (B.6), and utilizing (3.15), we have

$$\frac{do(\mu_{\mathcal{H}}(p))}{dp} = o^{'}(\mu_{\mathcal{H}}(p))\frac{d\mu_{\mathcal{H}}(p)}{dp},$$

$$\frac{d\Pi(\mu_{\mathcal{H}}(p))}{dp} = \mu_{\mathcal{H}} - \frac{\theta R(1-\delta(\mu_{\mathcal{H}}))}{[R(1-\delta(\mu_{\mathcal{H}}))-p]^2} = \mu_{\mathcal{H}} - \frac{1-\delta(\mu_{\mathcal{H}})}{\delta^{'}(\mu_{\mathcal{H}})} = -\frac{o^{'}(\mu_{\mathcal{H}})}{\delta^{'}(\mu_{\mathcal{H}})}.$$

Recall that $\mu_{\mathcal{H}}(p)$ is decreasing in p and $\mu_{\mathcal{H}}(\bar{p}) = \mu^*$. As $o(\mu)$ is quasi-concave in μ and attains its maximum at μ^* as shown in Lemma 11, $o^{'}(\mu_{\mathcal{H}}(p)) < 0$ when $p < \bar{p}$ and $o^{'}(\mu_{\mathcal{H}}(p)) > 0$ when $p > \bar{p}$. Therefore, both $do(\mu_{\mathcal{H}}(p))/dp$ and $d\Pi(\mu_{\mathcal{H}}(p))/dp$ are positive when $p < \bar{p}$ and they are both negative when $p > \bar{p}$. Therefore, $o(\mu_{\mathcal{H}}(p))$ and $\Pi(\mu_{\mathcal{H}}(p))$ are both quasi-concave in p and achieve their maximum at $p = \bar{p}$. According to Corollary 12, $\lambda(\mu^*, \bar{p}) < \lambda_b$. ∎

Proof of Proposition 12. Utilizing (3.11), (3.12) and (3.16), $SW(p)$ can be reexpressed as

$$SW(p) = [p + \beta(1-\delta(\mu_{\mathcal{H}}(p)))]\left[\mu_{\mathcal{H}}(p) - \frac{\theta}{R(1-\delta(\mu_{\mathcal{H}}(p)))-p}\right] - \beta\Lambda.$$

By noting that $\mu_{\mathcal{H}}(p) = \mu_{at}(p)$, $\mu_{\mathcal{H}}(p_{\mathcal{S}}) = \mu_{\mathcal{S}}$ and utilizing (3.15), the equilibrium outcome $p_{\mathcal{S}}$ must satisfy the following FOC:

$$\begin{aligned}
\left.\frac{dSW(p)}{dp}\right|_{p=p_{\mathcal{S}}} &= \left[1-\beta\delta^{'}(\mu_{\mathcal{S}})\left.\frac{d\mu_{\mathcal{H}}(p)}{dp}\right|_{p=p_{\mathcal{S}}}\right]\left[\mu_{\mathcal{S}}-\frac{\theta}{R(1-\delta(\mu_{\mathcal{S}}))-p_{\mathcal{S}}}\right] \\
&\quad -\frac{\theta[p_{\mathcal{S}}+\beta(1-\delta(\mu_{\mathcal{S}}))]}{[R(1-\delta(\mu_{\mathcal{S}}))-p_{\mathcal{S}}]^2} \\
&= \left[1-\beta\delta^{'}(\mu_{\mathcal{S}})\left.\frac{d\mu_{\mathcal{H}}(p)}{dp}\right|_{p=p_{\mathcal{S}}}\right]\left[\mu_{\mathcal{S}}-\frac{R(1-\delta(\mu_{\mathcal{S}}))-p_{\mathcal{S}}}{R\delta^{'}(\mu_{\mathcal{S}})}\right] \\
&\quad -\frac{p_{\mathcal{S}}+\beta(1-\delta(\mu_{\mathcal{S}}))}{R\delta^{'}(\mu_{\mathcal{S}})} \\
&= 0, \qquad \text{(B.11)}
\end{aligned}$$

or

$$\mu_{\mathcal{S}}=\frac{R(1-\delta(\mu_{\mathcal{S}}))-p_{\mathcal{S}}}{R\delta^{'}(\mu_{\mathcal{S}})}+\frac{p_{\mathcal{S}}+\beta(1-\delta(\mu_{\mathcal{S}}))}{R\delta^{'}(\mu_{\mathcal{S}})[1-\beta\delta^{'}(\mu_{\mathcal{S}})\frac{d\mu_{\mathcal{H}}(p)}{dp}|_{p=p_{\mathcal{S}}}]}. \qquad \text{(B.12)}$$

Therefore,

$$\begin{aligned}
o^{'}(\mu_{\mathcal{S}}) &= 1-\delta(\mu_{\mathcal{S}})-\mu_{\mathcal{S}}\delta^{'}(\mu_{\mathcal{S}}) \\
&= \frac{p_{\mathcal{S}}}{R}-\frac{p_{\mathcal{S}}+\beta(1-\delta(\mu_{\mathcal{S}}))}{R[1-\beta\delta^{'}(\mu_{\mathcal{S}})\frac{d\mu_{\mathcal{H}}(p)}{dp}|_{p=p_{\mathcal{S}}}]} \\
&= \frac{\beta(1-\delta(\mu_{\mathcal{S}}))(-\frac{p_{\mathcal{S}}\delta^{'}(\mu_{\mathcal{S}})}{1-\delta(\mu_{\mathcal{S}})}\frac{d\mu_{\mathcal{H}}(p)}{dp}|_{p=p_{\mathcal{S}}}-1)}{R[1-\beta\delta^{'}(\mu_{\mathcal{S}})\frac{d\mu_{\mathcal{H}}(p)}{dp}|_{p=p_{\mathcal{S}}}]}.
\end{aligned}$$

Because $\mu_{\mathcal{S}}(p)$ is decreasing in p as stated in Corollary 12, the denominator of the right-hand side of the above equation is positive. Then utilizing (B.10), we have $o^{'}(\mu_{\mathcal{S}}) < 0$. As $o(\mu)$ is quasi-concave and maximized at μ^* (Lemma 11), we have $\mu_{\mathcal{S}} > \mu^*$. Since $\mu_{\mathcal{H}}(p_{\mathcal{S}}) = \mu_{\mathcal{S}}$, the part 1 of Corollary 12 implies that $p_{\mathcal{S}} < \bar{p}$.

Plugging $\lambda(\mu_{\mathcal{S}},\, p_{\mathcal{S}})$ as stated in (3.11) into (B.4), we have

$$\left.\frac{dSW(\lambda,\, \mu_{\mathcal{S}})}{d\lambda}\right|_{\lambda=\lambda(\mu_{\mathcal{S}},\, p_{\mathcal{S}})} = R - \frac{o(\mu_{\mathcal{S}})[R(1-\delta(\mu_{\mathcal{S}})) - p_{\mathcal{S}}]^2}{\theta(1-\delta(\mu_{\mathcal{S}}))^2} + \beta.$$

By utilizing (3.15) and (B.12) and noting that $o(\mu) = \mu(1-\delta(\mu))$, the above equation can be simplified as

$$\begin{aligned}\left.\frac{dSW(\lambda,\, \mu_{\mathcal{S}})}{d\lambda}\right|_{\lambda=\lambda(\mu_{\mathcal{S}},\, p_{\mathcal{S}})} &= R - \frac{R\delta^{'}(\mu_{\mathcal{S}})\mu_{\mathcal{S}}}{1-\delta(\mu_{\mathcal{S}})} + \beta. \\ &= \frac{p_{\mathcal{S}}}{1-\delta(\mu_{\mathcal{S}})} - \frac{p_{\mathcal{S}} + \beta(1-\delta(\mu_{\mathcal{S}}))}{(1-\delta(\mu_{\mathcal{S}}))[1-\beta\delta^{'}(\mu_{\mathcal{S}})\frac{d\mu_{\mathcal{H}}(p)}{dp}\big|_{p=p_{\mathcal{S}}}]} + \beta.\end{aligned}$$

As $\frac{d\mu_{\mathcal{H}}(p)}{dp}\big|_{p=p_{\mathcal{S}}} < 0$ (Corollary 12), $1 - \beta\delta^{'}(\mu_{\mathcal{S}})\frac{d\mu_{\mathcal{H}}(p)}{dp}\big|_{p=p_{\mathcal{S}}} > 1$. Therefore,

$$\left.\frac{dSW(\lambda,\, \mu_{\mathcal{S}})}{d\lambda}\right|_{\lambda=\lambda(\mu_{\mathcal{S}},\, p_{\mathcal{S}})} > 0.$$

Next, from (B.3) we can further derive

$$\frac{\partial^2 SW(\lambda,\, \mu)}{\partial\lambda\partial\mu} = \frac{\theta o^{'}(\mu)}{[o(\mu)-\lambda]^2} + \frac{2\lambda\theta o^{'}(\mu)}{[o(\mu)-\lambda]^3}.$$

As $o(\mu)$ is quasi-concave in μ and achieves its maximum at $\mu = \mu^*$, $\frac{\partial^2 SW(\lambda,\, \mu)}{\partial\lambda\partial\mu} < 0$ if $\mu > \mu^*$. Therefore, $\frac{\partial SW(\lambda,\, \mu)}{\partial\lambda}$ is decreasing in μ when $\mu > \mu^*$. Recall that $\mu_{\mathcal{S}} > \mu^*$. Thus,

$$\left.\frac{dSW(\lambda,\, \mu^*)}{d\lambda}\right|_{\lambda=\lambda(\mu_{\mathcal{S}},\, p_{\mathcal{S}})} > \left.\frac{dSW(\lambda,\, \mu_{\mathcal{S}})}{d\lambda}\right|_{\lambda=\lambda(\mu_{\mathcal{S}},\, p_{\mathcal{S}})} > 0.$$

Since $SW(\lambda,\, \mu^*)$ is concave in λ (see the proof of Proposition 8) and achieves its optimum at $\lambda = \lambda_b$, $\lambda(\mu_{\mathcal{S}},\, p_{\mathcal{S}}) < \lambda_b$.

Below we prove that $\lambda(\mu^*, \bar{p}) < \lambda(\mu_{\mathcal{S}},\, p_{\mathcal{S}})$ by contradiction. Suppose that

$\lambda(\mu^*, \bar{p}) \geq \lambda(\mu_{\mathcal{S}}, p_{\mathcal{S}})$. The part 1 of Corollary 12 shows that $\mu_{\mathcal{H}}(\bar{p}) = \mu^*$. Thus,

$$\lambda(\mu_{\mathcal{H}}(\bar{p}), \bar{p}) = \lambda(\mu^*, \bar{p}) \geq \lambda(\mu_{\mathcal{H}}(p_{\mathcal{S}}), p_{\mathcal{S}}) = \lambda(\mu_{\mathcal{S}}, p_{\mathcal{S}}).$$

However, Proposition 11 shows that $\Pi(\mu_{\mathcal{H}}(p_{\mathcal{S}})) < \Pi(\mu_{\mathcal{H}}(\bar{p}))$. Then,

$$\begin{aligned} SW(p_{\mathcal{S}}) &= \Pi(\mu_{\mathcal{H}}(p_{\mathcal{S}})) - \beta\left(\Lambda - \lambda(\mu_{\mathcal{H}}(p_{\mathcal{S}}), p_{\mathcal{S}})\right) \\ &< \Pi(\mu_{\mathcal{H}}(\bar{p})) - \beta\left(\Lambda - \lambda(\mu_{\mathcal{H}}(\bar{p}), \bar{p})\right) = SW(\bar{p}), \end{aligned}$$

which contradicts the definition of $p_{\mathcal{S}}$. Therefore, we must have $\lambda(\mu^*, \bar{p}) < \lambda(\mu_{\mathcal{S}}, p_{\mathcal{S}})$. ∎

Proof of Proposition 13. For notational convenience, we use $\Pi(\mu|p)$ to represent the HCP's profit when the price p is charged. Through direct comparison, we can know that $p^* < \bar{p}$. According to the first statement of Corollary 12, $\mu_{\mathcal{H}}(p^*) > \mu^*$. As $\Pi(\mu|p)$ is quasi-concave in μ, $\Pi(\mu|p^*)$ is increasing in μ for $\mu \leq \mu^*$.

We next show that given $\alpha > 1 - \lambda_T(\mu^*, p^*)/\bar{\lambda}_T$, $\hat{\mu} = \mu^*$ and the socially optimal price p^*, the best response of the HCP is to choose μ^*. If $\hat{\mu} = \mu^*$, then $\mathbf{1}_{\mu > \mu^*} = 0$ for $\mu < \mu^*$, and $\mathbf{1}_{\mu > \mu^*} = 1$ for $\mu > \mu^*$. Thus, when $\mu < \mu^*$,

$$\Pi_{\mathcal{R}}(\mu|p^*) = \Pi(\mu|p^*) < \Pi(\mu^*|p^*) = \Pi_{\mathcal{R}}(\mu^*|p^*);$$

while when $\mu > \mu^*$,

$$\begin{aligned} \Pi_{\mathcal{R}}(\mu|p^*) &= (1-\alpha)\Pi(\mu|p^*) \\ &\leq (1-\alpha)\Pi(\mu_{\mathcal{H}}(p^*)|p^*) = (1-\alpha)\bar{\lambda}_T p^* \\ &< \lambda_T\left(\mu^*, p^*\right) p^* = \Pi_{\mathcal{R}}(\mu^*), \end{aligned}$$

where the last inequality is due to that $\alpha > 1 - \lambda_T(\mu^*, p^*)/\bar{\lambda}_T$. Therefore, when $\alpha > 1 - \lambda_T(\mu^*, p^*)/\bar{\lambda}_T$, $\mu = \mu^*$ is the HCP's best response service rate and the health care system is socially optimal. ∎

Proof of Proposition 15. According to Proposition 11, we can easily know that when $\lambda(\mu^*, \bar{p}) < \Lambda$, the partial market coverage scenario arises and the HCP chooses $\mu_{\mathcal{P}} = \mu^*$ and $p_{\mathcal{P}} = \bar{p}$. And $\lambda(\mu_{\mathcal{P}}, p_{\mathcal{P}}) < \lambda_b$.

When $\lambda(\mu^*, \bar{p}) \geq \Lambda$, the solutions under the partial market coverage are no longer feasible. In this case, all the patients can receive treatment. Here, the HCP's optimization problem becomes

$$\max_{\mu, p} \Pi(\mu, p) = \Lambda n(\mu) p, \quad s.t. \;\; \lambda(\mu, p) \geq \Lambda.$$

Then, the optimal solutions should satisfy $\lambda(\mu, p) = \Lambda$. Otherwise, if $\lambda(\mu, p) > \Lambda$, the HCP can make more profit by slightly increasing the price without violating the constraint. As stated in (3.11), letting $\lambda(\mu, p) = \Lambda$, we get $p = R(1 - \delta(\mu)) - \theta(1 - \delta(\mu))/(o(\mu) - \Lambda)$. Then the HCP's optimization problem can be rewritten as

$$\max_{\mu} \Pi(\mu) = \Lambda n(\mu) \left(R(1 - \delta(\mu)) - \frac{\theta(1 - \delta(\mu))}{o(\mu) - \Lambda} \right) = \Lambda \left(R - \frac{\theta}{o(\mu) - \Lambda} \right).$$

Then,

$$\frac{d\Pi(\mu)}{d\mu} = \frac{\Lambda \theta o'(\mu)}{(o(\mu) - \Lambda)^2} = 0 \;\Rightarrow\; o'(\mu) = 0.$$

Based on Lemma 11, we can show that $\mu_{\mathcal{P}} = \mu^*$. Consequently, $p_{\mathcal{P}} = R(1 - \delta(\mu^*)) - \theta(1 - \delta(\mu^*))/(o(\mu^*) - \Lambda)$. Finally, compared with the social optimality as shown in Proposition 14, we can easily know that the health care system can achieve the benchmark social optimality iff $\Lambda \leq \lambda(\mu^*, \bar{p})$. ∎

Proof of Corollary 13. Because $\lambda(\mu, p)$ is quasi-concave in μ (see Lemma 9), for a given p, there exist at most two points, denoted by μ_1 and μ_2, respectively such that $\lambda(\mu, p) = \Lambda$. Without loss of generality, assume that $\mu_1 < \mu_2$. By noting that $\mu_{an}(p)$ maximizes $\lambda(\mu, p)$, we have $\mu_1 < \mu_{an}(p) < \mu_2$. Note that $\lambda(\mu_2, p) = \Lambda$.

According to the implicit function theorem, we have

$$\frac{d\mu_2}{dp} = -\frac{\frac{\partial\lambda(\mu_2,p)}{\partial p}}{\frac{\partial\lambda(\mu_2,p)}{\partial \mu}}. \tag{B.13}$$

Because $\lambda(\mu,\, p)$ is quasi-concave in μ and $\mu_{an}(p) < \mu_2$, $\partial\lambda(\mu_2,\, p)/\partial\mu < 0$. Taking the derivative of $\lambda(\mu_2,\, p)$ as stated in (3.11) with respect to p, we have

$$\frac{\partial\lambda(\mu_2,\, p)}{\partial p} = -\frac{\theta(1-\delta(\mu_2))}{(R(1-\delta(\mu_2))-p)^2} < 0. \tag{B.14}$$

Then (B.13) implies that $d\mu_2/dp < 0$; that is μ_2 is decreasing in p. According to Proposition 16, $\mu_{\mathcal{F}}(p)$ is also decreasing in p.

Then from (3.12), we have

$$\frac{d\lambda_T(\mu_{\mathcal{F}},\, p)}{dp} = \frac{\Lambda\delta^{'}(\mu_{\mathcal{F}})}{(1-\delta(\mu_{\mathcal{F}}))^2}\frac{d\mu_{\mathcal{F}}(p)}{dp} < 0.$$

And from (3.5) and (3.6), we have

$$\frac{dW(\Lambda,\, \mu_{\mathcal{F}})}{dp} = \frac{\partial W(\Lambda,\, \mu_{\mathcal{F}})}{\partial\mu}\frac{d\mu_{\mathcal{F}}(p)}{dp} \quad\text{and}\quad \frac{dTW(\Lambda,\mu_{\mathcal{F}})}{dp} = \frac{\partial TW(\Lambda,\mu_{\mathcal{F}})}{\partial\mu}\frac{d\mu_{\mathcal{F}}(p)}{dp}.$$

Recall that both $TW(\Lambda,\mu)$ and $W(\Lambda,\, \mu)$ are quasi-convex in μ, and $\mu_{\mathcal{TW}}$ and $\mu_{\mathcal{W}}$ are the ones minimizing $TW(\Lambda,\mu)$ and $W(\Lambda,\, \mu)$, respectively (see Proposition 6). Let $p = p_j$ be the one satisfying $\mu_{\mathcal{F}}(p) = \mu_j$, $j \in \{\mathcal{TW},\, \mathcal{W}\}$. Since $\mu_{\mathcal{F}}(p)$ is decreasing in p, $\mu_{\mathcal{F}}(p) > \mu_j$ for $p < p_j$. Therefore, when $p < p_{\mathcal{TW}}$, $\partial TW(\Lambda,\mu_{\mathcal{F}}(p))/\partial\mu > 0$ and thus $dTW(\Lambda,\mu_{\mathcal{F}}(p))/dp < 0$; when $p > p_{\mathcal{TW}}$, $\partial TW(\Lambda,\mu_{\mathcal{F}})/\partial\mu < 0$ and thus $dTW(\Lambda,\mu_{\mathcal{F}})/dp > 0$. Hence, $TW(\Lambda,\mu_{\mathcal{F}})$ is quasi-convex in p and minimized at $p = p_{\mathcal{TW}}$. Similarly, we can show that $W(\Lambda,\mu_{\mathcal{F}})$ is also quasi-convex in p and minimized at $p = p_{\mathcal{W}}$. Last, since $\mu_{\mathcal{TW}} < \mu_{\mathcal{W}}$ (see Proposition 6) and $\mu_{\mathcal{F}}(p)$ is decreasing in p, $p_{\mathcal{TW}} > p_{\mathcal{W}}$. ∎

Proof of Proposition 17. We first consider the scenario $\Lambda \leq \lambda(\mu^*, \bar{p})$. Corollary 12 shows that $\lambda(\mu^*, \bar{p}) = o(\mu^*) - \sqrt{\theta o(\mu^*)/R}$. According to Proposition 14, when $\Lambda \leq \lambda(\mu^*, \bar{p})$, $\lambda_b = \Lambda$ and $\mu_b = \mu^*$. Let

$$p_0 = R(1 - \delta(\mu^*)) - \frac{\theta(1 - \delta(\mu^*))}{o(\mu^*) - \Lambda}.$$

Plugging μ^* and p_0 into (3.11) yields that $\lambda(\mu^*,\, p_0) = \Lambda$. Below we prove that when $\Lambda \leq \lambda(\mu^*, \bar{p})$, then $p_0 \in \Theta$ and $\mu_{\mathcal{F}}(p_0) = \mu^*$. Therefore, by charging the price p_0, the social planner can make the health care system achieve social optimality. Then it is natural that $p_{\mathcal{SF}} = p_0$ and $\mu_{\mathcal{SF}} = \mu_{\mathcal{F}}(p_{\mathcal{SF}}) = \mu^*$.

By noting that $\lambda(\mu^*, \bar{p}) = o(\mu^*) - \sqrt{\theta o(\mu^*)/R} \geq \Lambda$ and $o(\mu^*) = \mu^*(1 - \delta(\mu^*))$,

$$p_0 \;=\; R(1 - \delta(\mu^*)) - \frac{\theta(1 - \delta(\mu^*))}{o(\mu^*) - \Lambda} \geq R(1 - \delta(\mu^*)) - \frac{\sqrt{R\theta}(1 - \delta(\mu^*))}{\sqrt{o(\mu^*)}} = \bar{p}. \tag{B.15}$$

Then based on the part 1 of Corollary 12, we have $\mu_{\mathcal{H}}(p_0) \leq \mu_{\mathcal{H}}(\bar{p}) = \mu^*$. Next, according to Lemma 9 and Proposition 10, we have $\mu_{\mathcal{H}}(p) = \mu_{at}(p) > \mu_{an}(p)$. Hence, $\mu_{an}(p_0) < \mu_{\mathcal{H}}(p_0) \leq \mu^*$. Recall that $\lambda(\mu,\, p)$ is quasi-concave in μ and achieves its maximum at $\mu_{an}(p)$. Then

$$\lambda(\mu_{an}(p_0),\, p_0) > \lambda(\mu_{\mathcal{H}}(p_0),\, p_0) \geq \lambda(\mu^*,\, p_0) = \Lambda.$$

Thus, $p_0 \in \Theta$ and $\mu = \mu^*$ is the maximum service rate that satisfies $\lambda(\mu,\, p_0) \geq \Lambda$. According to Proposition 16, $\mu_{\mathcal{F}}(p_0) = \mu^*$. Therefore, $p_{\mathcal{SF}} = p_0$ and $\mu_{\mathcal{SF}} = \mu_{\mathcal{F}}(p_{\mathcal{SF}}) = \mu^*$.

Now consider the case that $\Lambda > \lambda(\mu^*, \bar{p}) = o(\mu^*) - \sqrt{\theta o(\mu^*)/R}$. We first prove that when $\Lambda > \lambda(\mu^*, \bar{p})$, $\mu_{\mathcal{H}}(p) > \mu^*$ and $p < \bar{p}$ for all $p \in \Theta$. Suppose that there exists a $p^{'} \in \Theta$ such that $p^{'} \geq \bar{p}$. Then according to part 3 of Corollary 12,

$$\lambda(\mu_{\mathcal{H}}(p^{'}), p^{'}) \leq \lambda(\mu_{\mathcal{H}}(\bar{p}), \bar{p}) = \lambda(\mu^*, \bar{p}) < \Lambda,$$

which implies that $p^{'} \notin \Theta$. This leads to a contradiction of our assumption that $p^{'} \in \Theta$. Therefore, $p < \bar{p}$ for all $p \in \Theta$. Then part 1 of Corollary 12 implies that if $p \in \Theta$, then $\mu_{\mathcal{H}}(p) > \mu_{\mathcal{H}}(\bar{p}) = \mu^*$.

We next show that $\mu_{\mathcal{F}}(p) > \mu^*$ for all $p \in \Theta$. According to the definition of Θ, if $p \in \Theta$, then $\lambda(\mu_{\mathcal{H}}(p),\, p) \geq \Lambda$. Because $\mu_{\mathcal{F}}(p)$ is the maximum service rate that satisfies $\lambda(\mu,\, p) = \Lambda$, according to Lemma 9, $\mu_{\mathcal{F}}(p) \geq \mu_{\mathcal{H}}(p)$ for $p \in \Theta$. Recall that $\mu_{\mathcal{H}}(p) > \mu^*$ for all $p \in \Theta$, we have $\mu_{\mathcal{F}}(p) \geq \mu_{\mathcal{H}}(p) > \mu^*$ for $p \in \Theta$. As $p_{\mathcal{SF}} \in \Theta$, $\mu_{\mathcal{SF}} = \mu_{\mathcal{F}}(p_{\mathcal{SF}}) > \mu^*$. Recall that under the full market coverage scenario, the social planner shall set a price so that $\mu_{\mathcal{F}}(p)$ is as close to μ^* as possible. Because $\mu_{\mathcal{F}}(p)$ is decreasing in p and $\mu_{\mathcal{F}}(p) > \mu^*$ for all $p \in \Theta$, the social planner shall choose the largest price in the set Θ. ∎

Proof of Proposition 18. Because λ_m represents the smallest equilibrium effective arrival rate of new patients under the partial coverage scenario, when $\Lambda < \lambda_m$, the optimal solution under the partial market coverage is no longer feasible, therefore, the health care system will end up covering the market fully.

When $\Lambda > \lambda_m$, let $\hat{\lambda} = \max_p \{\lambda(\mu_{\mathcal{H}}(p),\, p)\}$ represent the largest effective arrival rate of new patients under the partial market coverage scenario. If $\hat{\lambda} < \Lambda$, then Θ is empty and the health care system ends up with the partial market coverage. If $\hat{\lambda} \geq \Lambda$, both the full market coverage scenario and the partial market coverage scenario may occur. In equilibrium, the social welfare under the partial market coverage scenario is

$$SW(p_{\mathcal{S}}) = \Pi(\mu_{\mathcal{S}}) - \beta \left(\Lambda - \lambda(\mu_{\mathcal{S}},\, p_{\mathcal{S}})\right),$$

and that under the full market coverage scenario is

$$SW(p_{\mathcal{SF}}) = \Lambda \left[R - \frac{\theta}{o(\mu_{\mathcal{SF}}) - \Lambda} \right].$$

Below we are going to show that $SW(p_{\mathcal{S}}) > SW(p_{\mathcal{SF}})$ for all $\beta \geq 0$.

Note that when $\beta = 0$, the social welfare under the partial market coverage $SW(p)$ is equal to $\Pi(\mu_{\mathcal{H}}(p))$, which indicates that the objective of the HCP is aligned with that of the social planner. Then it is natural that the performance of the health care system will end up being socially optimal. The second statement of Proposition 17 shows that when $\Lambda > \lambda(\mu^*, \bar{p})$, $\mu(p_{\mathcal{SF}}) > \mu^*$. Thus, the equilibrium outcome under the full market coverage scenario is not socially optimal. Therefore, when $\beta = 0$, $SW(p_{\mathcal{S}}) > SW(p_{\mathcal{SF}})$.

When $\beta > 0$, with a slight abuse of notion, we assume that under the partial market coverage scenario, $\lambda_m = \lambda(\mu_{\mathcal{S}}, p_{\mathcal{S}})$. Below we prove by contradiction that $\lim_{\beta\to+\infty} \lambda(\mu_{\mathcal{S}}, p_{\mathcal{S}}) = \Lambda$. Suppose that the above is not true. Then there exists a $\varepsilon > 0$ such that for any $\hat{\beta} > 0$, there exists a $\beta_0 > \hat{\beta}$ such that $\Lambda - \lambda(\mu_{\mathcal{S}}, p_{\mathcal{S}}) > \varepsilon$. Then, according to Proposition 11, at $\beta = \beta_0$,

$$SW(p_{\mathcal{S}}) = \Pi(\mu_{\mathcal{S}}) - \beta_0 (\Lambda - \lambda(\mu_{\mathcal{S}}, p_{\mathcal{S}})) < \Pi(\mu_{\mathcal{H}}(\bar{p})) - \hat{\beta}\varepsilon.$$

Because $\lambda_m < \Lambda \leq \hat{\lambda}$, due to the continuity of $\lambda(\mu_{\mathcal{H}}(p), p)$, there exists a p_1 such that $\lambda(\mu_{\mathcal{H}}(p_1), p_1) = \Lambda$ and thus, $SW(p_1) = \Pi(\mu_{\mathcal{H}}(p_1)) > 0$. When $\hat{\beta}$ is large enough, we have

$$SW(p_{\mathcal{S}}) < \Pi(\mu_{\mathcal{H}}(\bar{p})) - \hat{\beta}\varepsilon < SW(p_1).$$

Due to the continuity of $SW(p)$ in p, we can find a p_2 sufficiently close to p_1 such that $\lambda(\mu_{\mathcal{H}}(p_2), p_2) < \Lambda$ and

$$SW(p_{\mathcal{S}}) < SW(p_2).$$

This contradicts the definition of $p_{\mathcal{S}}$. Hence, $\lim_{\beta\to+\infty} \lambda(\mu_{\mathcal{S}}, p_{\mathcal{S}}) = \Lambda$. Thus, when β approaches infinity, the partial market coverage scenario is reduced to the full market coverage scenario. Therefore, $\lim_{\beta\to+\infty} p_{\mathcal{S}} = p_{\mathcal{SF}}$, $\lim_{\beta\to+\infty} \mu_{\mathcal{S}} = \mu_{\mathcal{SF}}$ and $\lim_{\beta\to+\infty} SW(p_{\mathcal{S}}) =$

$SW(p_{\mathcal{SF}})$.

Finally, we are going to show that $SW(p_{\mathcal{S}})$ is decreasing in β. Recall that $SW(p)$ is not unimodal in p. For a fixed β, let $X(\beta)$ represent the set of prices that satisfy the FOC of $SW(p)$ and let $p_{\mathcal{S}}(\beta)$ represent the price that maximizes $SW(p)$. Then $p_{\mathcal{S}}(\beta) \in X(\beta)$. Suppose $\beta_1 < \beta_2$. Denote $p_{\mathcal{S}}(\beta_2) = p_d(\beta_2) \in X(\beta_2)$. It is worth noting that at $\beta = \beta_1$, $p_d(\beta_1)$ may not maximize $SW(p)$ anymore. However, based on the Envelop theory, we have

$$\frac{dSW(p_d(\beta))}{d\beta} = -(\Lambda - \lambda(\mu_{\mathcal{H}}(p_d(\beta)), p_d(\beta))) < 0.$$

Thus, $SW(p_d(\beta_2)) < SW(p_d(\beta_1))$. By noting that $p_{\mathcal{S}}(\beta)$ maximizes $SW(p)$, $SW(p_{\mathcal{S}}(\beta_2)) = SW(p_d(\beta_2)) < SW(p_d(\beta_1)) \leq SW(p_{\mathcal{S}}(\beta_1))$. Hence, $SW(p)$ is decreasing in β. Consequently, $SW(p_{\mathcal{S}}) > SW(p_{\mathcal{SF}})$ for any $\beta > 0$. ∎

Proof of Proposition 19. According to Proposition 17, when $\Lambda < \lambda(\mu^*, \bar{p})$, $p_{\mathcal{SF}} = R(1-\delta(\mu^*)) - \theta(1-\delta(\mu^*))/(o(\mu^*) - \Lambda)$ and $\mu_{\mathcal{SF}} = \mu^*$. Substituting them into (3.11) yields $\lambda(p_{\mathcal{SF}}, \mu_{\mathcal{SF}}) = \Lambda$. Then Proposition 14 implies that when $\Lambda < \lambda(\mu^*, \bar{p})$, the health care system under the full market coverage is already socially optimal. Furthermore, according to Propositions 12 and 17, when $\Lambda \geq \lambda(\mu^*, \bar{p})$, the equilibrium service rate is larger than the socially optimal one. Therefore, the health care system is not socially optimal. ∎

References

Abelson, R. 2013. Hospitals question medicare rules on readmissions. *The New York Times* http://www.nytimes.com/2013/03/30/business/hospitals-question-fairness-of-new-medicare-rules.html?pagewanted=all&_r=0 accessed 29 March 2013.

Alizamir, S., F. de Véricourt, P. Sun. 2013. Diagnostic accuracy under congestion. *Management Science* **59**(1) 157-171.

An, J., S.-H. Cho, C. Tang. 2014. Aggregating smallholder farmers for alleviating poverty in emerging economies: models and analyses. Forthcoming in *Production and Operations Management*.

Anand, K. S., M. F. Pac, S. Veeraraghavan. 2011. Quality-speed conundrum: Tradeoff in customer-intensive services. *Management Science* **57**(1) 40-56.

Armantier, O. and O. Richard. 2003. Exchanges of cost information in the airline industry. *RAND Journal of Economics* **34**(3) 461-477.

Bagnoli, M., and T. Bergstrom. 2005. Log-Concave Probability and Its Applications. *Economic Theory* **26**(2) 445-469.

Berwick, D. M., A. D. Hackbarth. 2012. Eliminating waste in US health care. *The Journal of the American Medical Association* **307**(14) 1513-1516.

Bylund, P. 2014. What Sweden can teach us about ObamaCare. *The Wall Street Journal* http://online.wsj.com/news/articles/SB10001424052702303603904579491971221879580 accessed 17 April 2014.

Burke, P. J. 1956. Output of a queuing system. *Operations Research* **4**(6) 699-704.

Carter, C. A., D. MacLaren. 1994. Alternative oligopolistic structures in international commodity markets: price or quantity competition? Working Paper, Department of Agriculture, The University of Melbourne.

Chan, C. W., V. F. Farias, N. Bambos, G. Escobar. 2012. Optimizing icu discharge decisions with patient readmissions. *Operations Research* **60**(6) 1323-1342.

Chan, C. W., G. B. Yom-Tov, G. Escobar. 2014. When to use speedup: An examination of service systems with returns. *Operations Research* **62**(2) 462-482.

Chen, Y. J., J. G. Shanthikumar, Z. J. Shen. 2013a. Training, production, and channel separation in ITC's e-Choupal network. *Production and Operations Management* **22**(2) 348-364.

Chen, Y. J., J. G. Shanthikumar, Z. J. Shen. 2013b. Incentive for peer-to-peer information sharing in Avaaj Otalo. Working paper, University of California at Berkeley.

Chen, Y. J., C. S. Tang. 2013. The economic value of market information for farmers in developing economies. Working paper, UCLA Anderson School.

Chollet, D., A. Barrett, T. Lake. 2011. Reducing hospital readmissions in New York State: a simulation analysis of alternative payment incentives. Princeton, NJ: Mathematica Policy Research.

Conis, E. 2009. Impact of a state-level all-payer system. *Health Policy Monitor* Available at http://www.hpm.org/survey/us/a14/5.

Creswell, J., R. Abelson. 2013. New laws and rising costs create a surge of supersizing Hospitals. *The New York Times.* http://www.nytimes.com/2013/08/13/business/bigger-hospitals-may-lead-to-bigger-bills-for-patients.html?pagewanted=all&module=Search&mabReward=relbias%3Ar%2C%7B%222%22%3A%22RI%3A13%22%7D accessed 12 August 2013.

Dawande, M., S. Gavirneni, M. Mehrotra, V. Mookerjee. 2013. Efficient Distribution of Water Between Head-Reach and Tail-End Farms in Developing Countries. *Manufacturing & Service Operations Management* **15**(2) 221-238.

Debo, L. G., L. B. Toktay, L. N. Van Wassenhove. 2008. Queuing for expert services.

Management Science **54**(8) 1497-1512.

Deodhar, S. Y., I. M. Sheldon. 1996. Estimation of imperfect competition in food marketing: A dynamic analysis of the German banana market. *Journal of Food Distribution and Research* **27**(3) 1-10.

Dong, F., T. Marsh, K. Stiergert. 2006. State trading enterprises in a differentiated environment: The case of global malting barley markets. *American Journal of Agricultural Economics* **88**(1) 90-103.

Dimakou, S. 2013. Waiting time distributions and national targets for elective surgery in UK: theoretical modelling and duration analysis. Doctoral thesis, Department of Economics, City University London, UK.

de Vricourt, F., Y. Zhou. 2005. Managing response time in a call-routing problem with service failure. *Operations Research* **53**(6) 968-981.

Edwards, C. 2009. Agricultural subsidies. http://www.downsizinggovernment.org/agriculture/subsidies accessed June 2009.

Fafchamps, M., B. Minten. 2011. The impact of SMS-based agricultural information on India. Working paper, Department of Economics, University of Oxford.

Feldman, J., N. Liu, H. Topaloglu, S. Ziya. 2014. Appointment scheduling under patient preference and no-show behavior. Forthcoming in *Operations Research.*

Fethke, C. C., I. M. Smith, N. Johnson. 1986. "Risk" factors affecting readmission of the elderly into the health care system. *Medical Care* **24**(5) 429-437.

Fetter, R. B., J. D. Thompson. 1996. Patients' waiting time and doctors' idle time in the outpatient setting. *Health Service Research* **1**(1) 66-90.

Friedman, B., J. Basu. 2004. The rate and cost of hospital readmissions for preventable conditions. *Medical Care Research and Review* **61**(2) 225-240.

Fudenberg, D., J. Tirole. 1991. *Game Theory.* Cambridge, Massachussetts: The

MIT Press.

Fuloria, P. C., S. A. Zenios. 2001. Outcomes-adjusted reimbursement in a health-care delivery system. *Management Science* **47**(6) 735-751.

Gal-Or, E. 1985. Information sharing in oligopoly. *Econometrica* **53**(2) 329-343.

Gal-Or, E. 1986. Information transmission – Cournot and Bertrand equilibria. *Review of Economic Studies* **53**(1) 85-92.

Glynn, N., K. Bennett, B. Silke. 2011. Emergency medical readmission: long-term trends and impact on mortality. *Clinics in Medicine* **11**(2) 114-118.

Guo, P., R. Lindsey, Q. Qian. 2013. Efficiency of subsidy schemes in reducing waiting times for public health-care services. Working paper, Sauder School of Business, University of British Columbia, Vancouver.

Hasija, S., E. Pinker, R. A. Shumsky. 2009. Work expands to fill the time available: Capacity estimation and staffing under parkinson's law. *Manufacturing Service Operations Management* **12**(1) 1-18.

Hassin, R., M. Haviv. 2003. To queue or not to queue: equilibrium behavior in queueing systems. Kluwer Academic Publishers, Norwell, MA.

Hassin R., S. Mendel. 2008. Scheduling arrivals to queues: A single-server model with no-shows. *Management Science* **54**(3) 565-572.

Heggestad, T., S. E. Lilleeng. 2003. Measuring readmissions: focus on the time factor *International Journal for Quality in Health Care* **15**(2) 147-154.

Hopp, W. J., S. M. R. Iravani, G. Y. Yuen. 2007. Operations systems with discretionary task completion. *Management Science* **53**(1) 61-77.

Hsia, R. Y., S. M. Asch, R. E. Weiss, et al. 2011. Hospital determinants of emergency department left without being seen rates. *Annals of Emergency Medicine* **58**(1) 24-32.

Hueth, B., P. Marcoul. 2006. Information sharing and oligopoly in agricultural markets: The role of the cooperative bargaining association. *American Journal of Agricultural Economics* **88**(4) 866- 881.

Humphreys, A. 2013. Canada must offer private options along with universal health-care to combat long wait times: report. http://news.nationalpost.com/2013/10/28/canada-must-offer-private-options-along-with-universal-health-care-to-combat-long-wait-times-report/ accessed 28 October 2013.

Ikegami, N., G. F. Anderson. 2012. In Japan, all-payer rate setting under tight government control has proved to be an effective approach to containing costs. *Health Affairs* **31**(5) 1049-1056.

Jansen, J. 2010. Strategic information disclosure and competition for an imperfectly protected innovation. *The Journal of Industrial Economics* **58**(2) 349-372.

Jensen, R. 2007. The digital provide: Information (technology), market performance, and welfare in the south Indian fisheries sector. *The Quarterly Journal of Economics* **122**(3) 879-924.

Jiang, H., Z. Pang, S. Savin. 2012. Performance-based contracts for outpatient medical services. *Manufacturing Service Operations Management* **14**(4) 654-669.

Johnson, M., P. Hazell. 2002. Cutting hunger in Africa through smallholder-led agricultural growth. An IFPRI technical paper in support of USAIDs Agricultural Initiative to cut hunger in Africa (AICHA). Available at: http://pdf.usaid.gov/pdf_docs/Pnacr575.pdf accessed 23 August 2002.

Jones, P., L. Chalmers, S. Wells, et al. 2012. Implementing performance improvement in New Zealand emergency departments: the six hour time target policy national research project protocol. *BMC Health Service Research* **12**(45).

KC, D. S., C. Terwiesch. 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management*

14(1) 50-65.

Klein, E. 2012. So it's the prices, but what do we do? http://www.pnhp.org/news/2012/march/so-its-the-prices-but-what-do-we-do accessed 2 March 2012.

Kim, S. H., C. W. Chan, M. Olivares, G. Escobar. 2013. Icu admission control: An empirical study of capacity allocation and its implication on patient outcomes. Working paper, Columbia Business School.

Kim, S., I. Horowitz, K. K. Young, T. A. Buckley. 1999. Analysis of capacity management of the intensive care unit in a hospital. *European Journal of Operational Research* **115**(1) 36-46.

Kociol, R. D., R. D. Lopes, R. Clare, L. Thomas, R. H. Mehta, P. Kaul, et al. 2012. International variation in and factors associated with hospital readmission after myocardial infarction. *The journal of the American medical association* **307**(1) 66-74.

Kostami, V., S. Rajagopalan. 2014. Speed quality tradeoff in a dynamic model. *Manufacturing & Service Operations Management* **16**(1) 104-118.

Lee, DK. K., S. A. Zenios. 2012. An evidence-based incentive system for Medicares End-Stage Renal Disease program. *Management Science* **58**(6) 1092-1105.

Li, L. 2002. Information sharing in a supply chain with horizontal competition. *Management Science* **48** (9) 1196-1212.

Liu, N., S. Ziya, V. G. Kulkarni. 2010. Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing & Service Operations Management* **12**(2) 347-364.

Liu, N., S. Ziya. 2014. Panel Size and Overbooking Decisions for Appointment-based Services under Patient No-shows. Forthcoming in *Production and Operations Management*.

Luo, J., V. Kulkarni, S. Ziya. 2012. Appointment scheduling under patient no-shows and service interruptions. *Manufacturing & Service Operations Management* **14**(4) 670-684.

Mills, A. F., N. Argon, S. Ziya. 2013. Resource-based patient prioritization in mass-casualty incidents. *Manufacturing & Service Operations Management* **15**(3) 361-377.

Mittal, S., S. Gandhi, G. Tripathi. 2010. Socio-economic impact of mobile phones on Indian agriculture. Working paper 246, Indian Council for Research and International Economic Relations.

Morrow-Howell, N., E. K. Proctor. 1993. The use of logistic regression in social work research. *Journal of Social Service Research* **16**(1-2) 87-104.

Nieuwoudt, W. L. 1987. Allocation of beef permits and quotas. *South African Journal of Economics* **55** (3) 182-187.

Oostenbrink, J. B., F. F. Rutten. 2006. Cost assessment and price setting of inpatient care in The Netherlands. The DBC case-mix system. *Health Care Management Science* **9**(3) 287-294.

Patrick, J., M. Puterman. 2008. Reducing waiting time through operations research: Optimizing the use of surge capacity. *Healthcare Policy* **3**(3) 75-88.

Paç, M. F., S. Veeraraghavan. 2010. Strategic diagnosis and pricing in expert services. Working paper, Wharton School, University of Pennsylvania, Philadelphia.

Pierson, D. 2011. Amid China's economic transformation, many farmers struggle. http://articles.latimes.com/2011/jul/07/business/la-fi-china-farming-20110707 accessed 07 July 2013.

Polevoi, S. K., J. V. Quinn, N. R. Kramer. 2005. Factors associated with patients who leave without being seen. *Academic Emergency Medicine* **12**(3) 232-236.

Purdy, S., S. Paranjothy., A. Huntley et al. 2012. Interventions to reduce unplanned hospital admission: a series of systematic reviews. University of Bristal, UK.

Raith, M. 1996. A general model of information sharing in oligopoly. *Journal of Economic Theory* **71** (1) 260-288.

Reinhardt, U. E. 2006. The pricing of US hospital services: chaos behind a veil of secrecy. *Health Affairs* **25**(1) 57-69.

Roy, A. 2011. Hospital monopolies: the biggest driver of health costs that nobody talks about. *Forbes.* http://www.forbes.com/sites/theapothecary/2011/08/22/hospital-monopolies-the-biggest-driver-of-health-costs-that-nobody-talks-about/ accessed 22 August 2011.

Sanmartin, C., J-M. Berthelot, C. N. Mcintosh. 2007. Determinants of unacceptable waiting times for specialized services in Canada. *Healthcare Policy* **2**(3) 140-154.

Schwartz, M. 1999. Buyer power concerns and the Aetna-Prudential Merger. U.S. Department of Justice, Washington, DC.

Sharma, A. P. 2013. 34% marginal farmers in Punjab living below poverty line. http://timesofindia.indiatimes.com/india/34-marginal-farmers-in-Punjab-living-below-poverty-line/articleshow/26176039.cms accessed 22 November 2013.

Smith, V. 2012. Rich farmers don't need farm bill's Welfare. http://www.usnews.com/debate-club/should-congress-pass-the-farm-bill/rich-farmers-dont-need-farm-bills-welfare accessed 14 September 2012.

So, K. C., C. S. Tang. 2000. Modeling the impact of an outcome-oriented reimbursement policy on clinic, patients, and pharmaceutical firms. *Management Science* **46**(7) 875-892.

Sodhi, M., C. S. Tang. 2012. Research opportunities in supply chains with micro-entrepreneurs in developing economies. *Production and Operations Management* forthcoming.

Stavrunova, O., O, Yerokhin. 2011. An equilibrium model of waiting times for elective surgery in NSW public hospitals. *Economic Record* **87**(278) 1475-4932.

Street, A., J. O'Reilly, P. Ward, A. Mason. 2011. DRG-based hospital payment and efficiency: Theory, evidence, and challenges. In *Diagnosis-Related Groups in Europe - Moving Towards Transparency, Efficiency and Quality in Hospitals*, Busse R, Geissler, A, Quentin W, Wiley M (eds). Open University Press: Maidenhead; 93-114.

Tsang, E. 2014. 21,000 hospital patients wait three years for specialist care. http://www.scmp.com/news/hong-kong/article/1487664/21000-three-year-wait-specialist-care?page=all accessed 19 April 2014.

Thukral, N., C. Packham. 2014. Australian farmers hold back wheat sales, may miss the boat. http://www.reuters.com/article/2014/06/16/us-australia-wheat-farmers-idUSKBN0-ER0KC20140616 accessed 16 June 2014.

Tong, C., S. Rajagopalan. 2014. Pricing and operational performance in discretionary services. *Production and Operations Management* **23**(4) 689-703.

Varkevisser, M., S. A. van der Geest, F. T. Schut. 2012. Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands. *Journal of Health Economics* **31** 371-378.

Vives, X. 1988. Aggregation of information in large Cournot markets. *Econometrica* **56** (4) 851-876.

Wang, X., L. G. Debo, A. Scheller-Wolf, S. F. Smith. 2010. Design and analysis of diagnostic service centers. *Management Science* **56**(11) 1873-1890.

Wilke, J. R. 2004. How driving prices lower can violate antitrust statutes. *The Wall Street Journal.* http://online.wsj.com/news/articles/SB107515998826512191 accessed 27 January 2004.

Winnett, R. 2011. Scandal of NHS 'production line' as readmissions soar. http://www.telegraph.co.uk/health/healthnews/8983505/Scandal-of-NHS-production-line-as-readmissions-soar.html accessed 29 December 2011.

Xu, Y., A. Scheller-Wolf, K. Sycara. 2012. The benefit of introducing variability in quality based service domains. Working paper, Tepper School of Business, Carnegie Mellon University, Pittsburgh.

Yom-Tov, G. B., A. Mandelbaum. 2014. Erlang-R: A time-varying queue with reentrant customers, in support of healthcare staffing. *Manufacturing & Service Operations Management* **16**(2) 283-299.

Zhu, K. 2004. Information transparency of business-to-business electronic markets: A game-theoretic analysis. *Management Science* **50**(5) 670-685.