THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l' Université Toulouse 1 Capitole
Discipline : Sciences Economiques

Présentée et soutenue par

Zi LI

Le 6 Juin 2013

Titre :

Essay on Public Policy for Innovation and Quality

JURY

Emmanuelle Auriol, professeur, Université Toulouse 1
Sara Biancini, professeur, Université de Caen, (rapporteur)
Robert Gary-Bobo, professeur, Ecole Polytechnique, (rapporteur)
Guido Friebel, professeur, Goethe Universitât Frankfurt
Wilfried Sand-Zantman, professeur, Université Toulouse 1

Ecole doctorale : Toulouse School of Economics
Unité de recherche : IDEI - TSE
Directeur de Thèse : Emmanuelle Auriol

Contents

Acknowledgements

This dissertation would not have been possible without the support and help from many people. Those people to whom I owe a great deal of thanks and I would like to address my appreciation:

First and foremost, I want to gratefully thank my advisor Emmanuelle Auriol for many helpful discussions and encouragements. It is her wiseness, kindness and great patience the most important source, from which I gain my strength to pursue the career of the economics researcher. She does not only advise me on research methods and skills, but also set an outstanding example for me about keeping optimistic mood towards braving new ideas.

During the last several years spending in the Toulouse School of Economics, I was provided the chance to attend numerous courses, seminars and workshops. Step by step, I learned how to formally analyze issues from the real world, and to understand the underlying trade-off behind various economic models. It is here in Toulouse, I, who had been only good at taking exams, was reborn as a junior economics researcher. Therefore, I want to thank the faculty for their effort to establish such an excellent PhD program.

I would like to express my great appreciation to Sara Biancini and Robert Gary-Bobo, who generously invested their time in reading my thesis. Their reports encouraged me to publish my works and provided a lot of excellent suggestions to modify them. I am also very thankful to Wilfried Sand-Zantman and Guido Friebel for their participating in the jury. Moreover, the first chapter of the dissertation has been benefited from the helpful discussions with Helmuth Cremer, Jean-Marie Lozachmeur and David Bardey.

Thanks to Toulouse School of Economics for the financial supports. Thanks to Aude Schloesing, who is always accessible to me and helped me in solving many problems. Many thanks to my wife and good friends with whom I shared a very nice time in Toulouse.

Finally, I would like to thank my parents and other family members for staying in close contact. And this dissertation is in memory of my grandfather in heaven.

Chapter 1

Introduction

Generally, people enjoy better lives today than who did one hundred years ago. One among many reasons behind this fact is the high quality services and products, which we now have access to, could only exist in scientific fictions at the beginning of the last century. At least two things make this change happened. One is the constantly emerging scientific innovations. The other is the importance of quality emphasized through markets. As a matter of fact, numerous episodes in the history of economics theory have suggested the positive effect of innovations on economic growth and of quality on welfare of human beings. Regulators continuously attempted to design public policies to encourage innovations and improve quality provision in a vertical structure. On the downstream, where the technologies are taken as granted, regulators seek for efficiency through governance mechanisms which incentivize the supply of high quality goods or services in various markets. But no one can be certain that these policies have served their purposes, without looking into the detailed information about the regulated markets. Different mechanisms may lead to diverse outcomes. Regulators have also to determine the approaches by which they can maximize the positive effect of quality on efficiency. Chapter two and three of this dissertation will discuss the policies concerning the quality of health and food markets, as the picked examples from the downstream. On the upstream, where the update of technologies should be induced to an efficient level, granting the patent rights to innovators is now considered as the main policy to promote innovations. Regulators have ameliorated the patent systems gradually along with their efforts. However, Heller and Eisenberg (1998) suggested a tragedy of anti-commons that the whole patent systems are flawed, because the veto power of original patent holders is able to delay and decline the sequential findings. Should we end the patent systems once for all according to anti-commons? Note that this tragedy has been challenged by recent studies which suggested its impact to be limited. Chapter four will try to give a new explanation about why the debate is inconclusive.

The second chapter is a joint work with Lijun Zhao, a former PhD candidate of Toulouse School of Economics. We investigate what happens when high quality providers (physicians, hospitals) are allowed to patronize multiple health plans (multi-homing) and compare it to single-homing in a two-sided framework. As put by Bardey and Rochet (2010), health plans' competition has two-sided nature. By this they mean that "health insurance markets are characterized by indirect network externality between providers and policyholders." Specifically, health plans limit their enrollees' access to only their contracted providers (physicians, clinics and hospitals) and policyholders (potential patients) prefer access to a large number of providers to increase their chances of getting proper treatment. Similarly, providers may prefer to have access to a large pool of potential patients depending on the incentive of providers embodied in payment scheme (fixed salary or fee for service (FFS)).

In 1990s, single-homing staff-mode HMOs (Health Maintenance Organizations) faded away in the U.S.. With the staff model, physicians may only see HMO patients. Now a captive group model replaces the staff-model. Providers multi-home and continue to treat non-HMO patients. It is said that a driven force of this change would be the market demand for the broad choices of high quality providers. The non-HMO policyholders might be willing to accept relatively high premiums, when they can select high quality providers outside of the list of providers affiliated to HMOs. More recently, China's government also began to encourage high quality providers to treat the patients, who used to fully pay for the care by themselves, through enforcing the health plans of those patients to cover some of the expenses. The multi-homing of high quality providers is supposed to promote fairness and policyholders' welfare. However, there is not much formal analysis of this issue with two-sided settings.

Regarding the multi-homing in two-sided markets, the model of "competitive bottleneck" (Armstrong and Wright (2007) and Armstrong (2006)) is famous. In that model, the whole provider side view platforms as homogenous and joins in all of them when multi-homing is allowed. As the authors suggested, the platforms do not compete directly on the multi-homing side, instead competing indirectly by subsidizing the single-homing side to join in. In equilibrium, the single-homing (consumer or policyholder) side is treated favorably (indeed, its price is necessarily no higher than its cost), while the multi-homing (provider or seller) side has its entire surplus extracted by the platforms. This result reflects the reality under some circumstances, but may not for high quality goods and services. In fact, the "competitive bottleneck" is also a warning from platforms to providers: showing disloyalty by multi-homing will cost them. The suppliers with high quality should be expected to incur costs when joining in any platform, thus are not as easily captured as assumed by the "competitive bottleneck". For instance, one should never anticipate that the luxury brands, such as Louis Vuitton and Hermes, will humiliate themselves by giving a discount in Auchan or Carrefour, because patronizing the supermarkets together with low quality products will demean their fashionably leading status. Therefore, it is more appropriate to extend the framework of Bardey and Rochet (2010) to discuss the multi-homing and single-homing of high quality providers, in which the providers are different in their preferences of participating in health plans.

We consider two types of providers, H (high quality) and L (low quality), exerting different quality externality to policyholders. And we define the health plan's quality level as the percentage of H providers in this plan. The health plans under investigation are asymmetric that one plan is always with strictly larger provider size than the other. The first result in this chapter is that multi-homing of H providers yields less competitive intensity between health plans than single-homing does, thus both plans prefer multi-homing than single-homing. Second, allowing H providers working for multiple plans increases the size and quality of both plans. Third, contradicting to the "competitive bottleneck", the policyholders, who always single-home, may not be benefited by multi-homing of H providers, due to that the premiums paid by policyholders may be relatively high in average compared to the increase in externality of multi-homing. Forth, the efficiency of multi-homing depends on the payment scheme providers receiving. Multi-homing increases both size and quality externality enjoyed by all policyholders if they stay in their original plan. Under fixed salary scheme, providers do not worry about the number of patient they treat. There is no network externality on provider side, so this side should be internalized. The plans optimally adjust the prices for policyholders, according to the change on provider side, so that policyholders will stay in their original choice of health plan. As a result, multi-homing under a fixed salary scheme improves efficiency through increasing the quality and size externalities for every policyholder. However, under the scheme of FFS, provides are incentivized by the number of their potential patients, thus network externality also appears on the provider side. The plan with strictly larger size and quality externality is less motivated by multi-homing to offer a large patient pool to its enrolled providers than single-homing, because multi-homing lowers the competition intensity. Therefore, this plan subscribes fewer policyholders with multi-homing than single-homing. Some policyholders who would choose this larger health plan with single-homing, turn to the smaller one with multi-homing. These policyholders' loss of the size and quality externality may exceed the benefit from more quality and size externalities. Social welfare may be worse off by multi-homing under an FFS scheme.

In **the third chapter**, we are interested in optimal structure of regulations on food quality. Particular attentions are paid to two regulatory approaches: the traditional product inspection (PI) and the new process certification (PC). PI involves the sampling and testing of products and signals the quality accurately to the consumers. PC

focuses on verifiable production process control and provides relatively noisy signals. As Auriol and Schilizzi (2003) pointed out, the regulatory costs affects the public certification structure for food quality. Due to the relatively low monitoring cost of PC, this new method are gradually becoming more common. There is a noticeable amount of literature, which demonstrated that PC is more cost-effective than PI in numerous cases (see for instance, Unnevehr and Jensen (1996), Crutchfield et al., (1997) and Roberts et al., (1996)).

It is common to categorize quality into search, experience and credence attributes in economic literature. The quality of search goods, such as color and size, is easily detected before consumption. For experience goods, like taste and suitability, consumers are able to discern its intrinsic quality after consumption. Finally, for credence (or trust) goods, quality can never be known by consumers with certainty (see Nelson (1970)). And the last one is usually affiliated with fraud and quality problems, which can lead to socially costly inefficiencies (see Darby and Karni (1973)). The credence feature of food safety and nutrition is main focus in this paper, which follows the understanding of Caswell and Mojduszka (1996). A question that should be asked is whether new method PC may only substitute for, or be supplemented by PI in optimal regulation, as suggested by some surveys (see for instance, Unnevehr and Jensen (1998)). To date as I know, there is not much effort exerted to provide some formal answers to this question.

Similar to Auriol and Schilizzi (2003), we consider a symmetric Cournot-oligopoly structure. The difference is that the total supply of high quality food and number of firms with the high quality are random. In practice, there is no manufacture procedure could completely prohibit the low quality food to be produced. One should expect the incidents leading to low quality outputs to happen with positive probabilities. The result suggests that PI should never be combined with PC to be adopted by the regulator. PI impacts the outcome on three aspects. First, adopting PI with high frequency, the regulator can induce firms to using the production process, which prevents the incidents as much as possible. If this incentive effect on firms' strategy of PI is established, then PC, which monitors the safety (or quality) of firms' production processes, becomes unnecessary. Second, PI is pro-competitive. Knowing PI being adopted, all firms should expect that some of their rivals may recall their products due to their bad quality. They should also increase quantities and prepare to take over their competitors' market shares. With quantity competition of Cournot model, the total supply will be distorted to less than efficient levels. Thus, enhancing quantity is a positive factor for social welfare. Third, PI directly prevents some low quality food from achieving consumers' table. Therefore, it increases the credibility of public regulation. Because of its relatively large monitoring cost, one may expect that adopting PI by some small frequency along with PC would be optimal. However, the second and third effect of PI timing together may cause the social benefit from PI is convex in its frequency. That is, if the expenditure on PI is worthy for improving efficiency, it should be conducted as frequent as possible; otherwise, it should be stopped. Therefore, if PI's cost is small enough (less than some threshold), it prevails in the optimal regulatory scheme, where PC is not needed. On the contrary, if PI's cost exceeds that threshold, PI should be excluded from the public policies. Moreover, the threshold decreases with competitiveness in the food market. The reason behind this result is that the quantity distortion, which leads to the profitability of the firms, becomes insignificant when the competition in the food market is intensified. The pro-competitive effect of PI is then diluted. The advantage of PC becomes most apparent in a complete competitive situation.

In **chapter four**, we try to provide a novel justification for the anti-commons hypothesis of Heller and Eisenberg (1998), which is suggested to be over-stated by some recent studies. Specifically, anti-commons highlights the patent protection as a two-edged sword: it spurs scientific research by securing scientists the fruits of their labors, but it also gives patent holders potential power to restrict how others conduct research based on the protected knowledge. It clearly implies that new innovations and investments in research and development (R&D) activities could be hindered by patent rights. However, some empirical studies, such as Mowery and Ziedonis (2002), Sampat et al. (2003) and Murray and Stern (2007), suggested oppositely. They used the citation data of patented innovations within relatively long period and showed that the average citation rate for a scientific publication did not fall dramatically after the formal patent rights associated with that publication had been granted. Furthermore, evidence for patent rights lowering the R&D activities has not been convincing enough (see Refferty (2008)).

Many theoretical contributions like Scotchmer (1991), Green and Scotchmer (1995), Chang (1995), Scotchmer (1996), focus on sequential innovation, where the patentee and licensee work on sequential and different projects. With such a setting, anti-commons rarely appears because the patentee should sell his or her technology out before the patent rights expire. Bessen and Maskin (2009) consider the case when both patent holder and licensee research for the subsequent innovation based on the originally patented knowledge. The patent holder may be better off by exterminate the competition in a future market of new innovation than issuing license and allowing a rival to research. In this sense, anti-commons may take place. The negative effect of patent protection is enlarged when the patent holder blocked others' R&D program and then failed in his or her own program. The further studies is delayed, since patent holder him or herself has no new knowledge as the necessary input. But their support for anti-commons cannot explain why its effect is suggested to be insignificant by the studies mentioned above.

Keeping the assumption of two scientists (or innovators) competing in a common new innovation of Bessen and Maskin (2009), here a common value situation is introduced: the scientists' valuation of the new innovation relies on each others' perspectives. Moreover, the patent holder can signal his perspective, which is private information, through the investment in his own R&D program. Auriol and Laffont (1992) pointed out that full informational disclosure usually requires positive social cost in a common value model. As in the Spence (1974) education model, a highly productive worker may invest in wasteful education to avoid being considered as a less capable one. Following these ideas, this chapter presents an outcome with over-investment in R&D: the patentee tries to fool the licensee by concealing his relatively bad perspective and signaling the good news through over-investment in his R&D. It is profitable for the patentee when his high expectation of the new innovation can largely enhance the licensee's valuation, thus a high license fee can be expected. The over-investment is the positive social cost, even though the information has not been disclosed. And it is eventually paid by the licensee through a license contract.

Anti-commons may also take place, because the patentee would like to screen licensee's private information. Specifically, the license fee would be larger when the licensee expects a fruitful future than a barren one. A blockade might be introduced in the license menu proposed by the patentee. It incentivizes the licensee to claim a good expectation by blocking him if he reports the bad news. As the result, anti-commons and over-investment in R&D can ex ante co-exist when each scientist's perspective can change the common valuation substantially. Under this circumstance, the patentee is eager to disclose licensee's information and the licensee is easily confused by over-investment of the patentee. The result suggests that anti-commons may be hardly observable by using citation and R&D investment data of patents, because it could be masked by the co-existed another form of inefficiency: the over-investment in R&D.

Chapter 2

Multi-homing of High Quality Providers in Two-sided Health Market

2.1 Introduction

[1]The theory of two-sided markets has been developed to study market structures in which two groups of agents interact via platforms. The importance of network externality, which occurs when the benefit enjoyed by a member of one group depends on the size of the other group with whom they connect, has been stressed. Platform members may also care about the quality of the other side agents they are going to meet.[2] When agents of any side present different qualities, platforms not only differ in relative size but also in the mixture of agents, resulting in quality externality.

The health plan market[3] is two-sided, because health plans limit enrolled policyholders' (potential patients) access to only their contracted providers (physicians, clinics and hospitals, etc.).[4] To increase their chances of getting proper treatment, policyholders prefer access to a large number of medical care providers. Similarly, providers may prefer to have access to a large pool of potential patients. The health markets are sensitive

[1]It is a joint work with Lijun Zhao

[2]Internet dating sites and night clubs (Damiano and Li (2005)) are two examples. Users of dating sites care about their potential dates' looks and backgrounds and the wealthy go to night clubs which serve expensive drinks.

[3]The health plan market is prevailing. Gaynor (1999) indicates that in the U.S., 83 percent of physicians had a contract with at least one health plan in 1995, up from 61 percent in 1990 (Emmons and Simon, 1996).

[4]For example, in China, the public health plans limit their policyholders to seeking care in hospitals, which are on the lists of their chosen providers.

to providers' quality.[5] The providers' quality may be heterogeneous in several aspects: general practitioners and specialists, experienced physicians and recent graduates, multilingual providers and one language providers, etc.. These characteristics can be known *ex ante* by health plans with little cost and providers can be reimbursed accordingly. In seeking treatment, policyholders care not only about the size of the network but also the concentration of high quality providers.[6]

As Evans (2003) notes, "*most two-sided markets we observe in the real world appear to have several competing two-sided firms and at least one side appears to multi-home.*" Evidence of physicians holding multiple contracts with different health plans can be found all over the world.[7] Compared to the low quality providers, those of high quality are more easily to get multiple offers from health plans.[8] With the new reform of medical care system (2009 - 2011), China's government officially encourages the multi-homing in health plans of high quality providers. In China, the national public health care system was decentralized in 1990s. The local health plans are mainly built on mandatory contributions (premiums) from policyholders. They regulated their health services such that policyholders got treatments and they covered the expenses, only if the treatments are conducted by some providers within their administrative localizations. Therefore, the high quality providers had to single-home in the health plans of their districts. And they only treated patients from distant locations, who could afford their medical care without using the health plans. The new reform relaxed this previous regulation. It allows the patients to seek for high quality providers with reimbursements from their local plans, regardless the difference in localizations between the providers and themselves. Consequently, the high quality providers are now more flexible in choosing health plans to join in. The government believes that the new reform will increase policyholders' welfare by allowing policyholders a broader choice of high quality providers, especially when *high quality providers are scarce* and *other available providers are with relatively low quality.* Thus, multi-homing of high quality providers is supposed to promote fairness and efficiency. However, there is no formal analysis of this issue.

[5] See for instance, Ma and Burgess (1993)

[6] The quality of medical care is usually regarded as credence attributes in economics literature (see Emons (1997)). It is hard to be verified by policyholders on individual levels. Policyholders need quality signals from health plans. Or they may rely on information disclosed by other institutions, such as Consumer Assessment of Health Plans Study (CAHPS), which signal the quality of health plans. The health plans and the institutions can identify the quality of medical providers through verifying physicians' diplomas and carefully recording doctors' and hospitals' previous performances, etc..

[7] In France, almost all clinics accept all kinds of health insurances. In the U.S., single-homing staff-model HMOs (Health Maintenance Organizations) became inactive in the 1990s. Multi-homing group model HMOs took their place.

[8] For example, the experienced specialists often have multiple jobs in different clinics and hospitals. The hospitals, identified as high quality ones, usually have more contracts with different health insurances than those of low quality.

In this paper, we extend Bardey & Rochet's framework (2010) by introducing quality differentiation among providers that policyholders value. Specifically, there are two types of providers, H (high quality) and L (low quality), each exerting different levels of quality externality on policyholders. The two health plans, which are for profit platforms, compete in price on two sides: the provider side (side 1) and the policyholder side (side 2). Another characteristic of the health market is that policyholders are heterogenous and have private information about their propensity to use services. Therefore, they are assumed different in their likelihood of getting a disease, which will determine their choice for network size and quality.

Health plans may influence each provider through administrative controls. For a provider, joining a health plan is often associated with the cost of administrative controls and corresponding paperwork. Geographic localizations may also matter. As in staff-mode Health Maintenance Organizations (HMOs), the contracted doctors should only treat patients inside the HMOs' buildings. Moreover, the health plans are often seen as separate goods by physicians even without cost differences. When physicians make one period contract decision, staying in the same health plan may be more attractive than switching to other plans because of the relationships they have developed with their current patients. Therefore, providers are expected to have different preferences for participating in health plans, which also explains the existence of independent practice providers not affiliated with any health plans or affiliated with only one plan when multi-homing is possible. The model of this chapter also includes this feature: providers are assumed to locate on a Hotelling line with two health plans at each extreme and must incur transportation costs to join in any plan.

We focus on the situation when only some of H providers will multi-home, if multi-homing is allowed. We compare the multi-homing equilibrium to when single-homing of providers is mandatory. For the purposes of this study, we use the percentage of H providers in one plan to indicate this plan's *quality level*. We show that multi-homing leads to higher quality and larger size of both health plan than single-homing. Multi-homing also yields less competitive intensity between health plans than single-homing does, thus both plans prefer multi-homing than single-homing. If single-homing prevents some H providers from multi-homing, allowing them working for multiple plans broadens their choices of actions and thus, increases the surplus of their side. However, the policyholders' welfare may be undermined by multi-homing, because the premium would be very high. Furthermore, we define the *relative quality* of the low quality as the substitution rate of one L provider to one H in each policyholder's expected utility. Contrary to the intention of China's government of encouraging multi-homing of H providers, we show that the effect of multi-homing on policyholders' welfare will be limited by the *low relative quality* and *few* high quality providers available in average.

Thus, increasing health plans' *quality level* through multi-homing may not significantly improve the policyholders' welfare, if there is some improvement. It suggests that to improve fairness and efficiency (of policyholders), more efforts should be exerted to enhance the quality of L providers, which could eventually increase the population of H providers.

In this paper, two commonly used schemes of the payment on the provider side are considered: fixed salary, and fee-for-service (FFS). Under a fixed salary scheme, a physician is paid independently of the number of treatments he performs; while under FFS, physicians are reimbursed on a per-service basis.[9] We show that compared to single-homing, multi-homing under a fixed salary scheme improves efficiency. It is because multi-homing increases both size and quality externality enjoyed by policyholders if they stay in their original plan. And they will stay when providers receive fixed salaries. Under this payment scheme, providers do not worry about the number of patient they treat: there is no network externality on provider side. The plans internalize this side through adjusting the premiums for policyholders, so none of policyholders will change their original choice of health plan. Therefore, multi-homing increases the externality enjoyed by every policyholder.

However, when providers are compensated with FFS, they care about the number of their potential patients: network externality also appears on the provider side. Additionally, the equilibrium allows the coexistence of two asymmetric health plans. The plan serving high risk policyholders is always a plan with strictly larger aggregate externality than the other one. With the network effect on provider side, the better plan with strictly larger aggregate externality is less motivated by multi-homing to offer a large patient pool to its providers than single-homing, because multi-homing lowers the competition intensity. The subscribed number of policyholders of this better plan is less with multi-homing than single-homing. Some policyholders who would choose the better plan with single-homing, turn to the lower quality one with multi-homing. These policyholders' loss may exceed the benefit of having more H providers in both plans. So social welfare may be worse off by multi-homing under an FFS scheme.

The remainder of the paper is organized as follows. Section 2.2 is the literature review. In Section 2.3, we introduce our model. We determine the single-homing and multi-homing equilibrium when providers are remunerated via salaries in Section 2.4, followed

[9] There are some other payment schemes on provider side. For example, capitation is another commonly used payment rule, where a physician receives a fixed predetermined payment based on each policyholder assigned to him, whether or not that policyholder seeks care. When there is no heterogeneity among policyholders' demand for treatment, FFS and capitation are basically same since the only concern is the number of policyholders. When policyholders are heterogenous in their treatment demand, the physicians must share the risks with the health plans with capitation. however, in this paper, all providers and health plans are risk neutral.

by the comparison between these two. We then continue to the case in which providers are remunerated through a FFS in Section 2.5. Finally, we conclude in Section 2.6.

2.2 Related Literature

As stressed in Rochet and Tirole (2006), whenever price structures (from which side a platform generates most its profit) matter, the markets are two-sided. A large part of two-sided market literature focuses on price structures. For instance, Armstrong and Wright (2007) and Armstrong (2006) provide a model of "competitive bottlenecks" in which the whole provider side joins all platforms in equilibrium when multi-homing is allowed, because there is no transportation costs for providers to join in platforms. They show that in equilibrium, the single-homing (consumer) side is treated favorably (indeed, its price is necessarily no higher than its cost), while the multi-homing (provider) side has its entire surplus extracted. Caillaud and Jullien (2003), Rochet and Tirole (2003) also provide similar results: multi-homing on one side intensifies price competition on the other side as platforms use low prices in an attempt to steer end users on the latter side toward an exclusive relationship. However, this competitive bottleneck is contestable: platforms neither necessarily generate most revenues on the multi-homing side as Rasch (2007) shows, nor strengthen the competition on the other side as in this paper. Especially when the provider side presents horizontal differentiation over platforms, providers are no longer captive and some of them choose to single-home if they are allowed to multi-home. In this case, platforms cannot extract the whole providers' surplus with multi-homing. Instead, they may weaken the competition on the policyholder side to generate more profit.

To date, the discussion of platforms' favor over multi-homing or single-homing is not conclusive either. For instance, Armstrong and Wright(2007) suggests that when agents on one side of the market multi-home, platforms might offer exclusive contracts to them to prevent them from multi-homing. Such exclusive contracts can be inexpensive to offer since by tying up one side of the market, the platform attracts the other side, which reinforces the decision of that side to sign up exclusively. However, Caillaud and Jullien (2003) show that platforms have incentives to propose non-exclusive services, because with single-homing, the incumbent platform needs to set the prices very low to deter any competition. We follow the latter's comprehension.

Finally, as for welfare concerns, Rochet and Tirole (2003) demonstrate that without detailed information about the platforms' demand, one cannot not expect clear comparisons across governance mechanisms. Nevertheless, Caillaud and Jullien (2003) show that multi-homing may improve the efficiency through enhancing the aggregate externality, but may lead to inefficiency in market structure since some platform may not attract enough agents on the single-homing side. Following their steps, we derive clear results about the impact of multi-homing on efficiency conditional on the adopted payment scheme in a more complex environment, where agents may have heterogenous taste for platforms and present different qualities.

There is not much attention paid to quality levels of platforms. Viecens (2006) is an exception.[10] This paper allows for the two types of externality, the standard indirect network effect and the externality due to quality concerns, as we do. The author argues that the quality level of a platform as a reputation affect may keep providers of high quality from multi-homing. In our paper, this reputation impact of quality appears when providers are paid by FFS: health plans may undercut their enrollments of low quality physicians to provide those of high quality to a large patient pool. Moreover, this paper assumes that the two platforms are (*ex ante*) identical, i.e., there are no transportation costs on the provider side. It concludes that the equilibrium, where all high quality providers multi-home and others single-home, yields the highest quality externality and thus, the highest welfare. In contrast, we assume that the provider side presents horizontal differentiation over platforms. When there are only a few providers with incentive to multi-home and others prefer single-homing because of the transportation costs, the little increment of quality externality from multi-homing can be offset by the inefficient allocation of policyholders. Thus multi-homing may undermine welfare.

Our paper is close to Bardey and Rochet (2010), which characterizes the indirect network externality between the provider and policyholders sides. In practice, a health plan allowing access to a large number of providers attracts policyholders characterized by a higher risk than the average population. In a one-sided analysis, this risk segmentation would only be a disadvantage for this health plan because of the higher number of reimbursements generated. However, in their two-sided framework, such a health plan may realize the highest profit in equilibrium. The reason behind this result is that the indirect network externality allows this health plan to charge a higher premium to policyholders and negotiate a lower FFS rate with providers. However, this paper does not discuss the case of providers choosing to multi-home. We extend their model by adding a quality heterogeneity on providers. Our goal is to understand how quality and efficiency change with the shift of providers' constraint on participating in multiple health plans.

[10]Gabszewicz and Wauthy (2004) and Damiano and Li (2005) also consider the two-sided platforms with endogenous quality differentiation. However, the former defines the quality in one market as the size of the network in the other market, thus there is no quality externality. On the other hand, the latter excludes the indirect network externality.

2.3 The Model

We consider a health care market delivering treatment for any illness any policyholder could possibly get.

The model is based on Bardey and Rochet (2010). There are two health plans $i = \{A, B\}$ competing for two groups of agents: policyholders and providers. Policyholders are heterogeneous in the probability of needing medical care and are fully covered by health plans. Each policyholder could only get treatment from providers in the same health plan.[11] Additionally, we assume that there are two types of providers who exert different quality externality on policyholders. One type is high quality and labeled H and the other is low quality and labeled L. Providers choose whether to join one or two (if multi-homing is allowed) or none of the health plans. They get paid by the health plans, not directly from patients. We assume that health plans can obtain the information of a provider's type (H or L) without cost.[12]

2.3.1 Provider Side

Providers are uniformly located on the Hotelling line and providers' location is private information. Plan A is located at 0 and plan B at 1. The total number of H type is q, the remaining $1-q$ are L type. To be simple, the distribution of quality is independent of location on Hotelling line.[13] Providers are assumed to be risk neutral. Their reservation utility is normalized to zero. Because providers' qualities can be costlessly verified by plans, type dependent payment from health plans to providers can then be used.

Let T_i^j denote the expected payment from health plan i to a j provider, where $j = H, L$, $i = A, B$.[14] The utility of a j provider located at x and joining health plan A (or B respectively) is $T_A^j - tx$ (or $T_B^j - t(1-x)$ respectively). Here t is the marginal transportation cost.For simplicity, we adopt the linear transportation cost.

We define $\widetilde{x}_i^j$ as the location of marginal j ($j = H, L$) provider who is indifferent between joining in health plan i ($i = A, B$) and one of the other two strategies: 1) joining in none of plans and getting utility 0, 2) only joining in the other plan. The first being weakly

[11] To make things simple, we keep this restriction in the study of our model.

[12] For example, managed care plans have explicit standards for selecting providers.

[13] That is $\Pr(H|x) = q$ for any location x on the Hotelling line. Under some circumstances, it may be more appropriate to assume that providers choose their locations strategically. For example, high quality providers are more possible to work in the big cities with the large population and high average income, such as Beijing, Paris and London, than in small cities. However, besides the geographic localizations, health plans also differ in other aspects, which are valued by providers as suggested in the introduction. The interdependence between providers' tastes of health plans and their quality may not be strong.

[14] The payments can be stochastic when fee-for-service is adopted.

dominated by joining in plan i ($i = A, B$)for providers $\widetilde{x}_i^j$ implies that the following inequalities (referred as *individual rationality constraints*) must hold:

$$(IR_A^j) : T_A^j - t\widetilde{x}_A^j \geq 0 \text{ and } (IR_B^j) : T_B^j - t(1 - \widetilde{x}_B^j) \geq 0. \tag{2.1}$$

If multi-homing is allowed, the second strategy does not matter. (IR_i^j), $j = H, L$, $i = A, B$ are sufficient for all j providers with $x \leq \widetilde{x}_A^j$ joining in plan A and those with $x \geq \widetilde{x}_B^j$ joining in plan B. [15] However, if single-homing is mandatory, the second strategy will be dominated. Besides (IR_i^j), the plans also need to meet the *incentive constraints*:

$$(IC_A^j) : T_A^j - t\widetilde{x}_A^j \geq T_B^j - t(1 - \widetilde{x}_A^j) \text{ and } (IC_B^j) : T_B^j - t(1 - \widetilde{x}_B^j) \geq T_A^j - t\widetilde{x}_B^j. \tag{2.2}$$

Otherwise, plan i's marginal j provider may obtain a strictly higher utility by patronizing only the plan $-i$. In this single-homing case, all j providers with $x < \widetilde{x}_A^j$ join in plan A and those with $x \geq \widetilde{x}_B^j$ join in plan B.[16]

We refer $\widetilde{x}_A^j$ (or $1 - \widetilde{x}_B^j$) as the j *providers' configuration* of plan A (or B). The equilibrium sizes of plan A and B are

$$n_A = (1 - q)\widetilde{x}_A^L + q\widetilde{x}_A^H \text{ and } n_B = (1 - q)(1 - \widetilde{x}_B^L) + q(1 - \widetilde{x}_B^H).$$

We define the *quality level* of health plan i ($i = A, B$) as a percentage of H providers in plan i, which is endogenously determined as

$$q_i = \frac{n_i^H}{n_i}, \tag{2.3}$$

where $n_A^H = q\widetilde{x}_A^H$ and $n_B^H = q(1 - \widetilde{x}_B^H)$ are the sizes of H providers in the health plan A and B respectively.

To study multi-homing of H providers, we focus on the equilibrium where L providers always single-home when multi-homing is allowed. Note that an L provider with location x prefers multi-homing, if and only if $T_A^L - tx \geq 0$ and $T_B^L - t(1 - x) \geq 0$. Our focus on multi-homing equilibrium satisfies

$$T_B^L - t(1 - \widetilde{x}_A^L) < 0. \tag{2.4}$$

[15] It is because plans cannot discriminate against providers' location, i.e., T_i^j is independent to x. All j providers on the left of $\widetilde{x}_A^j$ patronize plan A and those on the right of $\widetilde{x}_B^j$ patronize plan B.

[16] Without loss of any generality, we assume that the marginal j provider of plan A only patronize plan B if single-homing is mandatory.

So L providers in plan A never multi-home when multi-homing is allowed, since (2.4) implies that for all $x \leq \widetilde{x}_A^L$, we have $T_B^L - t(1-x) < 0$. Moreover, there are only two plans. Holding of (2.4) implies that L providers always single-home if they choose any plan as home.

To be consistent, we also restrict our single-homing analysis with (2.4). In the single-homing case, $(IR_A^L) : T_A^L - t\widetilde{x}_A^L \geq 0$ must hold. Thus (2.4) also implies that $(IC_A^L) : T_A^L - t\widetilde{x}_A^L > T_B^L - t(1 - \widetilde{x}_A^L)$ holds strictly. Similarly, we have (IC_B^L) holds strictly. Therefore, both (IC_A^L) and (IC_B^L) will be slack in plans' single-homing problems. Moreover, taking (2.4) into the single-homing case does not require any more technic assumptions than the multi-homing do.

2.3.2 Policyholder Side

As in Bardey and Rochet (2010), there is a continuum of policyholders, indexed by θ, independently drawn from $(0, 1]$ according to a cumulative distribution function F, which is continuously differentiable and satisfies the monotone hazard rate property, i.e., $\frac{1-F(\theta)}{f(\theta)}$ decreases with θ. Here θ represents the policyholder's probability of becoming ill and is private information, but the distribution is common knowledge. Every policyholder chooses only one plan and pays a fixed premium P_i $(i = A, B)$ to it.[17]

The driving force behind health plans is the "thick market" effect; a large network provides better chances for finding a trading partner. The expected utility of a policyholder with θ patronizing plan i $(i = A, B)$ takes the following separable form:

$$Eu_i(\theta) = \omega - P_i + \theta n_i \left[\lambda + q_i\right],$$

where ω is a fixed utility gain from policyholders participating in one plan and is assumed to be high enough that premium P_i charged by health plan i can be strictly positive.[18] Moreover, this fixed utility gain cannot be doubled by joining two plans.[19] $\lambda \geq 0$ is policyholders' marginal preference over a plan's size compared to its quality. Thus $n_i \left[\lambda + q_i\right]$ is plan i's aggregate externality for policyholders. Moreover, from the

[17]According to Bardey and Rochet (2010), the fixed premium (the imperfect risk adjustment) is justified by preventing health plans from setting premiums on an individual risk adjustment fashion to achieve fairness. As long as the out-of-pocket payment is only a premium, every policyholder will choose the only one health plan which gives him the highest utility level.

[18]The policyholders' reserve utility is also normalized at 0.

[19]Here is another reason why policyholders usually choose to single-home in health plan markets. Governments sometimes compensate policyholders financially when they join in a health plan program, but do not compensate multiple times. If the health plan program is mandatory, ω can be also regarded as the common financial budget of policyholders.

definition of q_i (formula (2.3)), we have

$$\theta n_i [\lambda + q_i] = \theta(1+\lambda) \left[\frac{\lambda}{1+\lambda} n_i^L + n_i^H \right], \tag{2.5}$$

where $i = A, B$ and $n_i^L = n_i - n_i^H$ is the size of plan i's L providers. The right side of (2.5) implies that the externality exerted by a H provider is strictly higher than a L provider, the quality difference arises. When λ is small, compared to H providers, L providers' *relative quality* $\frac{\lambda}{1+\lambda}$ (which is the substitutability of a L provider to a H provider) is negligible.[20]

Policyholders care about having access to high quality providers. One can consider L providers as general practitioners and H providers as specialists. All physicians can identify and treat a specific illness if it is detected early, but only specialists can deliver proper treatment at a later stage. Another example is that an experienced doctor can obtain a definite diagnosis quickly and easily from a patient's description of the symptoms and regular checkups, while a novice physician may take a longer time to reach the same conclusion or make more mistakes. Moreover, $\theta n_i [\lambda + q_i]$ also can be considered as a reduced form of a policyholder's utility resulting from a sub-game where all policyholders prefer to be treated by an H physician, but could not observe their types.[21]

Risk segmentation: We focus on asymmetric health plans which lead to adverse selection. Without loss of any generality, we assume that plan B always provides better services with a higher premium than plan A, that is $n_A [\lambda + q_A] < n_B [\lambda + q_B]$ and $P_B - P_A > 0$.[22] A policyholder with risk θ chooses plan A over plan B if

$$\theta n_A [\lambda + q_A] - P_A > \theta n_B [\lambda + q_B] - P_B$$

In equilibrium, policyholder θ will choose plan A if $\theta < \tilde{\theta}$, or goes to plan B otherwise, where

$$\tilde{\theta} = \frac{P_B - P_A}{\lambda(1-q)(1-\tilde{x}_B^L - \tilde{x}_A^L) + (\lambda+1)q(1-\tilde{x}_B^H - \tilde{x}_A^H)} \tag{2.6}$$

This $\tilde{\theta}$ represents the marginal policyholder who is indifferent with either plan. The equilibrium demand of plan A (or B) on the policyholder side is $F(\tilde{\theta})$ (or $1 - F(\tilde{\theta})$).

[20] Please do not confuse *relative quality* with the *quality level* introduced above. They both indicate providers' quality and can be introduced as one index. However, for simplicity, we sperate them from each other. *Quality level* refers to percentage of H providers in one plan and *relative quality* refers to the impact of providers' qualities on policyholders' utility.

[21] The problem is withheld information. Policyholders can't verify each physician's claim of being more competent. Although the health plan here acts as agent of all policyholders and could endorse each doctor, it is not at his interest to tell the policyholder all the information. The detail of the sub-game can be found in Appendix.

[22] If $n_A [\lambda + q_A] = n_B [\lambda + q_B]$, the private information θ is useless for any policyholder.

2.3.3 Two Competing Health Plans

We assume that total treatment costs are in proportion to the total number of treatments. c is the fixed cost per treatment. Since all costs related to treatment are eventually paid by the plans, the expected profits of plan A and B are given as follows:

$$\begin{aligned}\pi_A &= P_A F(\tilde{\theta}) - (1-q)T_A^L \widetilde{x}_A^L - qT_A^H \widetilde{x}_A^H - c\int_0^{\tilde{\theta}} \theta f(\theta) d\theta \\ \pi_B &= P_B \left[1 - F(\tilde{\theta})\right] - (1-q)T_B^L(1-\widetilde{x}_B^L) - qT_B^H(1-\widetilde{x}_A^H) - c\int_{\tilde{\theta}}^1 \theta f(\theta) d\theta.\end{aligned}$$

The equilibrium is given by $(P_i, T_i^j, \widetilde{x}_i^j, \tilde{\theta})$ for all $j = H, L$,$i = A, B$, where $\tilde{\theta}$ is determined by $(P_i, \widetilde{x}_i^j)$ according to (2.6), $P_A, T_A^j, \widetilde{x}_A^j$ maximizes π_A and $P_B, T_B^j, \widetilde{x}_B^j$ maximizes π_B to satisfy their respective constraints.[23]

As described in subsection 2.3.1, in the multi-homing case, plan i only faces the *individually rationality constraints* (IR_i^j) in formula (2.1); while in the single-homing case, besides (IR_i^j), plan i also faces *incentive constraints* (IC_i^j) in formula (2.2). Moreover, we focus on cases with slack (IC_i^L) (see subsection 2.3.1). Plan i's single-homing problem is constrained by (IR_i^L), (IR_i^H) and (IC_i^H).

2.4 Equilibrium under Fixed Salary Scheme

When providers receive a fixed salary, each of them gets a constant payment which is unrelated to the actual number of treatment he or she offers.[24] Let W_i^j denote a type j provider's salary if he chooses plan i, where $j = H, L$, $i = A, B$. Then j provider's payment from health plan i is $T_i^j = W_i^j$. The profits of plans can be written as following:

$$\begin{aligned}\pi_A &= P_A F(\widetilde{\theta}) - (1-q)W_A^L \widetilde{x}_A^L - qW_A^H\ \widetilde{x}_A^H - c\int_0^{\widetilde{\theta}} \theta f(\theta) d\theta \text{ and} \\ \pi_B &= P_B(1 - F(\widetilde{\theta})) - (1-q)W_B^L(1-\widetilde{x}_B^L) - qW_B^H(1-\widetilde{x}_B^H) - c\int_{\widetilde{\theta}}^1 \theta f(\theta) d\theta.\end{aligned}$$

The marginal j $(j = H, L)$ providers' *individually rationality constraints* are

$$(IR_A^j) : W_A^j - t\widetilde{x}_A^j \geq 0 \text{ and } (IR_B^j) : W_B^j\ - t(1-\widetilde{x}_B^j) \geq 0. \tag{2.7}$$

[23] We assume that these are both concave problems.

[24] Staff-model HMOs usually select this kind of payment scheme.

And their *incentive constraints are*

$$(IC_A^j): W_A^j - t\widetilde{x}_A^j \geq W_B^j - t(1-\widetilde{x}_A^j) \text{ and } (IC_B^j): W_B^j - t(1-\widetilde{x}_B^j) \geq W_A^j - t\widetilde{x}_B^j, \quad (2.8)$$

Here (IR_i^j) and (IC_i^j) are independent with the risk segmentation $\tilde{\theta}$. This implies that the fixed salaries do not introduce any network externality on provider side. Participating in one plan or both is motivated by whether or not the providers' transportation cost is covered by the salary. If the changing environment on the provider side causes any variation in shares on the provider side, both plans will optimally adjust their premiums accordingly, such that equilibrium risk segmentation $\tilde{\theta}$ does not change. More precisely, with (2.6), the first order conditions for the premiums can be written as

$$\frac{\partial \pi_A}{\partial P_A} = F(\tilde{\theta}) - (\frac{P_A - c\tilde{\theta}}{P_B - P_A} f(\tilde{\theta})\tilde{\theta}) = 0, \quad (2.9)$$

$$\frac{\partial \pi_B}{\partial P_B} = 1 - F(\tilde{\theta}) - (\frac{P_B - c\tilde{\theta}}{P_B - P_A} f(\tilde{\theta})\tilde{\theta}) = 0. \quad (2.10)$$

Combining (2.9) and (2.10), we have $1-2F(\tilde{\theta}) \equiv f(\tilde{\theta})\tilde{\theta}$, for all t, q, λ and for both multi-homing and single-homing cases. Moreover, by offering better services (i.e., $n_A[\lambda + q_A] < n_B[\lambda + q_B]$), plan B secures a higher market share (i.e., $1 - F(\tilde{\theta}) > F(\tilde{\theta})$). Let $\widehat{\theta}$ be the *only* θ satisfying $\widehat{\theta} = \frac{1-2F(\widehat{\theta})}{f(\widehat{\theta})}$.[25] We make the following assumption:

Assumption A2.1. $t \in (\underline{t},\ \infty)$, where $\underline{t} = \max\{\frac{\lambda}{2}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}(1 - F(\widehat{\theta}))\}$.

With $t > \frac{\lambda}{2}\widehat{\theta}$, L providers never multi-home in all cases and (IC_i^L) $(i = A, B)$ will be slack. With $t > \frac{1+\lambda}{2}\widehat{\theta}(1 - F(\widehat{\theta}))$, we can focus on interior solutions, i.e., $0 < \widetilde{x}_i^j < 1$, for $j = H, L$,$i = A, B$. [26]

We first determine the equilibrium with single-homing and then with multi-homing. After that we make a comparison between the two.

[25]Because F satisfies the monotone hazard rate property, $\widehat{\theta}$ is uniquely defined. For example, consider a iso-elastic distribution $F(\theta) = \theta^{\varepsilon}$ with $\varepsilon > 0$, $\widehat{\theta} = (2+\varepsilon)^{-\frac{1}{\varepsilon}}$; if $\varepsilon = 1$, F is uniform distribution function, $\widehat{\theta} = \frac{1}{3}$.

[26]We exclude the workload limitations of providers and we can relax this assumption. If we assume that each physician can provide only 1 treatment, two other assumptions can be added, $t \in (\underline{t}, \overline{t})$, where $\underline{t} = \max\{\frac{\lambda}{2}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}(1 - F(\widehat{\theta}))\}$, $\overline{t} = (\lambda + \frac{1}{2})\frac{\widehat{\theta}}{E(\theta|\theta > \widehat{\theta})}$, and $E(\theta|\theta \leq \widehat{\theta}) + E(\theta|\theta > \widehat{\theta}) \leq \frac{1+\lambda}{\lambda}$.With these assumptions, health plans will ensure that every policyholder gets the treatment they need, i.e., the total number of physicians in each plan should be more than the expected number of treatments $(1-y)\widetilde{x}_A^L + y\widetilde{x}_A^H \geq \int_0^{\tilde{\theta}} \theta f(\theta)d\theta$, $(1-y)(1-\widetilde{x}_B^L) + y(1-\widetilde{x}_B^H) \geq \int_{\tilde{\theta}}^1 \theta f(\theta)d\theta$. And every physician holding multiple contracts will not be overloaded, i.e., $\frac{\int_0^{\tilde{\theta}} \theta f(\theta)d\theta}{(1-y)\widetilde{x}_A^L + y\widetilde{x}_A^H} + \frac{\int_{\tilde{\theta}}^1 \theta f(\theta)d\theta}{(1-y)(1-\widetilde{x}_B^L) + y(1-\widetilde{x}_B^H)} \leq 1$. These two additional assumptions will not change our conclusions.

.

2.4.1 Single-homing Equilibrium under Fixed Salary Scheme

Now, let us find the equilibrium *providers' configurations* when a provider can only sign with one plan. In this case, besides meeting *individual rationality constraints* (IR_i^j) in (2.7), both plans also need to make their *incentive constraints* (IC_i^H) in (2.8) hold, for all $j = H, L$, $i = A, B$.

We claim in the above subsection that under *A2.1*, (IC_i^L) $(i = A, B)$ is slack because formula (2.4) holds. If that is true, (IR_i^L) $(i = A, B)$ must be binding in equilibrium, because the salary W_i^L, which is considered a cost of plan i, should be as low as possible. This implies $W_A^L = t\widetilde{x}_A^L$ and $W_B^L = t(1 - \widetilde{x}_B^L)$. Then formula (2.4) is equivalent to

$$1 - \widetilde{x}_B^L + \widetilde{x}_A^L < 1. \tag{2.11}$$

The first order conditions with respect to $\widetilde{x}_A^L$ and $1 - \widetilde{x}_B^L$ are

$$\begin{aligned}
\frac{\partial \pi_A}{\partial \widetilde{x}_A^L} &= (1-q)\lambda\tilde{\theta}(\frac{P_A - c\tilde{\theta}}{P_B - P_A} f(\tilde{\theta})\tilde{\theta}) - 2t(1-q)\widetilde{x}_A^L = 0, \\
\frac{\partial \pi_B}{\partial (1 - \widetilde{x}_B^L)} &= (1-q)\lambda\tilde{\theta}(\frac{P_B - c\tilde{\theta}}{P_B - P_A} f(\tilde{\theta})\tilde{\theta}) - 2t(1-q)(1 - \widetilde{x}_B^L) = 0.
\end{aligned}$$

Combing these 2 first order conditions with (IR_i^L), (2.9), (2.10) and $\tilde{\theta} \equiv \widehat{\theta}$, we have L *providers' configurations* and L providers' salaries of both plans:

$$\begin{aligned}
\widetilde{x}_A^L &= \frac{\lambda}{2t}\widehat{\theta}F(\widehat{\theta}) \text{ and } W_A^L = \frac{\lambda}{2}\widehat{\theta}F(\widehat{\theta}); \\
1 - \widetilde{x}_B^L &= \frac{\lambda}{2t}\widehat{\theta}(1 - F(\widehat{\theta})) \text{ and } W_B^L = \frac{\lambda}{2}\widehat{\theta}(1 - F(\widehat{\theta})).
\end{aligned}$$

So formula (2.11) holds under *assumption A2.1*, because of $1 - \widetilde{x}_B^L + \widetilde{x}_A^L = \frac{\lambda}{2t}\widehat{\theta} < 1$ with $t > \frac{\lambda}{2}\widehat{\theta}$. Holding of (2.11) suggests that L providers are not fully covered, i.e., there is some L provider excluded by these two plans. This implies that each health plan has local monopoly power when hiring L providers.

Similarly, when (IC_i^H) $(i = A, B)$ is slack, (IR_i^H) must be binding in equilibrium and the following inequality holds:

$$1 - \widetilde{x}_B^H + \widetilde{x}_A^H < 1. \tag{2.12}$$

Therefore health plans are local monopolies on the provider side because both (2.11) and (2.12) hold. Moreover, the quality is independent with location. The equilibrium H *providers' configurations* are $\widetilde{x}_A^H = (\frac{\lambda}{1+\lambda})^{-1}\widetilde{x}_A^L$ and $1 - \widetilde{x}_B^H = (\frac{\lambda}{1+\lambda})^{-1}(1 - \widetilde{x}_B^L)$, where $\frac{\lambda}{1+\lambda}$ is L providers' *relative quality*, which is also the ability to substitute an L provider for an H provider. Furthermore, binding of (IR_i^H) demands $W_A^H = t\widetilde{x}_A^H$ and $W_B^H = t(1 - \widetilde{x}_B^H)$.

In this case, H *providers' configurations* and salaries are

$$\begin{aligned}\widetilde{x}_A^H &= \frac{1+\lambda}{2t}\widehat{\theta}F(\widehat{\theta}) \text{ and } W_A^H = \frac{1+\lambda}{2}\widehat{\theta}F(\widehat{\theta});\\ 1-\widetilde{x}_B^H &= \frac{1+\lambda}{2t}\widehat{\theta}(1-F(\widehat{\theta})) \text{ and } W_B^H = \frac{1+\lambda}{2}\widehat{\theta}(1-F(\widehat{\theta})).\end{aligned}$$

However, as (2.12) requires that H provider market is not fully covered, $1-\widetilde{x}_B^H+\widetilde{x}_A^H = \frac{1+\lambda}{2t}\widehat{\theta} < 1$ must hold. This requires $t > \frac{1+\lambda}{2}\widehat{\theta}$.

When $t \in (\underline{t}, \frac{1+\lambda}{2}\widehat{\theta}]$,[27] H providers will be fully covered, since (13) does not hold any more. Note that in single-homing case, $1-\widetilde{x}_B^H+\widetilde{x}_A^H \leq 1$ must hold, since for $i=A,B$, any provider enrolled by plan i cannot join in plan $-i$. Then we must have $\widetilde{x}_A^H = \widetilde{x}_B^H$, the marginal H provider of plan A is also that of plan B, because of $1-\widetilde{x}_B^H+\widetilde{x}_A^H = 1$. Now *incentive constraints* will affect the strategies of both plans. Furthermore, when t is very small, the competition can be strengthened enough such that the marginal H provider will be left with strictly positive utility, i.e., (IR_i^H) is slack and

$$(IC_i^H): W_A^H - t\widetilde{x}_A^H = W_B^H - t(1-\widetilde{x}_A^H) > 0. \tag{2.13}$$

Then $W_A^H = W_B^H + 2t\widetilde{x}_A^H - t$ and $W_B^H = W_A^H + 2t(1-\widetilde{x}_B^H) - t$. The first order conditions with respect to $\widetilde{x}_A^H$ and $1-\widetilde{x}_B^H$ are

$$\begin{aligned}\frac{\partial \pi_A}{\partial \widetilde{x}_A^H} &= q(1+\lambda)\tilde{\theta}(\frac{P_A - c\tilde{\theta}}{P_B-P_A}f(\tilde{\theta})\tilde{\theta}) - q(W_B^H + 4t\widetilde{x}_A^H - t) = 0,\\ \frac{\partial \pi_A}{\partial \widetilde{x}_A^H} &= q(1+\lambda)\tilde{\theta}(\frac{P_B - c\tilde{\theta}}{P_B-P_A}f(\tilde{\theta})\tilde{\theta}) - q(W_A^H + 4t(1-\widetilde{x}_B^H) - t) = 0.\end{aligned}$$

Combing these 2 first order conditions with (2.9), (2.10) and $\tilde{\theta} \equiv \widehat{\theta}$, H *providers' configurations* and H providers' salaries are

$$\begin{aligned}\widetilde{x}_A^H &= \frac{1}{2} - \frac{1+\lambda}{6t}\widehat{\theta}(1-2F(\widehat{\theta})) \text{ and } W_A^H = \frac{1+\lambda}{3}\widehat{\theta}(1+F(\widehat{\theta})) - t;\\ 1-\widetilde{x}_B^H &= \frac{1}{2} + \frac{1+\lambda}{6t}\widehat{\theta}(1-2F(\widehat{\theta})) \text{ and } W_B^H = \frac{1+\lambda}{3}\widehat{\theta}(2-F(\widehat{\theta})) - t.\end{aligned}$$

Moreover, only $t < \frac{1+\lambda}{3}\widehat{\theta}$ ensures (13) holds. This equilibrium only appears when $\underline{t} = \max\{\frac{\lambda}{2}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}(1-F(\widehat{\theta}))\} < \frac{1+\lambda}{3}\widehat{\theta}$.[28]

[27] $\underline{t} = \max\{\frac{\lambda}{2}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}(1-F(\widehat{\theta}))\} < \frac{1+\lambda}{2}\widehat{\theta}$

[28] $(\underline{t}, \frac{1+\lambda}{3}\widehat{\theta}) = \varnothing$ is possible. For example, if $\lambda \geq 2$, $\underline{t} \leq \frac{\lambda}{2}\widehat{\theta} \leq \frac{1+\lambda}{3}\widehat{\theta}$.

If $t \in [\frac{1+\lambda}{3}\widetilde{\theta}, \frac{1+\lambda}{2}\widetilde{\theta}]$, multiple equilibria exist. In the following proposition, we use subscript S to indicate single-homing and present a continuous equilibrium where $1-\widetilde{x}_B^H-\widetilde{x}_A^H$ decreases with t.

Proposition 2.1. *Under A2.1 and fixed salary scheme, for all $t \in (\underline{t}, \infty)$, the single-homing equilibrium risk segmentation is $\widehat{\theta} = \frac{1-2F(\widehat{\theta})}{f(\widehat{\theta})}$. On the provider side, the single-homing equilibrium configuration of L is,*

- *for all $t \in (\underline{t}, \infty)$, $\widetilde{x}_A^L = \frac{\lambda}{2t}\widehat{\theta}F(\widehat{\theta})$ and $1 - \widetilde{x}_B^L = \frac{\lambda}{2t}\widehat{\theta}(1 - F(\widehat{\theta}))$.*

And that of H providers is,

- *if $t \in (\frac{1+\lambda}{2}\widehat{\theta}, \infty)$, $\widetilde{x}_A^{H,S} = \frac{1+\lambda}{2t}\widehat{\theta}F(\widehat{\theta})$ and $1 - \widetilde{x}_B^{H,S} = \frac{1+\lambda}{2t}\widehat{\theta}(1 - F(\widehat{\theta}))$;*
- *if $t \in [\frac{1+\lambda}{3}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}]$, $\widetilde{x}_A^{H,S} = F(\widehat{\theta})$ and $1 - \widetilde{x}_B^{H,S} = 1 - F(\widehat{\theta})$;*
- *if $t \in (\underline{t}, \frac{1+\lambda}{3}\widehat{\theta}) \neq \varnothing$, $\widetilde{x}_A^{H,S} = \frac{1}{2} - \frac{1+\lambda}{6t}\widehat{\theta}(1-2F(\widehat{\theta}))$ and $1-\widetilde{x}_B^{H,S} = \frac{1}{2} + \frac{1+\lambda}{6t}\widehat{\theta}(1-2F(\widehat{\theta}))$.*

The Hotelling's framework is well known to be unsuitable for passing from uncovered to covered markets, where multiple equilibria arise.

Illustration example of multiple equilibria($\lambda = 0$)**:** Consider $F(\theta) = \theta$, i.e., policyholders' risks are uniformly distributed. When $\lambda = 0$, there is only quality externality. The *risk segmentation* is reduced from (6) to

$$\tilde{\theta} = \frac{P_B - P_A}{q(1 - \widetilde{x}_B^H - \widetilde{x}_A^H)}.$$

$1-\widetilde{x}_B^H-\widetilde{x}_A^H > 0$ must hold in equilibrium because of $P_B > P_A$. Moreover, the equilibrium *risk segmentation* $\tilde{\theta} \equiv \widehat{\theta} = \frac{1-2F(\widehat{\theta})}{f(\widehat{\theta})}$ for all $\lambda \geq 0$. Thus $\widehat{\theta} = \frac{1}{3}$ and $F(\widehat{\theta}) = \frac{1}{3}$. When $t \in [\frac{1+\lambda}{3}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}] = [\frac{1}{9}, \frac{1}{6}]$, IR_i^H and IC_i^H $(i = A, B)$ are all binding, that is

$$\widetilde{x}_A^H = \widetilde{x}_B^H \text{ and } W_B^H - t(1 - \widetilde{x}_B^H) = W_A^H - t\widetilde{x}_A^H = 0$$

To solve $(\widetilde{x}_A^H, \widetilde{x}_B^H, W_A^H, W_B^H)$, the system is under-identified. Since $\widetilde{x}_A^H = \frac{1+\lambda}{2t}\widehat{\theta}F(\widehat{\theta})$ when A is a local monopoly on the provider side and $1 - \widetilde{x}_B^H = \frac{1+\lambda}{2t}\widehat{\theta}(1 - F(\widehat{\theta}))$ when B is a local monopoly on the provider side, any $\widetilde{x}_A^H$ can then be equilibrium, if 1) $\widetilde{x}_A^H < \frac{1}{2}$ (since $1 - \widetilde{x}_B^H - \widetilde{x}_A^H > 0$); 2) $\widetilde{x}_A^H \leq \frac{1+\lambda}{2t}\widehat{\theta}F(\widehat{\theta}) = \frac{1}{18t}$; 3) $1 - \widetilde{x}_A^H \leq \frac{1+\lambda}{2t}\widehat{\theta}(1 - F(\widehat{\theta})) = \frac{1}{9t}$. Thus all possible equilibria $\widetilde{x}_A^H$ with $t \in [\frac{1}{9}, \frac{1}{6}]$ are in the shadow area of Figure 2.1. Note that in the above example $W_B^H = t(1 - \widetilde{x}_B^H) > W_A^H = t\widetilde{x}_A^H$. To be close to its local monopoly status $1 - \widetilde{x}_B^{H,S} = \frac{1+\lambda}{2t}\widehat{\theta}(1 - F(\widehat{\theta}))$, plan B feasibly offers a higher salary to steal H providers from plan A and charge a higher premium to remian profitability. Moreover, 1.pdf 1.pdf

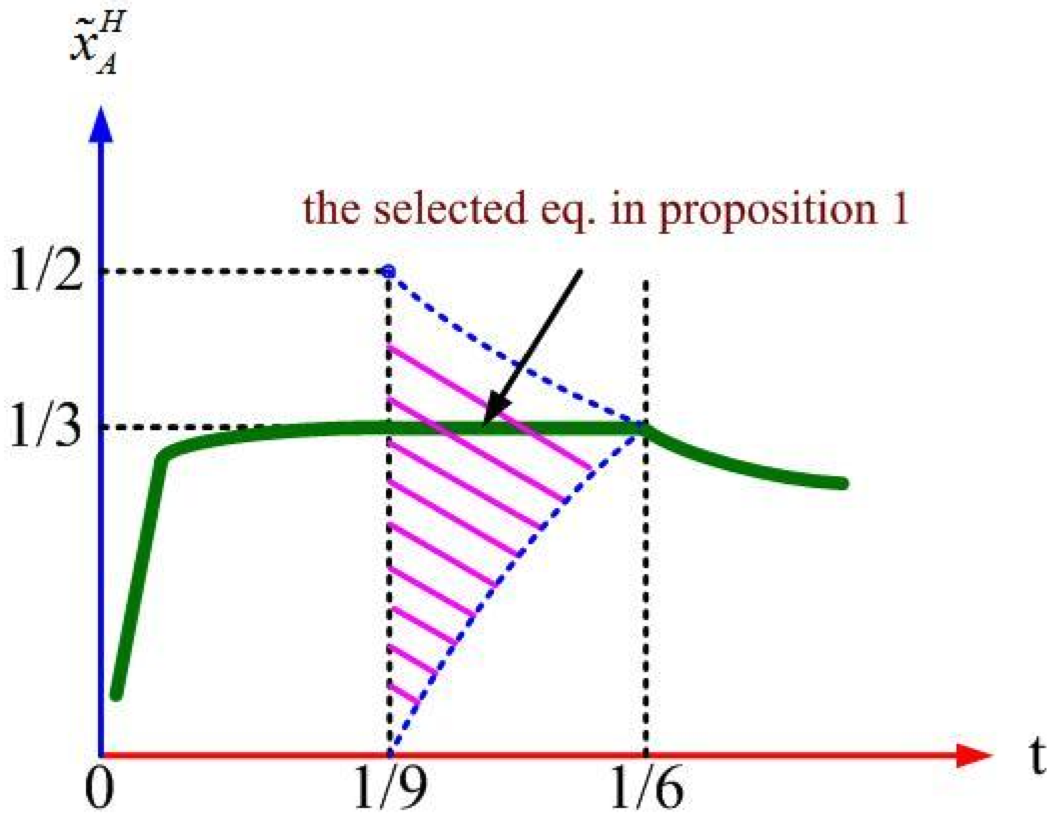

FIGURE 2.1: The equilibrium value of $\widetilde{x}_A^H$ with single-homing.

plan B's ability to steal H providers should increase as salary becomes cheaper, or as t decreases. Thus we should expect $1-\widetilde{x}_B^H-\widetilde{x}_A^H$ to (may not strictly) decrease with t. The equilibrium in proposition 2.1 is selected according to this intuition: for $t\in(\underline{t},\frac{1+\lambda}{3}\widehat{\theta})\neq\varnothing$, $1-\widetilde{x}_B^{H,S}-\widetilde{x}_A^{H,S}=\frac{1+\lambda}{3t}\widehat{\theta}(1-2F(\widehat{\theta}))$ decreases with t and strictly larger than $1-2F(\widehat{\theta})$, which the value of $1-\widetilde{x}_B^{H,S}-\widetilde{x}_A^{H,S}$ for all $t\in[\frac{1+\lambda}{3}\widehat{\theta},\frac{1+\lambda}{2}\widehat{\theta}]$; if $t\in(\frac{1+\lambda}{2}\widehat{\theta},\infty)$, $1-\widetilde{x}_B^H-\widetilde{x}_A^H=\frac{1+\lambda}{2t}\widehat{\theta}(1-2F(\widehat{\theta}))$ decreases with t and strictly less than $1-2F(\widehat{\theta})$. Then we have the following remark.

Remark 2.1 *The single-homing equilibrium selected in proposition 2.1 is the only continuous equilibrium characterized by* $1-\widetilde{x}_B^H-\widetilde{x}_A^H$ *decreasing with* t.

2.4.2 Multi-homing Equilibrium under Fixed Salary Scheme

Here multi-homing is allowed. Plans will maximize their profits subject to the providers' *individual rational constraints* (IR_i^j) in (2.7). In this case the equilibrium takes the same form as the single-homing case when *incentive constraints* (IC_i^H) $(i=A,B)$ is slack. We use subscript M to indicate multi-homing and present the result as follows:

2.pdf 2.pdf

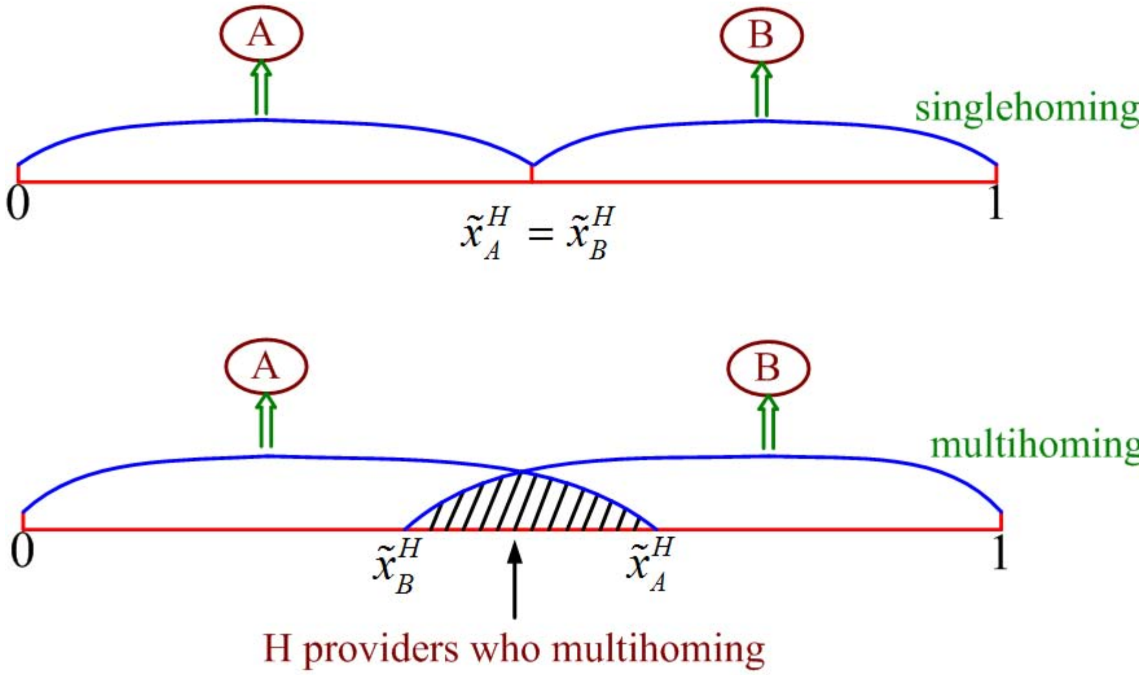

FIGURE 2.2: Single-homing and Multi-homing of H providers when $t \in (\underline{t}\ ,\ \frac{1+\lambda}{2}\widehat{\theta}]$.

Proposition 2.2. *Under A2.1 and the fixed salary scheme, for all* $t \in (\underline{t}, \infty)$*, the only multi-homing equilibrium is characterized by risk segmentation* $\widehat{\theta} = \frac{1-2F(\widehat{\theta})}{f(\widehat{\theta})}$ *and providers' configurations:*

- $\widetilde{x}_A^{L,M} = \frac{\lambda}{2t}\widehat{\theta}F(\widehat{\theta})$ *and* $1 - \widetilde{x}_B^{L,M} = \frac{\lambda}{2t}\widehat{\theta}(1-F(\widehat{\theta}))$*;*
- $\widetilde{x}_A^{H,M} = \frac{1+\lambda}{2t}\widehat{\theta}F(\widehat{\theta})$ *and* $1 - \widetilde{x}_B^{H,M} = \frac{1+\lambda}{2t}\widehat{\theta}(1-F(\widehat{\theta}))$*.*

Moreover, L providers will never have incentive to multi-home; multi-homing of H provider appears only when $t \in (\underline{t}\ ,\ \frac{1+\lambda}{2}\widehat{\theta}]$*.*

L providers will never have incentive to multi-home because of $1 - \widetilde{x}_B^{L,M} + \widetilde{x}_A^{L,M} = \frac{\lambda}{2t}\widehat{\theta} < 1$ with $t > \frac{\lambda}{2}\widehat{\theta}$. Furthermore, since $t > \underline{t} \geq \frac{1+\lambda}{2}\widehat{\theta}(1-F(\widehat{\theta}))$, for all $j = L, H$ and $i = A, B$, we have $0 < \widetilde{x}_i^{j,M} < 1$. Thus, the equilibrium given by proposition 2.2 is an interior solution. When $t \in (\underline{t}\ ,\ \frac{1+\lambda}{2}\widehat{\theta}]$, we have $\widetilde{x}_B^{H,M} \leq \widetilde{x}_A^{H,M}$ and only H providers with $x \in [\widetilde{x}_B^{H,M}, \widetilde{x}_A^{H,M}]$ multi-home. Figure 2.2 demonstrates single-homing and multi-homing H providers' configuration when $t \in (\underline{t}\ ,\ \frac{1+\lambda}{2}\widehat{\theta}]$.

2.4.3 Comparison between Multi-homing and Single-homing under Fixed Salary Scheme

In this subsection, we compare the multi-homing outcome with that of single-homing, when some H type providers have incentive to multi-home, i.e., $t \in (\underline{t}, \frac{1+\lambda}{2}\widehat{\theta}]$ (see proposition 2.2). First, L providers' configurations remain the same, i.e., $\widetilde{x}_i^{L,M} = \widetilde{x}_i^{L,S}$ for $i = A, B$ (see proposition 2.1 and 2.2). Intuitively, when a fixed salary scheme is adopted, providers are lack of (*ex post*) incentive to compete over the policyholders in the same health plan.[29] As a result, the only determinant of L type's involvement is the *risk segmentation*, which stays constant when the environment shifts from single-homing to multi-homing. Second, for any $t \in (\underline{t}, \frac{1+\lambda}{2}\widehat{\theta}]$, both health plans involve more of H providers with multi-homing than with single-homing, i.e., $\widetilde{x}_A^{H,M} \geq \widetilde{x}_A^{H,S}$ and $1-\widetilde{x}_B^{H,M} \geq 1-\widetilde{x}_B^{H,S}$. As defined in formula (2.3), a health plan's *quality level* is the percentage of H type providers in it. Thus, multi-homing of H providers results in the highest quality health plans among all outcomes. We present this result in the following proposition.

Proposition 2.3. *Under A2.1 and fixed salary scheme, both health plans' quality levels are* $\frac{(1+\lambda)q}{\lambda+q}$ *with multi-homing, which is higher than* q*, the average quality level in the whole provider population and health plans' quality levels with single-homing.*

When multi-homing is allowed, *quality level*s of both health plan are higher than q, the average *quality* of the physician market. This explains why managed care, due to its ability to identify a physician's competence, offers a higher quality of care than traditional indemnity insurance.

The fact that multi-homing leads to higher *quality levels* than single-homing may suggest why single-homing staff-mode HMOs (Health Maintenance Organizations) faded away in the U.S.. In the staff model, physicians have offices in HMO buildings and may only see HMO patients. Two rules have hit staff-model HMOs particularly hard: market demands for high quality (broad choice of high quality providers) and greater geographic accessibility. Now a captive group model has replaced the staff-model. Providers continue to treat non-HMO patients and physicians can travel from one hospital to another. However, there are criticisms. The problem is congestion caused by providers' multi-homing. One cannot assume a provider's quality to be consistent when some providers treat many patients in a short period. Moreover, we may find other negative impact of multi-homing than "congestion".

Note that with multi-homing, both health plans choose the same quality. The reason is that by allowing multi-homing, health plans do not directly impact their rivals' strategies

[29] In contrast, this interdependency among providers becomes stronger when FFS is adopted,

of contracting providers, thus both of them can choose a blend of providers with mixture rates $(\frac{\widetilde{x}_A^L}{\widetilde{x}_A^H}, \frac{1-\widetilde{x}_B^L}{1-\widetilde{x}_B^H})$ equal to $\frac{\lambda}{1+\lambda}$, the policyholders' substitution rate of a L provider to a H provider.[30] It also implies that plans' competitive intensity is decreased by giving providers the permission to multi-home. This increases both plans' profits.[31]

Corollary 2.1 *Under A2.1 and the fixed salary scheme, both health plans prefer multi-homing than single-homing.*

With multi-homing), salaries of H providers are $W_A^{H,M} = \frac{1+\lambda}{2}\widehat{\theta}F(\widehat{\theta})$, $W_B^{H,M} = \frac{1+\lambda}{2}\widehat{\theta}(1-F(\widehat{\theta}))$. With single-homing, if $t \in [\frac{1+\lambda}{3}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}]$, the salaries are $W_A^{H,S} = tF(\widehat{\theta})$, $W_B^{H,S} = t(1-F(\widehat{\theta}))$; or if $t \in (\underline{t}, \frac{1+\lambda}{3}\widehat{\theta}) \neq \varnothing$, $W_A^{H,S} = \frac{1+\lambda}{3}\widehat{\theta}(1+F(\widehat{\theta}))-t$, $W_B^{H,S} = \frac{1+\lambda}{3}\widehat{\theta}(2-F(\widehat{\theta}))-t$. Remember that $F(\widehat{\theta}) < \frac{1}{2}$ and $t > \frac{1+\lambda}{2}\widehat{\theta}(1-F(\widehat{\theta}))$. Thus, for any $t \in (\underline{t}, \frac{1+\lambda}{2}\widehat{\theta}]$, we have $W_i^{H,M} \geq W_i^{H,S}$, $i = A, B$. Moreover, $W_i^{L,M} = W_i^{L,S}$. Let us define the surplus on the provider side (side 1) as the aggregate utility of all providers:

$$CS^1 = \sum_{j=L,H} q^j (\int_0^{\widetilde{x}_A^j} (W_A^j - tx)dx + \int_{\widetilde{x}_B^j}^1 (W_B^j - t(1-x))dx),$$

where $q^H = q$, $q^L = 1-q$. Because $\widetilde{x}_i^{L,M} = \widetilde{x}_i^{L,S}$ for $i = A, B$ and $\widetilde{x}_A^{H,M} \geq \widetilde{x}_A^{H,S}$ and $1 - \widetilde{x}_B^{H,M} \geq 1 - \widetilde{x}_B^{H,S}$, multi-homing equilibrium leads to a larger provider surplus than the single-homing does.[32]

Corollary 2.2 *Under A2.1 and the fixed salary scheme, the provider side receives a higher total surplus with multi-homing than with single-homing.*

Now, we turn to welfare comparison. The utilitarian form of the social welfare function is chosen. Thus, social welfare is simply the sum of the policyholders' benefit minus the social cost of providing health care service.

$$\begin{aligned} SW &= (n_A(\lambda + q_A)) \int_0^{\tilde{\theta}} \theta f(\theta)d\theta + (n_B(\lambda + q_B)) \int_{\tilde{\theta}}^1 \theta f(\theta)d\theta - cE(\theta) \\ &\quad - \frac{t}{2}\left\{(1-q)\left[\widetilde{x}_A^{L2} + (1-\widetilde{x}_B^L)^2\right] - q\left[\widetilde{x}_A^{H2} + (1-\widetilde{x}_B^H)^2\right]\right\} \end{aligned} \tag{2.14}$$

Remember that change in exclusivity on the provider side does not change the equilibrium *risk segmentation*, the social aggregate externality

$$(n_A(\lambda + q_A)) \int_0^{\tilde{\theta}} \theta f(\theta)d\theta + (n_B(\lambda + q_B)) \int_{\tilde{\theta}}^1 \theta f(\theta)d\theta$$

[30] Our assumption that the distribution of quality level is independent with that of location is crucial.

[31] Mathematically, when allowing multi-homing, both plans' optimization problems are on less constrained spaces than the case of prohibiting multi-homing.

[32] Remember that providers' individually rational constraints are $W_A^j - t\widetilde{x}_A^j \geq 0$ and $W_B^j - t(1-\widetilde{x}_B^j) \geq 0$, thus policies shifting towards allowing multi-homing on the provider side leads to a *Pareto* improvement on this side.

increases with multi-homing, because it increases both the size and quality level of both plans. As long as t is relatively small ($t \in (\underline{t}, \frac{1+\lambda}{2}\widehat{\theta}]$), the (transportation) cost of enrolling more physicians will be covered by the increase in social benefit. We get the following result. A detailed proof is in appendix.

Proposition 2.4. *Under A2.1 and the fixed salary scheme, multi-homing of H providers leads to higher welfare than single-homing.*

Unlike the competitive-bottleneck scenario (see Armstrong (2006)), here an improvement in efficiency (social welfare) does not necessarily lead to an improvement of policyholders' welfare. In a competitive-bottleneck scenario, which assumes transportation cost is 0 on the provider side, all providers will multi-home if multi-homing is allowed. Once the policyholders are attracted by the plans, both plans have local monopoly power to connect the providers to policyholders (theoretically, plans can use 0 salary to employ all physicians). As a result, the provider side is left with no surplus. Consequently, the competitive intensity over policyholders is increased, and they can benefit from multi-homing on the other side. In contrast, here we assume $t > 0$ to be realistic. Some H providers do not multi-home because of the transportation cost. This means that these H providers are no longer captive thus reducing the health plans' market power. As we see in corollary 2.2, instead of leaving zero surplus to providers, the providers prefer multi-homing to single-homing. Moreover, corollary 2.1 suggests that multi-homing of H providers increases both plans' profits. There might be cases where the total increase in plans' profits and providers' surplus from multi-homing exceeds that in the total social welfare, thus cause losses on the policyholder side.

Specifically, we define the welfare of the policyholder side (side 2) as the average utility of all policyholders, that is

$$\begin{aligned} CS^{2,k} \quad = \quad & \int_0^{\widehat{\theta}} (\omega - P_A^k + \theta \left[\lambda(1-q)\widetilde{x}_A^{L,k} + (1+\lambda)q\widetilde{x}_A^{H,k}\right])f(\theta)d\theta + \\ & \int_{\widehat{\theta}}^1 (\omega - P_B^k + \theta \left[\lambda(1-q)(1-\widetilde{x}_B^{L,k}) + (1+\lambda)q(1-\widetilde{x}_B^{H,k})\right])f(\theta)d\theta \end{aligned} \tag{2.15}$$

where $k = S, M$ indicating single-homing or multi-homing respectively. The policyholder side's net benefit of multi-homing is

$$\begin{aligned} CS^{2,M} - CS^{2,S} \quad = \quad & \int_0^{\widehat{\theta}} (P_A^S - P_A^M + \theta(1+\lambda)q(\widetilde{x}_A^{H,M} - \widetilde{x}_A^{H,S}))f(\theta)d\theta + \\ & \int_{\widehat{\theta}}^1 (P_B^S - P_B^M + \theta(1+\lambda)q(\widetilde{x}_B^{H,S} - \widetilde{x}_B^{H,M}))f(\theta)d\theta. \end{aligned}$$

From the first order conditions (2.9) and (2.10), equilibrium premiums are

$$\begin{aligned} P_A &= c\widehat{\theta} + \frac{F(\widehat{\theta})}{f(\widehat{\theta})}(\lambda(1-q)(1-\widetilde{x}_B^L - \widetilde{x}_A^L) + (\lambda+1)q(1-\widetilde{x}_B^H - \widetilde{x}_A^H)), \\ P_B &= c\widehat{\theta} + \frac{1-F(\widehat{\theta})}{f(\widehat{\theta})}(\lambda(1-q)(1-\widetilde{x}_B^L - \widetilde{x}_A^L) + (\lambda+1)q(1-\widetilde{x}_B^H - \widetilde{x}_A^H)). \end{aligned}$$

When $t \in [\frac{1+\lambda}{3}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}]$, we have $\widetilde{x}_A^{H,S} = \widetilde{x}_B^{H,S} = F(\widehat{\theta})$, $\widetilde{x}_A^{H,M} = \frac{1+\lambda}{2t}\widehat{\theta}F(\widehat{\theta})$ and $\widetilde{x}_B^{H,M} = 1 - \frac{1+\lambda}{2t}\widehat{\theta}(1-F(\widehat{\theta}))$, where $\widehat{\theta} = \frac{1-2F(\widehat{\theta})}{f(\widehat{\theta})}$, so

$$\begin{aligned} P_A^S - P_A^M &= (\lambda+1)qF(\widehat{\theta})\widehat{\theta}(1-\frac{1+\lambda}{2t}\widehat{\theta}) < 0 \text{ and} \\ P_B^S - P_B^M &= (\lambda+1)q(1-F(\widehat{\theta}))\widehat{\theta}(1-\frac{1+\lambda}{2t}\widehat{\theta}) < 0. \end{aligned}$$

Here policyholders face higher premiums with multi-homing than with single-homing. Intuitively, multi-homing of providers leads to larger costs that health plans must incur, but better services (higher quality and larger size) that policyholders can obtain. This increase in price on the policyholder side is not surprising. However, for policyholders, it is questionable whether the enhancement of the premiums is worth the improvement in service. The problem is that multi-homing of some physicians does not result in a *Pareto* improvement on the policyholder side. Policyholders with lowest risk may pay much more for the service improvement than that they can benefit from it, because their marginal benefits from it are their very low risks. Especially when $t \in [\frac{1+\lambda}{3}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}]$, all of plan A's policyholders prefer single-homing over multi-homing. In detail, a policyholder in plan A can get an expected utility

$$Eu_A^k(\theta) = \omega - P_A^k + \theta\left[\lambda(1-q)\widetilde{x}_A^{L,k} + (1+\lambda)q\widetilde{x}_A^{H,k}\right],$$

where $k = S, M$ indicating single-homing or multi-homing respectively. From the above two subsections, we have $\widetilde{x}_A^{H,M} - \widetilde{x}_A^{H,S} = (\frac{1+\lambda}{2t}\widehat{\theta} - 1)F(\widehat{\theta})$, thus a plan A's policyholder with probability θ strictly prefers multi-homing of H providers if and only if

$$Eu_A^M(\theta) - Eu_A^S(\theta) = (\theta - \widehat{\theta})(\lambda+1)qF(\widehat{\theta})(\frac{1+\lambda}{2t}\widehat{\theta} - 1) > 0.$$

Because only when $\theta < \widehat{\theta}$, a policyholder will take part in plan A, we must have $Eu_A^M(\theta) - Eu_A^S(\theta) \leq 0$ as long as $t \leq \frac{1+\lambda}{2}\widehat{\theta}$. This means that multi-homing of H providers impacts every policyholder in plan A negatively. Moreover, we have the following result.

Proposition 2.5. *Under A2.1 and the fixed salary scheme, the policyholders' welfare may be higher with single-homing than with multi-homing, if* $t \in [\frac{1+\lambda}{3}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}]$ *and*

$F(\widehat{\theta})^2[\widehat{\theta} - E(\theta|\theta \leq \widehat{\theta})] > [1 - F(\widehat{\theta})]^2[E(\theta|\theta \geq \widehat{\theta}) - \widehat{\theta}]$.

Proof. When $t \in [\frac{1+\lambda}{3}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}]$, from the equilibrium selected by proposition 2.1, we have

$$CS^{2,M} - CS^{2,S} = (\lambda + 1)q(\frac{1+\lambda}{2t}\widehat{\theta} - 1)[F(\widehat{\theta})\int_0^{\widehat{\theta}}(\theta - \widehat{\theta})f(\theta)d\theta + (1 - F(\widehat{\theta}))\int_{\widehat{\theta}}^1(\theta - \widehat{\theta})f(\theta)d\theta]. \tag{2.16}$$

So $CS^{2,M} - CS^{2,S} \geq 0$, only when

$$F(\widehat{\theta})^2[\widehat{\theta} - E(\theta|\theta < \widehat{\theta})] \leq [1 - F(\widehat{\theta})]^2[E(\theta|\theta \geq \widehat{\theta}) - \widehat{\theta}].$$

□

This proposition suggests that if policies that allow H providers to multi-home does, in fact, benefit policyholders, then this benefit may be dependent upon the distribution of policyholders' risks. In the appendix, we provide an example of the distribution.[33] Moreover, formula (2.16) suggests that even if the policyholders' welfare is increased by multi-homing, the benefit will be very small when q (the total population of H providers) and $\frac{\lambda}{1+\lambda}$ (the relative quality of L providers) is very low.

corollary 2.3 *If the policyholder's welfare can be improved by multi-homing of H providers under A2.1 and the fixed salary scheme, that improvement cannot be significant unless the total population of H providers is relatively large and L providers are of relatively high quality.*

In many developing countries, high quality health care services are considered scarce resources, i.e., q is small. Additionally, when compared to high quality ones, other health care services may be very poor. i.e., $\frac{\lambda}{1+\lambda}$ is low. [34] For these countries, corollary 2.3 suggests that even if there is some increase of policyholders' welfare from H providers multi-homing, such an increase is diluted by the countries' low average qualities and low substitution rate of L providers to H providers. If one considers any additional patient costs incurred by changing the market environment,[35] such small increment can be offset easily. In all, proposition 2.5 and corollary 2.3 suggest that for developing countries,

[33] The example may imply that multi-homing of H providers cannot improve the *ex ante* policyholders' welfare, if the disease under investigation is very common or highly contagious, i.e., there are $\frac{2}{3}$ in whole population are with the higher than $\frac{8}{9}$ risk of being ill.

[34] This relative quality effect can be measured by the fixed wage ratio of a L provider to a H provider in the multi-homing case according to the above analysis, i.e., $\frac{\lambda}{1+\lambda} = \frac{W_i^L}{W_i^H}$.

[35] Such as additional transportation and waiting cost when a patient decides to see a H provider instead of a nearby L provider. Remember that a policyholder must choose one health plan and pay the premium before becoming a patient. Here the decision of see a doctor is sometimes made after that. $CS^{2,M} - CS^{2,S}$ in formula (2.16) is *ex ante*, but the additional cost is *ex post*.

policies using health plans to allocate health care resources and allowing multi-homing of H providers may not work to benefit policyholders.

Nevertheless, Proposition 2.4 suggests that policies allowing multi-homing of high quality providers are still justified because they lead to more efficient outcomes than those prohibiting it. This, however, depends on a fixed salary payment scheme for the provider. In the next section, another payment scheme, fee-for-service (FFS) is investigated and we find that multi-homing may no longer be more efficient than single-homing.

2.5 Fee-for-service Scheme and Efficiency of Single-homing

We now consider the same market structure as before except that providers receive a fee-for-service (FFS).[36] Using FFS reimbursement, each provider first pays the treatment cost, then is compensated by the health plan. As in Bardey & Rochet (2010), we denote the FFS rate as R_i^j, if the provider is j type and is affiliated with network i ($i = A, B$ and $j = H, L$). The providers' expected payment equals the product of the profit margin offered by the network (FFS R_i^j minus unit cost of treatment c) by the level of activity that the physician expects to have if he joins the plan. This expected activity level is equal to the expected number of treatments in the network, divided by the number of providers in the network, as we assume that the policyholders are randomly assigned to providers in the same plan. So the expected activity levels of providers in plan A and B are

$$\Phi_A = \frac{\int_0^{\tilde{\theta}} \theta f(\theta) d\theta}{(1-q)\widetilde{x}_A^L + q\widetilde{x}_A^H} \text{ and } \Phi_B = \frac{\int_{\tilde{\theta}}^1 \theta f(\theta) d\theta}{(1-q)(1-\widetilde{x}_B^L) + q(1-\widetilde{x}_B^H)},$$

where $\tilde{\theta}$ is the risk segmentation. The expected payment to a j provider if he join in plan i ($i = A, B$) is

$$T_i^j = (R_i^j - c)\Phi_i.$$

Here FFS introduces an inter-group competition on provider side. With more providers in the network, the expected activity level Φ_i will be less. These formulas also reveal the second indirect externality in our model, this time from policyholders to providers: Φ_i depends on the number and risk of policyholders who join plan i.

The plans' profits can be rewritten as follows:

[36] There is a large amount of literature that focus on the comparison of the payment schemes with the concern of how these schemes affect the quantity of medical care services. A flat scheme such as salary is believed to reduce more unnecessary service than FFS. In our model, the quantity is 1 for each patient (policyholder when he or she is ill) and a provider has no power to determine the amount of treatments, so there is no such incentive problem.

$$\begin{aligned}
\pi_A &= P_A F(\widetilde{\theta}) - c\int_0^{\tilde{\theta}} \theta f(\theta)d\theta - [(1-q)(R_A^L - c)\widetilde{x}_A^L + q(R_A^H - c)\ \widetilde{x}_A^H]\Phi_A \\
\pi_B &= P_B(1 - F(\tilde{\theta})) - c\int_{\tilde{\theta}}^1 \theta f(\theta)d\theta - [(1-q)(R_B^L - c)(1 - \widetilde{x}_B^L) + q(R_B^L - c)(1 - \widetilde{x}_B^H)]\Phi_B.
\end{aligned}$$

Now, health plan i $(i = A, B)$ chooses $(P_i, R_i^j, \widetilde{x}_i^j)$ to maximize its profit.

When multi-homing is allowed, health plans only need to meet providers' *individually rationality constraints*, for all $j = H, L$:

$$(IR_A^j) : (R_A^j - c)\Phi_A - t\widetilde{x}_A^j \geq 0 \text{ and } (IR_B^j) : (R_B^j - c)\Phi_B - t(1 - \widetilde{x}_B^j) \geq 0.$$

These constraints must be binding, i.e., $T_A^j = t\widetilde{x}_A^j$ and $T_B^j = t(1 - \widetilde{x}_B^j)$. Remember that under fixed salary scheme with multi-homing (and single-homing when $t > \frac{1+\lambda}{2}\widehat{\theta}$), we also have $T_A^j = W_A^j = t\widetilde{x}_A^j$ and $T_B^j = W_B^j = t(1-\widetilde{x}_B^j)$. This implies that the multi-homing equilibrium under the FFS scheme is the same as that under fixed salary scheme.[37] We present this result in proposition 2.6.

Proposition 2.6. *Under A2.1 and FFS scheme, multi-homing equilibrium is equivalent to that under fixed salary.*

Now we turn to the single-homing case. Let $\widetilde{\theta}(t)$ be the single-homing equilibrium risk segmentation. Consider t is slightly less than $\frac{1+\lambda}{2}\widehat{\theta}$,[38] The constraints IC_i^L $(i = A, B)$ remain their slackness because of the continuity of the system.[39] In this case, plans need to deal with marginal H provider's *incentive constraints*:

$$\begin{aligned}
(IC_A^H) &: (R_A^H - c)\Phi_A - t\widetilde{x}_A^H \geq (R_B^H - c)\Phi_B - t(1 - \widetilde{x}_A^H), \\
(IC_B^H) &: (R_B^H - c)\Phi_B - t(1 - \widetilde{x}_B^H) \geq (R_A^H - c)\Phi_A - t\widetilde{x}_B^H.
\end{aligned}$$

The equilibrium *risk segmentation* $\widetilde{\theta} = \widehat{\theta}$ may no longer hold.

Intuitively, single-homing *risk segmentation* under an FFS scheme should shift towards the left as transportation cost t decreases. That is, plan B expands on the policy-holder side with t deceasing. As in the previous section, with smaller t, plan B has incentive to hire more H providers to be close to its local monopoly status, $1 - \widetilde{x}_B^H = \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1-F(\widetilde{\theta}(t)))$. To encourage more H providers to join in plan B, this plan can

[37] Each plan faces the same problem with the same constraint.

[38] As in previous section, only when $t \leq \frac{1+\lambda}{2}\widehat{\theta}$, does single-homing equilibrium will take a different form than multi-homing.

[39] Note that when $t = \frac{1+\lambda}{2}\widehat{\theta}$, we have $1 - \widetilde{x}_B^L + \widetilde{x}_A^L = \frac{\lambda}{2t}\widehat{\theta} < 1$. Because of the continuity of the system, when t is slightly smaller than $\frac{1+\lambda}{2}\widehat{\theta}$, formula (2.11): $1 - \widetilde{x}_B^L + \widetilde{x}_A^L < 1$ holds.

1) increase its fee-for-service rate of a H provider R_B^H, 2) downsize its enrollment of L provider $1-\widetilde{x}_B^L$, 3) expand its share on policyholder side $(1-F(\widetilde{\theta}))$, or do a combination of them. Here, method 2 and 3 increase the expected activity level Φ_B. The method of 1) increasing R_A^H and 2) downsizing $\widetilde{x}_A^L$ can be easily adopted by plan A. However, plan B always offers better service than A ($n_A\,[\lambda+q_A] < n_B\,[\lambda+q_B]$). This gives plan B relative advantage in the competition on the policyholder side. Using method 3 of increasing $F(\widetilde{\theta})$ is difficult for plan A and plan B will increase $(1-F(\widetilde{\theta}))$ as the most effective way to recruit H providers. On the other hand, plan B increases its number of H provider while also increasing its advantage to the policyholder, which will in turn makes itself more attractive for H providers. Generally, we should expect that with single-homing, the better plan B will expand on policyholder side (i.e., $\widetilde{\theta} < \widehat{\theta}$) when transportation cost t is smaller than $\frac{1+\lambda}{2}\widehat{\theta}$.

There are multiple equilibria (for H provider configuration) when t is slightly less than $\frac{1+\lambda}{2}\widehat{\theta}$. We present one in the following proposition which leads to the largest social aggregate externality among all equilibria. The detailed proof of the proposition can be found in appendix.

Proposition 2.7. *Under A2.1 and FFS scheme, if t is slightly less than $\frac{1+\lambda}{2}\widehat{\theta}$, there exists a single-homing equilibrium where*

***a**. risk segmentation* $\widetilde{\theta}^S(t) < \widehat{\theta} = \frac{1-2F(\widehat{\theta})}{f(\widehat{\theta})}$;

***b**. L providers' configuration is* $\widetilde{x}_A^{L,S} = \frac{\lambda}{2t}\widetilde{\theta}(t)F(\widetilde{\theta}(t))$ *and* $1-\widetilde{x}_B^{L,S} = \frac{\lambda}{2t}\widetilde{\theta}(t)(1-F(\widetilde{\theta}(t)))$;

***c**. H providers' configuration is* $\widetilde{x}_A^{H,S} = 1-\frac{1+\lambda}{2t}\widetilde{\theta}(t)(1-F(\widetilde{\theta}(t)))$ *and* $1-\widetilde{x}_B^{H,S} = \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1-F(\widetilde{\theta}(t)))$.

In proposition 2.7, plan B is a local monopoly for providers with the largest enrollment of H providers among all possible equilibria. Thus this equilibrium is with the largest quality of plan B since

$$q_B = \frac{n_B^H}{n_B} = \frac{q(1-\widetilde{x}_B^{H,S})}{(1-q)(1-\widetilde{x}_B^{L,S})+q(1-\widetilde{x}_B^{H,S})}.$$

Because of $\widetilde{\theta}^S < \widehat{\theta}$ and $F(\widehat{\theta}) < \frac{1}{2}$, we have $\int_{\tilde{\theta}^S}^1 \theta f(\theta)d\theta > \int_0^{\tilde{\theta}^S} \theta f(\theta)d\theta$. Then this equilibrium leads to the largest socially aggregate externality

$$(n_A(\lambda+q_A))\int_0^{\tilde{\theta}^S} \theta f(\theta)d\theta + (n_B(\lambda+q_B))\int_{\tilde{\theta}^S}^1 \theta f(\theta)d\theta,$$

for any $\tilde{\theta}^S$, because the marginal increase of aggregate externality with n_B and q_B is higher than that with n_A and q_A.

Note that both $\widetilde{x}_A^L = \frac{\lambda}{2t}\widetilde{\theta}F(\widetilde{\theta})$ and $\widetilde{x}_A^H = \frac{\lambda}{2t}\widetilde{\theta}(1-F(\widetilde{\theta}))$ increase with $\widetilde{\theta}$ and $\widetilde{\theta}^S < \widehat{\theta}$. Under FFS scheme, single-homing of H providers leads to less L provider enrollment in both plans than multi-homing. Nevertheless, compared to L providers, multi-homing expands both plans' sizes of H provider with expansion rate $\frac{dn^H}{dn^L} = (\frac{\lambda}{1+\lambda})^{-1} > 1$. So multi-homing leads to higher quality than single-homing.

Corollary 2.4 *Under A2.1 and FFS scheme, the plans' quality is higher with multi-homing than single-homing. However, when compared to single-homing, multi-homing leads to more L provider enrollment in both plans, thus none offer a higher than* $\frac{(1+\lambda)y}{\lambda+y}$ *quality level to policyholders.*

As stated above, if single-homing is mandatory, health plans may attempt to take more H providers on board by promising them more potential treatments, which may lead to a reduction in the number of L providers in each plan. If H providers multi-home, it directly weakens the plans' competition over them, thus the plans may then increase the number of L providers to lower H providers' payment. Furthermore, only contracting more H providers may not be as profitable for each plan as enrolling more L and H providers, due to the transportation cost. However, with a fixed salary, such a negative intergroup effect among providers vanishes because providers are swayed by their expected activity. So multi-homing of H providers will not change the L providers' configuration. In all, even though services from H providers and that from L providers are substitutable for policyholders, health plans will not exclude any of their L providers to give policyholders a very high *quality level* with multi-homing which increases H providers' enrollment.

Now let us see why the single-homing equilibrium given by proposition 7 may be more efficient than the multi-homing equilibrium. Here we continue to adopt the utilitarian social welfare function (see formula (2.14)). From proposition 2.7, when t is slightly less than $\frac{1+\lambda}{2}\widehat{\theta}$, equilibrium risk segmentation with single-homing $\widetilde{\theta}^S(t)$ is strictly less than that with multi-homing $\widehat{\theta}$. Single-homing may lead to a more efficient market structure than multi-homing on policyholder side, because policyholders with $\theta \in [\widetilde{\theta}^S(t), \widehat{\theta})$ who would choose plan A if multi-homing is allowed, now turn to plan B when providers single-home. When the environment shifts from single-homing to multi-homing, the loss of externality enjoyed by these policyholders is

$$\theta\left[\lambda\widetilde{x}_B^{L,S} + (1+\lambda)\widetilde{x}_B^{H,S}\right] - \theta\left[\lambda\widetilde{x}_A^{L,M} + (1+\lambda)\widetilde{x}_A^{H,M}\right] \approx \theta\frac{\lambda^2 + (1+\lambda)^2}{2t}\widehat{\theta}(1-2F(\widehat{\theta})).$$

The average policyholders' loss is not negligible if $\widehat{\theta}$ and $1-2F(\widehat{\theta})$ is relatively large. Moreover, we have the following proposition with a detailed proof in the Appendix.

Proposition 2.8. *Under A2.1 and FFS scheme, for some t less than $\frac{1+\lambda}{2}\widehat{\theta}$, multi-homing of H providers yields strictly lower social welfare than the single-homing equilibrium given in proposition 2.7, if*

$$\widehat{\theta} \geq \max\{2E(\theta|\theta \leq \widehat{\theta}), \frac{(1-F(\widehat{\theta}))F(\widehat{\theta})}{(1-2F(\widehat{\theta}))^2}E(\theta|\theta \geq \widehat{\theta})\}.$$

Note that $\widehat{\theta} < E(\theta|\theta \geq \widehat{\theta})$, therefore, according to proposition 2.8, $\frac{(1-F(\widehat{\theta}))F(\widehat{\theta})}{(1-2F(\widehat{\theta}))^2} < 1$ must hold, or $F(\widehat{\theta}) < \frac{\sqrt{37}-5}{6} \approx 0.1805$. This proposition suggests that when providers are remunerated by FFS, the single-homing outcome given by proposition 2.7 may be more efficient than the multi-homing equilibrium, if plan A only gets a relatively small share on the policyholder side (with single-homing or multi-homing). Under this circumstance, plan A offers relatively small externality to policyholders, i.e.,

$$\begin{aligned} n_A\left[\lambda + q_A\right] &= \frac{\lambda^2}{2t}\widetilde{\theta}(t)F(\widetilde{\theta}(t)) + (1+\lambda)(1 - \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1-F(\widetilde{\theta}(t)))) \\ &< \frac{\lambda^2 + (1+\lambda)^2}{2t}\widehat{\theta}F(\widehat{\theta}) < 0.1805\frac{\lambda^2 + (1+\lambda)^2}{2t}\widehat{\theta}. \end{aligned}$$

For a policy maker, this plan should be designed as a complementary health plan to plan B and is only suitable for a few low-risk policyholders ($\theta < \widetilde{\theta}(t)$ and $F(\widetilde{\theta}(t))$ is small) who are unlikely to get sick. However, by allowing multi-homing, this plan can subscribe more high quality providers to increase its *quality level* q_A and providers' size n_A, then attract high-risk policyholders ($\theta \in [\widetilde{\theta}^S(t),\widehat{\theta})$) to join in. Thus, to achieve the maximum social welfare, policies using an FFS scheme on provider side and prohibiting providers to multi-home might be the solution.

2.6 Conclusion

In this paper, we study the multi-homing of high quality providers (physicians) in a two-sided health market in this paper. We modify the Bardey and Rochet (2010) framework by introducing a quality differentiation among providers: H as high quality and L as low. The policyholders are potential patients who differ in their risk of being sick θ. We focus on the equilibrium where the *risk segmentation* $\widetilde{\theta}$ exists such that all policyholders with $\theta < \widetilde{\theta}$ go to one plan (plan A) and those with $\theta \geq \widetilde{\theta}$ go to the other (plan B). Besides the quality differentiation, providers also have different strategies for choosing plans and are assumed to locate on a Hotelling line with two plans at the endpoints. A provider prefers to single-home in one plan only if he locates too far from the second, i.e., his location is close to either endpoint of the Hotelling line. A provider has incentive to multi-home only when his additional transportation costs of joining in a second health plan can be covered by the payment given by that plan (he locates in the middle part of the Hotelling line).

We define the *quality level* of a health plan as the percentage of H providers in that plan. We find that when providers are allowed to multi-home, both plans are more profitable than when providers are constrained to single-home, because multi-homing leads to less competition between these two plans. Allowing multi-homing also works in line with providers' interests, since it gives them more freedom. Because more H providers enroll in both plans under multi-homing, it increases the quality externality of each plan.

However, proposition 2.5 suggests that multi-homing of providers does not guarantee an increase in policyholder's surplus. In fact, there is an outcome where all policyholders in one health plan are harmed by multi-homing. Corollary 2.3 further suggests that if the high quality providers are scarce resources or the performance of low quality providers are relatively poor, multi-homing cannot improve policyholders' welfare significantly. These two results imply that multi-homing of high quality providers in health markets may not benefit the consumers.

The efficiency of multi-homing relies on the payment scheme on the provider side. Under a fixed salary scheme, multi-homing under a fixed salary scheme improves efficiency compared to single-homing, because it increases the externality enjoyed by every policyholder. However, when providers are compensated with fee-for-service (FFS), the social welfare may be decreased by multi-homing under FFS.

Another interesting question is how multi-homing of H providers impacts L providers. First, we assume that services from L providers can be substituted by those from H providers. Second, multi-homing leads to a larger amount of H providers in each health plan. Therefore, one may expect that multi-homing of H providers limits the subscribing number of L providers compared to single-homing. On the contrary, compared to single-homing, multi-homing of H providers never decreases L provider enrollment in all outcomes of this paper. This is due to transportation costs. For each plan, the strategy of subscribing an L provider locating close to the plan may take precedence over the hiring an H provider who locates further. As multi-homing leads to less competition intensity than single-homing, plans' incentive to cut the cost of hiring L providers and enhancing the payment of H providers is weaker with multi-homing than single-homing.

Chapter 3

Regulation on Food Quality: Process Certification or Product Inspection

3.1 Introduction

Effective and credible food quality regulatory systems are critically important in public policies. Consumers focus on public signals related to the intrinsic quality of food, such as safety and nutrition attributes.[1] Food safety has been increasingly suspected in recent years after outbreaks of food-borne diseases (e.g., E. Salmonella in the US and Japan) and human transmissions of zoonotic diseases (e.g., the mad cow disease in the UK). Developing countries also face this issue: in China, the Sanlu Scandal (2008) erupted when thousands of Chinese infants died because of poisoned milk. These events suggest a remarkable informational asymmetry between producers and consumers: food companies often know more about what they are selling than consumers. Regulations may serve to address this problem and provide consumers with quality signals they can trust.

The old-fashioned approaches to food quality regulation is *product inspection* (PI). Government regulators sample and test foods before they are sold. Findings of low quality foods are often followed by product recalls, which directly prevent the buyers from consuming low quality food. Nowadays, many governments are taking a new approach to ensure the quality of the food supply: *process certification* (PC). Through PC, regulatory systems focus on verifiable control of the manufacturing process. Only when the food

[1] In developed countries, extrinsic food quality, including animal welfare and environmental preservation is also a concern.

companies' production processes fulfill certain standards, are they allowed to sell their products in specific markets. For instance, food process control can be based on the label of Hazard Analysis Critical Control Point (HACCP), which is widely recognized as a regulatory approach in many developed countries (see Grijspaardt & Vink (1995), Morris (1997) and Peters (1997)).[2] The enthusiasm for *PC* also spread to developing countries where food safety requirements have been progressively put into place. Between 2001 and 2008, a label called National Inspection Exemption was granted by China's authority. Food companies, whose processes were certified by HACCP or ISO9000,[3] could easily be labeled and exempted from any final product testing.

This paper attempts to formally analyze the rationale for the above change in food quality regulations. Monitoring costs are the main reason that *PC* took the place of *PI*. First, *PI*s are often conducted in advanced and expensive laboratories to ensure the accurate testing results. *PC* does not incur this cost. Consider checking the safety of soup. Given other conditions are equal, *PI* may require counting bacteria under a microscope while *PC* may only require monitoring if the chef wears face mask covering his/her nose and mouth. Thus *PC* is less expensive.[4] Secondly, the detection of low quality products often fail to provide evidence of business fraud.[5] So it may not be strong enough to support the implementation of serious punishments. To provide incentive for food companies to improve quality, governments have to constantly inspect the products.[6] The cost for the old-fashion approach to be effective could be even larger. Third, *PC* standardizes the way firms organize their manufacturing, thus these firms are responsible for most of the costs. *PI,* however is usually a burden for governments, who use public funds and generate a welfare distortion. Forth, it is the production process that mainly determines the quality of products. *PI* cannot directly enhance the quality provision. Expenditures for sampling and testing may be superfluous. In fact, a number

[2] The European Union Directive 93/43, effective in December 1995, requires food companies to implement HACCP. In the United States, HACCP was mandated through regulation for seafood in 1994, for meat and poultry in 1996, and proposed for fresh fruit juice in 1998, with regulations for other food industries in the beginning of 21st century. Australia, New Zealand, and Canada also have mandatory (or voluntary) public programmes to encourage adoption of HACCP.

[3] ISO 9000 certification is issued by International Organization for Standardization, which ensures a consistent production process related to quality provision.

[4] For instance, HACCP, as one kind of *PC*, is widely recognized in the food industry as an effective approach to establish good manufacturing practices for the production of safe food. It is achieved through the identification of points in the production process that are most critical to monitor and control. It also establishes effective approaches to keep production records. With HACCP-based process certification, once it has been verified that effective food safety systems are in place, monitoring will rely largely on audits of production records. Hence, the cost of monitoring is more than likely to be lower than product inspection.

[5] In Unnevehr and Jensen (1999), the authors pointed that the producers may not know what they are selling. For example, bacteria, which are living organisms, may not be detectable when the foods are just produced. In cases, the food companies claimed that there were inevitable delays between initial occurrence of incidents and the detection of incidents.

[6] According to a report of the first US National Conference on Food Protection (CFP) in 1995, there was no indication of any decline in the incidence of food borne diseases by adopting modest product inspection.

of analysis has demonstrated that PC is more cost-effective than testing products and then destroying or reworking on them in numerous cases (see for instance, Unnevehr and Jensen (1996), Crutchfield et al., (1997) and Roberts et al., (1996)). Following this idea, the issue is evaluated based on that PI is a more costly regulatory approach than PC.

A variation of the Cournot model is considered, to which I introduce contingent valuations measuring consumers' willingness to pay for a specific quality attribute. Conventionally, two kinds of food (L and H) are presented to consumers, where L is low quality and H is high. The credence feature of food quality is main focus of this paper. Most foods can be categorized as credence goods (see Caswell and Mojduszka (1996)): buyers cannot possibly know the food quality even after consumption. Thus, food quality (H or L) is defined by performance standards and ascertained only through PI. However, it is food production methods that determine the quality provision: the probability that products are H (L). Through PC, a regulator standardizes the manufacturing process to enhance the percentage of H foods in average. It is assumed that there exists the safest process with minimum probability of products being L (but not zero). This safest process should be chosen as the process standard whenever PC being adopted. Intuitively, the efficiency of credence goods markets largely relies on the credibility of quality signal. Regulations may be optimal if the public policy mandates the adoption of the safest process. Food quality is crucial, so is the quantity. The shortage of food could be regarded as a more serious issue than unsafe food. Moreover, besides providing effective signals of quality, a regulator should also avoid strict and inflexible mechanisms with little marginal benefit for consumers. Therefore, the quantity in the Cournot setting is also important as a measurement of competitiveness.

Whether or not to stop PI is a current area of controversy surrounding changes to the regulatory system. Note that the safest process cannot guarantee high quality, so using PI after the monitoring process further decreases the probability of low quality products arriving consumers' tables. Accompanied with PC, PI still benefits consumers through increasing the credibility of the market. For example, the HACCP systems in the American meat industry also include testing for food-borne pathogens. After Sanlu scandal, China is trying to establish new regulations for the diary industry that combine PC and PI.

However, this paper suggests that PC may only substitute for, but not be supplemented by PI in optimal regulation. The reason behind this result is PI impacts the outcomes not only on market credibility, but also on firms' quantities. Knowing PI is being adopted, all firm expect that some of their rivals may recall their products, leading firms to increase quantities as they prepare to take over their competitors' market share.[7] With imperfect quantity competition, the total supply is distorted to less than efficient levels. Thus, enhancing quantity is also a positive factor for social welfare. The influence on market credibility and firms' quantities timing together causes a social benefit of PI that is convex in its frequency: the marginal benefit increases as PI becomes more frequent. That is, if the expenditure on PI is worthy for improving efficiency, it should be conducted as frequent as possible. Depending on PI's cost, it is optimal to adopt PI or to never implement it at all. Obviously, "never" implies the case where PI is optimally stopped due to its relatively large cost. But if the cost is small, PI should be optimally conducted so often that firms voluntarily adopt the safest process, regardless of any process standard. In such cases, PC becomes unnecessary. Consequently, combining PC and PI may not generate more cost-effective mechanisms. Moreover, the advantage of PC may become apparent when the competition is intensive and the quantity distortion is insignificant. Since consumers cannot prominently benefit from total supply enhancement by PI, PI may become less and less desirable when quantity competition is strengthened.

Despite the relatively small monitoring cost, PC requires set-up costs for adding technologies to make process control verifiable. These costs are usually borne by the food companies. Again consider the example of checking safety of soup. It is almost costless to monitor whether the chef wears face mask if the cooking process is perfectly recorded. But effective recording systems may require cameras and storage devices for digital film which are expensive. The scale of set-up costs also depends on whether or not firms must modify their processes to meet the process standard.[8] High set-up cost may exclude some companies from business (see MacDonald & Crutchfield (1996)). As mentioned above, PC is likely a better option for governments than PI with strong competition. However, set-up costs for PC may not be overcome by some firms. Then mandating process standards may downsize market competitiveness, as well as PC's advantage as a regulatory approach. Then a paradox may arise: PC is a more cost-effective approach than PI before PC's implementation; but after that, it is not.

The rest of this chapter is organized as follows. Section 3.2 is the literature review and section 3.3 is the model set-up. Section 3.4 studies a benchmark case, where the informational asymmetry between consumers and food companies can be solved without

[7]The effect of PI on firms' quantities also depends on the correlation between firms' qualities. In this chapter, I assume that the quality is independent among firms. When the quality is positively correlated, firms may decrease their quantities in responding to the PI. It will be briefly discussed later in this chapter.

[8]Specifically, Colatore (1998) investigated the actual costs of HACCP adoption among breaded fish producers in the state of Massachusetts. The total first-year costs of adoption averaged \$116000 per firm and the cost of meeting minimum requirements averaged \$34000. These two figures are set-up costs. The monitoring and certifying expenditure is considerably less than \$34000.

any regulation. In section 3.5, results for credence goods is presented. Section 3.6 is the conclusion and further developments.

3.2 Literature Review

In the economic literature, it is common to distinguish qualities according to the extent to which they can be identified by the consumers. The quality of search attributes, such as color and size, is easily detected before consumption. For experience attributes, like taste and suitability, consumers are able to discern its intrinsic quality after consumption. Finally, for credence (or trust) goods, quality can never be known by consumers with certainty (see Nelson (1970)). Markets can provide quality assurance for search and experience goods, which is relatively effective and robust enough. In the case of credence goods, however, the informational asymmetry between consumers and sellers is not easily remedied by markets. Thus, suboptimal equilibria exist and are more common (see Darby and Karni (1973)). Theoretical contributions on credence goods have emerged where they have discussed whether or not market mechanisms can prohibit business fraud and solve the informational asymmetry. The markets for experts (e.g., lawyers, medical doctors, auto mechanics service-persons) got a lot of attention. For instance, Wolinsky (1993 and 1995) showed that to discipline experts, customers may intentionally search for multiple opinions and that an expert's reputation is also essential for the customers' decision. Then Emons (1997 and 2001) showed that if consumers have enough information about market data, they are able to infer the expert's incentives. The market equilibrium resulting in non-fraudulent behavior does exist. In other cases, however, there is no trade because consumers anticipate fraudulent behavior. While many other papers assumed fraud is costless, Alger & Salanee (2006) allowed for fraud costs, which aggravate the inefficiency of fraudulent behavior. They suggested that when the cost is small relative to the economies of scope, the experts may always fraud: they pool the information and price the diagnosis high enough, in order to deter consumers seeking for second opinions. Reputation may also provide a solution. These important papers highlight the difficulties in achieving an efficient market for credence goods. Thus, certification signals might be a good thing in this area.

The credibility of certification is also a common topic discussed in the literature, such as Biglaiser (1993), Albano & Lizzeri (1997) and Lizzeri (1999). Because certification requires the involvement of an independent third party, they treate certification agents as intermediaries between producers and consumers in the process of quality provision. These papers investigate whether or not the certification intermediaries may strategically manipulate the accuracy of their provided signals. When competition among the intermediaries is intense, they may specify each producers' quality and thus lead to full information revelation. But a monopoly intermediary might not fully disclose the information about producers' quality. In order to certify producers with relatively low quality, they would rather provide noisy signals. These noisy signals still increase efficiency compared to situations without signals. On the contrary, when the reputations of certification intermediaries are considered, as suggested by Strausz (2005), credibility is easier to sustain when there is only one certifier. Such reputation models hinge on a trade–off between the short-run gain from concealing information and the long-run gain for honesty. When the monopoly certifier expects a larger future demand, the long-run gain rises and noisy signal becomes less attractive. In this sense, the governments are more reliable than private certifiers in providing credible quality signals.

Many papers are inspired by food quality policies and focus on their welfare impacts. For example, Zago & Pick (2002) considered the intervention of the European Union in agricultural products with specific qualities related to a production area or technology. They concluded that the impact may be detrimental. Producers specializing in low quality food may be worse-off, due to their geographical and technological conditions. The impact on consumers' welfare is also ambiguous and depends on the market power. However, their results are limited by only considering organoleptic characteristics and are not very constructive for food regulations with safety concerns. Auriol & Schilizzi (2003) pointed out the importance of certification costs for market structure. They compared welfare levels resulting from public and private certification programs and found that the higher the cost of certification, the higher the need for public intervention. Their analysis was restricted to the assumption of perfect signals transforming credence goods to search ones, and therefore does not consider inaccurate *process certification* which is now a widespread regulatory approach. The degree of credibility in the public signal has also inspired studies, such as Annania & Nistico (2004). Their results suggested that high quality producers may be forced to accept regulations with imperfect signals and share the market with low quality producers. There are other papers on similar topics, such as Caswell & Mojduszka (2000), Giannakas (2002), Giannakas & Fulton (2002), etc.. However, none of these papers have studied the controversy of stopping the traditional regulatory approach (*product inspection*), or discussed the effectiveness of *process certification* with respect to the competitiveness. The present paper aims to complement this literature through formally analyzing this issue.

3.3 The Model

I consider a variation of Cournot competition in one specific food market, where products' quality may be differentiated. The quality is either H (high quality) or L (low quality). And the *ex ante* probability of H depends on the production processes of firms (or producers, sellers). In practice, one unit of food is H if it fulfills some performance standards;[9] it is L if otherwise.

There are N ($N \geq 2$) identical firms and mass 1 of consumers; all are assumed to be *risk neutral*. Consumers can be introduced as being risk averse in a food quality context (see, for instance, Spence (1977)). With risk aversion, there are two potential distortions, one from risk reduction (the incentive for firms to produce more H commodities) and the other from risk allocation (the possibility that L products will be consumed). To achieve efficiency, policy instruments are needed to create the correct incentive for risk reduction and to ensure efficient risk allocation. However, consumers' awareness of quality (through quality signals lack thereof) may provide efficient risk allocation. The incentive for risk reduction can be modeled as through parties are risk neutral (see Miceli & Segerson (1995)).

Consumers may not observe the quality and public signals may be needed. Two approaches for signaling, *product inspection* (PI) and *process certification* (PC), is available for the regulator (she), who maximizes a utilitarian social welfare function. By PI, she samples and checks whether the quality meets the requirements of the performance standards. Through PC, she does not only monitor firms' production processes, but also sets some process standard that the firms' production has to achieve. Since the standards should be announced before firms choose their processes, the regulator is the first one to move in this model and firms are the second, then market price equalizes the consumers' demand and the supply. The rest of this section details the model setting.

3.3.1 Quality and Process

The most common quarantine rule in food business is: if one unit of food is proved to be unsafe or unqualified, then all other food produced in the same factory (or farm) and during the same production period, should be assumed unsafe or unqualified and be recalled or disposed of immediately. The contamination of food, by either chemicals or bacterium, is likely to be spread among all products sharing the same manufacture (or delivery) process. To address the concerns behind the quarantine rule, I assume that

[9] The performance standards are assumed as granted in this paper. Therefore H and L are well defined.

one firm's whole production shares the same quality,[10] which is assumption 1:

A3.1. If some products are H (L), the other products from the same firm are also H (L).

Additionally, quality is assumed to be independent among firms, because commodities from different firms do not share the same manufacturing procedures.

The probability of food quality being H solely depends on the process.[11] As a result, processes can be characterized by the probability of being high quality, which is denoted as $x = Pr(H)$ (and $1 - x = Pr(L)$). I determine that a process x is *safer* (or better qualified) than x', if $x > x'$. The production cost increases with safety. Provided with process x, a firm must incur cx (the marginal cost) to produce one additional unit of food. The quality is random, there is no process that can make H quality certain. Let $\overline{x}$ be the maximum probability of H products provision among all production processes and assume that

A3.2. For any process x, its marginal cost is cx and it satisfies $x \leq \overline{x} < 1$.

Under assumption 3.2, $\overline{x}$ can be referred to as the *safest process* (or best qualified process). Note that the cost for quality is linear on the process under *A3.2.* This assumption will prove useful for the adoption of safest process in any equilibrium. It seems quite restrictive when compared to a general setting allowing for strictly convex technology. However, strict convexity would add distortions into the analysis; it is possible that firms may overproduce the quality with respect to the social optimal level.[12] In the context of food quality (especially safety), policies clearly reducing quality should not exist: "the safer the better" is common sense for consumers. In this sense, the linear cost for the processes could be more realistic.

To adopt some specific production processes, firms should also pay for the start-up costs. In the long-run, however, only the marginal costs matter. Nevertheless, the effect of set-up costs for processes (and for PC) will be also considered in this paper.

[10] I only consider a one-period game, that is all food from one firm being produced at the same time.

[11] Sometimes, process standards also define production capacities. Production that exceeds capacity is considered to lower the quality provision. For simplicity, quality is assumed independent with quantity in this paper.

[12] Consider the case of quality being observable. The consumers will not be interested by the safety level of firms' processes. The only thing mattering to them is the quantity, as Cournot quantity competition is assumed. Since I assume that the quality is either H or L, the firms emphasize their processes to increase their chances to sell their products and recover their costs. Then the optimal process choice regarding the social welfare and the firms' profit may not be the same. The firms may overly invest in process comparing to that required by the social optimality.

3.3.2 Public Signals

To simplify the discussion, public interventions on the quality provision are focused: there is not any quantity or price regulation. Particular attention is paid to PC and PI:

Process Certification:

Adopting PC, the regulator first sets a process standard $\widetilde{x}$. Then she monitors the firms during their production and is informed about their exact processes. Only the firms choosing processes safer than $\widetilde{x}$ are qualified to enter the H market. Under *A3.2*, the standard $\widetilde{x} \leq \overline{x} < 1$ is an imperfect signal of quality: the proportion of H products in the whole production is equal to or larger than $\widetilde{x}$, but never exceeds $\overline{x}$. In the case of PC being abandoned by the regulator, the process standard is $\widetilde{x} = 0$.

As stressed in the introduction, monitoring firms' processes incurs much lower costs than sampling and inspecting their products for the regulator. Here the cost of PC is assumed to be small enough, thus can be ignored; while the cost of PI is larger.[13] Moreover, PC is free for the firms, since it is a mandatory public policy.[14]

Product Inspection:

PI can accurately identify the quality of each sampled product. The unit cost of PI is $K > 0$, which is the cost for sampling a negligible amount of food and testing the quality. Under *A3.1*, PI provides sufficient evidence to determine the quality of the inspected firm. Assume that before the firms' move, the regulator announces a common probability $0 \leq r \leq 1$ with commitment that each firm's products will be inspected. Then the expected cost of PI is rNK, which is assumed to be paid by the regulator.[15] If PI is not adopted, the regulator sets $r = 0$.

PI can be a public signal because it prevents some L products from being purchased as H products. Note that it is impossible to sample and test foods after they are consumed. It is also not optimal to conduct PI before firms put L products into the H market. So in this setting, there is a specific stage for PI intervention, which follows the firms' move and is followed by consumption. When PI finds L products in the H market, the regulator forces the inspected firms to call back and dispose their foods with 0 cost.[16]

Moreover, I assume that PI occurs independently among firms. Consider that there are firms who are not inspected, i.e., $r < 1$. Since passing PI perfectly signals the

[13] The monitoring cost of PC is normalized almost to 0 in this paper for simplicity. In fact, the main result in this paper remains true, as long as PI is much more costly than PC.

[14] Sometimes, the firms are charged by a fixed fee or per unit fee (see Auriol & Schilizzi (2003) and Crespi and Marrete (2001)). Because marginal PC costs almost nothing, there is no need to collect a fee from firms.

[15] If one considers the public funding cost, this cost is $(1+\lambda)rNK$, where λ is Laffont-Tirole parameter.

[16] If the disposing cost is not zero, it can be accounted into the inspection cost.

high quality, releasing such information may cause the market pricing differently. To circumvent this problem, I also assume that neither the consumers nor the firms can know about whose products have passed *PI*.

3.3.3 Fraud and Liability

The N ($N \geq 2$) identical firms simultaneously decide their processes and quantities before production. Let x_i and q_i ($i = 1, .., N$) denote firm i's process and quantity, where $x_i \leq \overline{x}$. If firm i wants to enter the H market, its process must be safer than the process standard set by the regulator, i.e., $x_i \geq \widetilde{x}$. Remember that cx_i is the marginal cost for process x_i under *A3.2*. Firm i's total cost is cx_iq_i. Meanwhile, each firm also decides whether it will commit business fraud.

Fraud

After production, firms are able to detect their own quality, and they have already decided whether or not to commit fraud if their quality is L. There are two strategies for dealing with L products: honesty I_h, i.e., a firm will dispose of them, and fraud I_f, i.e., a firm *attempts* to sell them as H products.

Liability

In many cases, such as bacterial contamination, firms may not be sure of the quality.[17] Thus, the strategy of I_f may not be applicable. To encompass these circumstances, I assume that all the firms are protected by an *endogenous limited liability*: punishment cannot exceed the benefit a firm's gain from fraud. Consider that firm i decides to fraud (i.e., $I_i = I_f$ ($i = 1, .., N$)):

i) PI caught the fraud. Remember that *PI* is conducted *before* consumers purchase the food. Even if the maximum punishment is imposed, the penalty is 0 for this attempted fraud, since there has been no benefit to firm i from using I_f . This creates an equivalent situation, where firms are exempted from any liability due to their inaccurate knowledge of their own quality.

Sometimes, shirking ($I_i = I_f$) detected by *PI* is affiliated with *exogenous* punishments. Theoretically, as long as the *exogenous* punishments are relatively small with respect to the consumers' benefit from purchasing H products and the *PI* cost K, a utilitarian regulator has no intention of collecting the punishment. She will only use them to reduce the *PI*'s frequency which induces firms' non fraudulent behavior, i.e., $I_i = I_h$ (see Baron

[17] It is difficult to prove the existence of fraudulent behavior.

and Myerson (1982)). Such small punishments cannot change the main result of this paper quantitatively.

ii) When fraud is caught *after* purchasing, firm i faces a penalty amounting to pq_i, where p is the price in the H market. Each consumer, who buys a L product from firm i, will get full reimbursement.

In countries like China, deliberately and successfully distributing L products (e.g., unsafe food) is punishable by death, i.e., there is no limited liability. However, both PC and PI are not instruments regulating the quality *ex post* the consumption. Hardly any large or extreme punitive rules for fraud could be enforced to strengthen their effectiveness. Because these two approaches are only preventive, the detection of bad performance through them cannot be used to punish the firms based on hypothesis, what if L products were consumed. Moreover, when the food is search goods, fraud does not happen because of the consumers' awareness of L products. When it is credence goods, fraud cannot be detectable after consumption, since consumers can never know its quality. Only in the scenario of experience goods may fraud exist and can be revealed. And the full reimbursement, like a customer product warranty, can make $I_i = I_f$ disappear in any equilibrium of experience goods: since there is no positive gain from fraud, firms do not commit it.

3.3.4 The Demand

Consumers acquire 0 or 1 unit of food. Consuming L products results in a 0 utility for all consumers, so they only purchase in the H market.[18] Consumers' utility for one unit of H product is represented by a parameter $\theta \in [0, \overline{\theta}]$, which is uniformly distributed over the population of consumers.[19] Moreover, $\overline{\theta} > c$ (the unit cost of H product) is assumed, since there should be consumers willing to pay for high quality.

If consumers can detect the quality before the consumption (e.g., the search goods case), only H products will be purchased. For any given price p in the H market, only consumers with $\theta \geq p$ are willing to buy them. The demand function is

$$D(p) = \frac{\overline{\theta} - p}{\overline{\theta}}.$$

In the case where consumers cannot know the quality before consumption, (e.g., the experience and credence goods cases), the demand relies on the proportion of H product

[18] Conventionally, consumers who do not purchase food in the H market can find substitution products in other markets. Or the L market's price is 0.

[19] The lowest value of θ is 0, because there is some consumer indifferent between L and H. It is standard in modeling food quality. The distribution of θ is assumed to be uniform for simplicity.

in the market. Let b denote the expected percentage of quality H, which depends on the firms' fraud strategies, processes and quantities, as well as public intervention. In this paper, I will focus on the *symmetric* equilibrium where all firms choose the same strategy (q, x, I_j) $(j = h, f)$. So b can be denoted as $b(r, \widetilde{x}, x, I_j)$, i.e., b does not change with q.[20] However, consumers cannot observe firms' strategies. Their belief of b, represented by $\widetilde{b}$, fully relies on the signal, i.e., $\widetilde{b} = \widetilde{b}(r, \widetilde{x})$. $\widetilde{b}$ reflects the *credibility* of the H market and its equilibrium value $\widetilde{b}^*$ equals to the "true" credibility b, i.e., firms' strategies are predictably provided with the public signal.

For any given price p in the H market, the risk neutral consumers who purchase are those who expect higher utility than the price i.e., $\widetilde{b}\theta \geq p$ or $\theta \geq p/\widetilde{b}$. The demand function is

$$D(p) = \frac{1}{\overline{\theta}}(\overline{\theta} - \frac{p}{\widetilde{b}}) = \frac{\overline{\theta}\widetilde{b} - p}{\overline{\theta}\widetilde{b}}.$$

Let Q denote the supply (the total quantity in the H market). The inverse demand function is $p(Q, \widetilde{b}) = \overline{\theta}\widetilde{b}(1 - Q)$. Thus for given $(Q, b, \widetilde{b})$, the consumers' surplus is defined as

$$\begin{aligned} CS(Q, b, \widetilde{b}) &= b \int_{\overline{\theta}(1-Q)}^{\overline{\theta}} \frac{\theta}{\overline{\theta}} d\theta - p(Q, \widetilde{b})Q \\ &= \overline{\theta}Q[b(1 - \frac{Q}{2}) - \widetilde{b}(1 - Q)], \end{aligned} \tag{3.1}$$

where the market's credibility b timing $\int_{\overline{\theta}(1-Q)}^{\overline{\theta}} \frac{\theta}{\overline{\theta}} d\theta$ is the "true" aggregate utility of consumers and $p(Q, \widetilde{b})Q$ is their total expenditure.

3.3.5 Timing

This game has 4 stages. The timing is summarized as follows:

The 1st stage (policy stage): regulator chooses $(r, \widetilde{x})$, where $\widetilde{x} \leq \overline{x}$. She commits to monitor each firm's process at *the 2nd stage* and inspects their quality at *the 3rd stage* with probability r.

[20] Generally, the expected proportion of quality H in the market is

$$b = (\sum_{i=1}^{N} x_i q_i)/(\sum_{i=1}^{N}(1_{i,h}x_i + (1 - 1_{i,h})(1 - r + rx_i)q_i)),$$

where $1_{i,h} = 0$ if $I_i = I_f$ and $1_{i,h} = 1$ if $I_i = I_h$. If $q_i = q$ for all $i = 1, .., N$, then b only depends on $x_i = x$ and $I_i = I_j$ $(j = h, f)$. Moreover, x_i and I_i may be affected by the public intervention $\widetilde{x}$ and r, hence $b = b(r, \widetilde{x}, x, I_j)$

The 2nd stage (production stage): firms choose (q_i, x_i) $(i = 1, .., N)$ based on their strategy of $I_i = I_j$, where $j = h, f$; the regulator monitors the process and observes x_i. She only certifies the firms with $x_i \geq \widetilde{x}$.

The 3rd stage (inspection stage): the regulator inspects each firm's products with probability r. If one firm's quality is inspected and proved to be L, then the firm is forced to recall all its products. At this stage, the punishment is 0.

The 4th stage (consumption stage): the price p equalizes the demand and the supply. Moreover, at the end of this stage, if a consumer detects that the purchased product is L (only in the experience goods case), he or she reports to the regulator and the firm pays the punishment p.

3.3.6 Definition of Equilibrium

The conception of equilibrium is the sub-game perfect Nash equilibrium (SPNE). I focus on *symmetric* equilibrium, where all firms choose the same strategy (q, x, I_j) $(j = h, f)$ at *stage 2*.

Consumers' Surplus Suppose that at the end of *stage 3*, k $(k \leq N)$ firms can sell their food in H market. Then the supply $Q = kq$. Rewriting formula (3.1), the consumer surplus in equilibrium is

$$CS(kq, \widetilde{x}) = \overline{\theta} kq[b(1 - \frac{kq}{2}) - \widetilde{b}^*(1 - kq)] = \frac{\overline{\theta}}{2} b(kq)^2,$$

because of $\widetilde{b}^* = b$ in equilibrium. Then the expected consumer surplus is

$$E(CS(r, \widetilde{x}, q, x, I_j)) = \frac{\overline{\theta}}{2} \sum_{k=0}^{N} b(kq)^2 Pr(Q = kq | I_j, r, x, \widetilde{x}). \tag{3.2}$$

Note that if $x < \widetilde{x}$, for any $I_j \in \{I_h, I_f\}$ and $r \geq 0$, the conditional probability of $Q = kq$ is

$$Pr(Q = kq | I_j, r, x, \widetilde{x}) = 1 \text{ if } k = 0, \text{ or } 0 \text{ if } k > 0.$$

The reason is that $x \geq \widetilde{x}$ is required by the regulator as one condition of entering H market. No matter how honest the firms intend to be or how frequent the PI is, the H market cannot exist given $x < \widetilde{x}$.

Now consider that the firms' processes are certified (i.e., $x \geq \widetilde{x}$) and all of them reveal their quality truthfully (i.e., $I_j = I_h$). $I_j = I_h$ implies that there are exactly k firms' products are H, given $Q = kq$. Remember that the quality is independent between any two firms and the probability of H is the process x. Then the conditional probability is

$$Pr(Q = kq|I_j, r, x, \widetilde{x}) = \frac{N!}{(N-k)!k!} x^k (1-x)^{N-k},$$

where $\frac{N!}{(N-k)!k!}$ is the combination number.

There is another interesting situation: $x \geq \widetilde{x}$ and $I_j = I_f$. In this case, $Q = kq$ implies that there are k firms, whose products are either H (with prob. x), or are L (with prob. $(1-x)$) but not inspected (with prob. $(1-r)$). Given x and r, a firm's products are sold in H market with *ex ante* probability

$$s(r, x) = x + (1-x)(1-r) = 1 - r + rx \geq x. \tag{3.3}$$

$s(r, x)$ equals x only when $r = 1$; otherwise, it is strictly larger than x. Consequently, the conditional probability is

$$\begin{aligned} Pr(Q &= kq|I_j, r, x, \widetilde{x}) = \frac{N!}{(N-k)!k!} s(r, x)^k (1 - s(r, x))^{N-k} \\ &= \frac{N!}{(N-k)!k!} (1 - r + rx)^k (r - rx)^{N-k}. \end{aligned}$$

Firms' Strategy $(r, \widetilde{x})$ is decided by the regulator at *stage 1*. Now consider that at *stage 2*, all firms other than i $(i = 1, .., N)$ choose (q, x, I_j) $(j = h, f)$ as their common strategy. Denote $y = (q, x, I_j, r, \widetilde{x})$ and $m \in \{j, -j\}$. Firm i's expected profit is $E(\pi_i(q_i, x_i, I_m, y))$, when its strategy is (q_i, x_i, I_m). If (q, x, I_j) maximizes firm i's expected profit, i.e.,

$$(q, x) \in \arg \max_{q_i, x_i \geq \widetilde{x}} E(\pi_i(q_i, x_i, I_j, y)) \text{ and}$$

$$E(\pi_i(q, x, I_j, y)) \geq \max_{q_i, x_i \geq \widetilde{x}} E(\pi_i(q_i, x_i, I_{-j}, y)),$$

then (q, x, I_j) is a symmetric sub-game equilibrium. To simplify the expression, I denote $E(\pi(q, x, I_j, r, \widetilde{x}))$ as the firms' common expected profit in this equilibrium.

Policy Let the best symmetric response of firms be $q_i^*(r, \widetilde{x}), x^*(r, \widetilde{x}), I_j^*(r, \widetilde{x})$ $(j = h, f)$. Given $(r, \widetilde{x})$, the expected social welfare is defined as

$$E(SW(r, \widetilde{x})) = E(CS(r, \widetilde{x}, q^*, x^*, I_j^*)) + NE(\pi(q^*, x^*, I_j^*, r, \widetilde{x})) - rNK, \tag{3.4}$$

where rNK is the expected cost of PI and the expected consumers' surplus $E(CS)$ is given by formula (3.2). The policy $(r, \widetilde{x})$ is chosen by the regulator to maximize this expected social welfare.

Definition 3.1 A symmetric equilibrium consists of r^*, $\widetilde{x}^*$, $q^*(r,\widetilde{x})$, $x^*(r,\widetilde{x})$, $I_j^*(r,\widetilde{x})$ $(j = h, f)$ and $\widetilde{b}^*(r,\widetilde{x})$, such that for any r and $\widetilde{x}$,

1. $\widetilde{b}^*(r,\widetilde{x}) = b(r,\widetilde{x}, x^*(r,\widetilde{x}), I_j^*(r,\widetilde{x}))$;

2. $(q^*(r,\widetilde{x}), x^*(r,\widetilde{x})) \in \arg\max_{q_i, x_i \geq \widetilde{x}} E(\pi_i(q_i, x_i, I_j^*(r,\widetilde{x}), y^*(r,\widetilde{x}))$;

3. $E(\pi(q^*(r,\widetilde{x}), x^*(r,\widetilde{x}), I_j^*(r,\widetilde{x}), r, \widetilde{x})) \geq \max_{q_i, x_i \geq \widetilde{x}} E(\pi_i(q_i, x_i, I_{-j}^*(r,\widetilde{x}), y^*(r,\widetilde{x}))$;

and 4. $(r^*, \widetilde{x}^*) \in \arg\max_{r,\widetilde{x}} E(SW(r,\widetilde{x}))$,

where y denotes $(q, x, I_j, r, \widetilde{x})$ and $E(SW(r,\widetilde{x}))$ is given in formula (3.4).

3.4 The Benchmark Result: Search and Experience Goods

Search Goods

Consider first the case of a search attribute, where consumers detect product quality before purchase. There is no quality signaling problem, since consumers do not acquire any L product. Specifically, firms can only sell H products at *stage 4* (i.e., $I_j^*(r,\widetilde{x}) = I_h$). Therefore, the regulator does not need PI to prevent L products from consumption, i.e., the optimal PI rate is $r^* = 0$. The credibility of H market and consumers' belief are $b = \widetilde{b}^* = 1$, for any process standard $\widetilde{x} \in [0, \overline{x}]$.

If all firms choose to produce q units in *stage 2* and only k sellers' quality is H, then the supply is $Q = kq$. Moreover, if all firms choose x as their process, the expected consumer surplus (formula (3.2)) is

$$\begin{aligned} E(CS(r^*, \widetilde{x}, q, x, I_j^*)) &= E(CS(0, \widetilde{x}, q, x, I_h)) \\ &= \frac{\overline{\theta}}{2} Nx((N-1)x+1)q^2. \end{aligned} \tag{3.5}$$

Now suppose that at *stage 2*, all firms other than i choose (q, x). Without loss of any generality, I focus on $x \geq \widetilde{x}$ (the regulatory standard). If there are $k-1$ $(\leq N-1)$ firms other than i whose products are H at *stage 4*, the supply in the market is

$$Q = kq \text{ if } i\text{'s products are } L \text{ or } Q = kq + q_i \text{ if } i\text{'s products are } H,$$

where q_i is i's quantity. Firm i can recover its cost cx_iq_i and thus is mindful of the market price, only when its quality is H. Note that the equilibrium price is $p = \overline{\theta}\widetilde{b}^*(1-Q)$ (see subsection 3.3.4) and here $\widetilde{b}^* = 1$. Thus, i's expected price with q_i is

$$\begin{aligned} E(p(q_i, q, x)) &= \overline{\theta} \sum_{k=0}^{N-1} (1 - kq - q_i) Pr(Q - q_i = kq) \\ &= \overline{\theta} \sum_{k=0}^{N-1} (1 - kq - q_i) \frac{(N-1)!}{k!(N-1-k)!} x^k (1-x)^{N-1-k} \\ &= \overline{\theta}(1 - q_i - x(N-1)q). \end{aligned}$$

Firm i's challenge is to maximize its expected profit, which is

$$\max_{q_i, x_i \geq \widetilde{x}} q_i x_i (Ep((q_i, q, x)) - c) = q_i x_i (\overline{\theta}(1 - q_i - x(N-1)q) - c),$$

where $x_i \geq \widetilde{x}$ is required by the regulator. Note that to achieve a positive expected profit, $(\overline{\theta}(1 - q_i - x(N-1)q) - c)$ has to be positive in equilibrium. Otherwise, firm i is better off by simply choosing $q_i = 0$ and staying out of the market. So i's expected profit increases with x_i. The only symmetric equilibrium process is

$$x_i^*(r, \widetilde{x}) = x^*(r, \widetilde{x}) = \overline{x},$$

for any r and $\widetilde{x}$. In other words, the process with the maximum provision of quality is adopted by all firms in equilibrium. Here motivated by consumers' demand and awareness of quality, firms voluntarily improve their production processes to provide more H foods. The assumption of a linear cost function $cq_i x_i$ makes it endogenous that safest process $\overline{x}$ would be used. However, with other conventional assumptions, one can also expect a relatively high production of H quality due to the firms' voluntary process choices.[21]

Furthermore, using the first order condition for q_i, the only symmetric equilibrium quantity is

$$q_i^*(r, \widetilde{x}) = q^*(r, \widetilde{x}) = \frac{\overline{\theta} - c}{\overline{\theta}(2 + (N-1)\overline{x})}, \tag{3.6}$$

for any r and $\widetilde{x}$.

The above analysis demonstrates that for search goods, regulation on food quality is not necessary. On one hand, PI is wasteful because consumers themselves can detect the quality. On the other hand, PC is also unnecessary since firms choose the safest process regardless of the regulatory standard.

Experience Goods

[21] For instance, Segerson (1999) uses cost function $c(x, q)$, where $c_x > 0$ and $c_{xx} \geq 0$. It suggests that the search goods assumption may lead to social optimal quality provision level without any regulation.

For experience goods, quality is an experience attribute, i.e., if it is observable only after purchase. There is a potential quality signaling problem; since the firms may shirk, the consumers are not ready to pay a high price for quality. The most common and cheapest signal consists in offering a warranty along with the products, which could be an *endogenous* punishment rule. Specifically, if a consumer detects the purchased product is L, he or she reports to the regulator. The regulator then withdraws the firm's ability to sell it. I assume that if a seller is indifferent between selling L products or not, he will choose $I_j^* = I_h$. As a result, all firms only sell H products at equilibrium. Because consumers know the exact quality before consumption: the *endogenous* limited liability turns the experience goods to search ones. Hence, the symmetric equilibrium with experience goods is the same as that with search goods.

Proposition 3.1. *Under search and experience goods assumptions and A3.1, A3.2, neither PI nor PC is necessary, i.e.,* $r^* = 0$ *and* $\widetilde{x}^* = 0$ *consists of the optimal regulation. Firms chooses the safest process* $\overline{x}$ *and quantity* $\frac{\overline{\theta}-c}{\overline{\theta}(2+(N-1)\overline{x})}$ *in symmetric equilibrium.*

Similar to Darby & Karni (1973), markets for search and experience goods easily remedy themselves from market failure caused by informational asymmetry between consumers and firms. In fact, the informational asymmetry does not exist with search goods because consumers can observe the quality before purchasing. In the case of search goods, liabilities can supply consumers with the signal that food actually does conform to the quality description given by firms, and then solve the problem. However without any regulation, the market for credence goods cannot avoid the market failure. This case will be discussed in the next section.

3.5 Credence Goods

In this section, I assume that consumers never observe the quality level of the products whether prior to or after they purchase credence goods. Both nutritional contends and safety level of foods are considered as quality attributes of this type: an extreme case of experience goods, where the lag between consumption and quality detection tends towards infinity. The firms are assumed unable to signal their process on their own. Only public regulation can get that job done.

If there is no regulation at all, i.e., $r = 0, \widetilde{x} = 0$, the market would be closed. To see this point, consider all firms other than i $(i = 1, .., N)$ choose $q > 0$, $x > 0$ and I_j $(j = h, f)$ as strategy. Since there is no product inspection, firms' L products can be sold in H market. Moreover, firms do not worry about the *ex post* punishment due to quality being credence attributes. Therefore, it is optimal for firm i to choose I_f to fraud. The equilibrium price is $p = \overline{\theta}\widetilde{b}^*(1-Q)$ for any supply Q. Then i's expected profit can be written as

$$q_i(\overline{\theta}\widetilde{b}^* E(1-Q) - cx_i).$$

This expected profit function is strictly decreasing on x_i, thus in equilibrium i chooses the process $x_i = 0$. Consequently, the only symmetric equilibrium is characterized by $x^* = 0$, i.e., there is no H food produced and the market's credibility is $b = 0$. In equilibrium, consumers' belief is $\widetilde{b}^* = b = 0$. Thus, consumers know that every unit of food sold in H market is actually L. They do not purchase any products in such a market.

Proposition 3.2. *Under credence goods assumption and A3.1, A3.2, regulation is necessary; no regulation (i.e., $r = 0, \widetilde{x} = 0$) leads to a market closure.*

With credence goods, consumers cannot discriminate between L and H products before or after purchase. Thus, a firm that would think of improving quality anticipates that it will not be able to recover its cost, due to the absence of quality signals. It then supplies the minimal level of H products: 0. On the other hand, consumers anticipate that since firms' profits decrease with higher probability of H, they are going to only offer L products, no matter what the prices are or which quality is claimed. Then the H market collapses. Consequently, regulation of the food quality is needed in this case.

The Regulation with Perfect Signaling of Quality:
Now consider the extreme case where regulation leads to a perfect signal of quality, i.e., $\widetilde{b}^* = b = 1$. If so, it is necessary that in equilibrium, all firms honestly deal with L products and choose $I_j^* = I_h$. To achieve that goal, the regulator needs to make the strategy I_f not profitable for any firm. Since the consumers can never detect the quality, PI is required.

Lemma 3.1 *In equilibrium, firms choose I_h if and only if $r = 1$.*

Proof. Suppose one firm's products are L at the end of production. If $r < 1$, the firm can sell them with probability $1 - r > 0$ at positive price. Moreover, even if the inspection revealed the fraud, there is no positive punishment due to the endogenous limited liability. Thus, the expected profit at the end of *stage 2* is strictly positive when the firm chooses I_f. However, if the firm chooses I_h, they expect a 0 profit. As the result, for any $r < 1$, the best response for any firm is $I_j^* = I_f$. □

Lemma 3.2 *If the regulation leads to the perfect signal of quality, then the rate of PI is $r = 1$. The symmetric sub-game equilibrium is the same as that of the benchmark case, where $q^*(r, \widetilde{x}) = \frac{\overline{\theta} - c}{\overline{\theta}(2+(N-1)\overline{x})}$ and $x^*(r, \widetilde{x}) = \overline{x}$, for any $\widetilde{x} \in [0, \overline{x}]$.*

The proof of lemma 3.2 is straightforward. Perfect signal of quality turns credence goods into a search good. The sub-game outcome is the same as the case with search goods. Proposition 3.2 suggests that if there is no PI (or PC), firms do not provide quality voluntarily. Lemma 3.2 suggests that when PI is most intensive (i.e., $r = 1$), firms improve their process to $\overline{x}$, regardless of the policy standard $\widetilde{x}$. Thus, PI is effective for provision of high quality with credence goods. Because sampling and testing requires a social cost up to rNK (see formula (3.4)), the inspection cost may lead to relatively low social welfare if r is relatively large (e.g., $r = 1$). In lemma 3.2, the outcome is a sub-game equilibrium, but may not be an equilibrium. The regulation with perfect signaling of quality may not be optimal.

The Regulation with Imperfect Signaling of Quality:

Consider that the regulation does not generate a perfect signal of quality. Lemma 3.1 implies that here the PI rate satisfies $r < 1$. All firms commit fraud if their products are L, i.e., $I_j^* = I_f$. Suppose all firms choose process x. With $I_j^* = I_f$, one firm's *ex ante* probability that his products can be sold in H market is $s(r, x) = 1 - r + rx$ (formula (3.3)). But their *ex ante* probability of the products being H is still x. The credibility of the market is the expected percentage of H products in the market, which is

$$b(r, \widetilde{x}, x, I_f) = \frac{\Pr(\text{products are } H)}{\Pr(\text{products can be sold in } H \text{ market})} = \frac{x}{1 - r + rx} < 1.$$

It is strictly less than 1 due to $r < 1$. If x is the equilibrium process, then consumers' belief

$$E(\widetilde{b}^*(r, \widetilde{x})) = b(r, \widetilde{x}, x, I_f) = \frac{x}{1 - r + rx}. \tag{3.7}$$

The PI cost is also absorbed by the regulator. She has another approach to enhance the quality with little cost: PC. When it is not optimal for her to inspect the products very often, she could set a high process standard $\widetilde{x}$ to increase the expected proportion of H products in the production. A dominant strategy of hers is $\widetilde{x} = \overline{x}$. With the safest process $\overline{x}$ being mandatory, the consumers' belief is the highest, because $E(\widetilde{b}^*(r, \widetilde{x}))$ (formula (3.7)) increases with firms' process x. And the demand $D(p) = 1 - p/(\widetilde{b}\overline{\theta})$ increases with $\widetilde{b}$, for any price p. Thus, setting $\widetilde{x} = \overline{x}$ leads to the largest demand and then may generate the highest social welfare. More precisely, consider all firms choose x as their process. Firms other than i choose q as their quantity. Given i being approved to enter the H market, the expected price is

$$E(p(q_i, r, \widetilde{x}, q, x)) = \overline{\theta} E(\widetilde{b}(r, \widetilde{x})) \sum_{k=0}^{N-1} (1 - kq - q_i) Pr(Q - q_i = kq), \tag{3.8}$$

where $E(\widetilde{b}(r,\widetilde{x}))(1-kq-q_i)$ is the expected price conditional on $Q = kq + q_i$ for any $k \leq N-1$. Firms independently enter the market with probability $s(r,x)$. And they expect an equilibrium consumer's belief. So the expected price in (3.8) is

$$\begin{aligned} E(p(q_i, r, \widetilde{x}, q, x)) &= E(\widetilde{b}^*(r,\widetilde{x}))(1-q_i-s(r,x)(N-1)q) \\ &= \frac{x}{1-r+rx}\overline{\theta}(1-q_i-(1-r+rx)(N-1)q). \end{aligned}$$

Consequently, firm i's expected profit given its quantity q_i is

$$\begin{aligned} E(\pi_i) &= q_i[s(r,x)E(p(q_i,r,\widetilde{x},q,x)) - cx] \\ &= q_i x(\overline{\theta}(1-q_i-(1-r+rx)(N-1)q)-c). \end{aligned}$$

If $q_i = q$ maximizes this expected profit, the symmetric quantity choice is

$$q = \frac{\overline{\theta}-c}{\overline{\theta}[2+(1-r+rx)(N-1)]}. \tag{3.9}$$

Taking this quantity to formula (3.4), the expected social welfare is

$$\begin{aligned} E(SW) &= \frac{\overline{\theta}}{2}xN\{[(1-r+rx)(N-1)+3] \\ &\quad [\frac{\overline{\theta}-c}{\overline{\theta}[2+(1-r+rx)(N-1)]}]^2\} - rNK. \end{aligned} \tag{3.10}$$

This expected social welfare is increasing with x. If the regulator picks the safest process as the standard, the only sub-game equilibrium process is $x^*(r,\widetilde{x}) = \widetilde{x} = \overline{x}$. Then the expected social welfare is maximized for any given r. Lemma 3.3 summarizes the above analysis.

Lemma 3.3 *For any given $r < 1$, $\widetilde{x} = \overline{x}$ (weakly) dominates any $\widetilde{x} \in [0,\overline{x}]$, i.e., $\widetilde{x} = \overline{x}$ generates symmetric sub-game equilibrium leading to higher or at least equal expected social welfare than any $\widetilde{x} \in [0,\overline{x}]$ does.*

Lemma 3.3 suggests that the regulator should set $\overline{x}$ as the process standard when the public signal is imperfect. The linear cost for the process is critical for this result. In the real world, regulatory standards may not provide maximum quality. As new technologies for food production, stocking and delivery are invented, the conception of the safest processes evolves. However, set-up costs for new technologies are usually enormous. Forcing food companies to constantly update their process may not be efficient. Nevertheless, lemma 3.3 fits the conventional wisdom about food safety issues: *i)* policies with the clear intention for reducing safety should not exist; and *ii)* the safest process should be always recommended by authorities. Moreover, in this paper, consumers are risk neutral. When modeling food quality, they can be introduced as risk averse, and they would like to pay more for higher credibility in H markets. Remember that the credibility (formula (3.7)) is maximized when $x = \widetilde{x} = \overline{x}$ for any PI rate. The public signal using $\widetilde{x} = \overline{x}$ may lead to largest demand and social welfare.

The Optimal Regulation:
As implied by lemma 3.3, all firms adopt process $\overline{x}$ in equilibrium if $r < 1$. Because $\widetilde{x} = \overline{x}$ is a dominant strategy of the regulator, if any firm deviates from $\overline{x}$, the regulator can just make this process mandatory. Taking $x = \overline{x}$ into formula (3.9), the equilibrium quantity is

$$q^*(r, \widetilde{x}) = \frac{\overline{\theta} - c}{\overline{\theta}[2 + (1 - r + r\overline{x})(N - 1)]}. \tag{3.11}$$

Furthermore, take $r = 1$ into formula (3.11). $q^*(r, \widetilde{x})$ equals to $(\overline{\theta} - c)/[\overline{\theta}(2 + (N - 1)\overline{x})]$, which is, according to lemma 3.2, the sub-game quantity outcome when the public signal is perfect. Remember that all firms also choose $x = \overline{x}$ in this case. Therefore, in any symmetric equilibrium, firms adopt the safest process and their quantity is given by formula (3.11).

Lemma 3.4 *In symmetric equilibrium, each firm's process is* $x^*(r^*, \widetilde{x}^*) = \overline{x}$ *and quantity is* $q^*(r^*, \widetilde{x}^*) = \frac{\overline{\theta} - c}{\overline{\theta}[2 + (1 - r + r\overline{x})(N - 1)]}$, *where* $(r^*, \widetilde{x}^*)$ *is the optimal policy.*

In the following, the optimal policy $(r^*, \widetilde{x}^*)$ is investigated. Note that setting $\widetilde{x} = \overline{x}$ is not the only way of making firms to provide maximum quality. Lemma 3.2 implies that PC may not be necessary, i.e., $\widetilde{x}^* = 0$. Intensive product inspection (e.g., $r = 1$) is the only incentive for firms to dispose of their L products honestly, i.e., $I^* = I_h$, and they adopt the safest process $\overline{x}$ to minimize their probability of producing L food. In fact, certifying process may not be necessary either with some $r < 1$, where according to lemma 3.1, firms try to fraud (i.e., $I^* = I_f$). Specifically, suppose that the regulator optimally sets $\widetilde{x}^* = 0$. As implied by lemma 3.4, the firms' symmetric best response to the policy satisfies $q^*(r, \widetilde{x}) = (\overline{\theta} - c)/[\overline{\theta}(2 + (1 - r + r\overline{x})(N - 1))]$ and $x^*(r, \widetilde{x}) = \overline{x}$ in equilibrium. And the equilibrium belief (formula (3.7)) is $\overline{x}/(1 - r^* + r^*\overline{x})$. Then a single firm i's $(i = 1, .., N)$ expected price with quantity $q_i = q^*$ is

$$\begin{aligned} & E(\widetilde{b}^*(r^*, \widetilde{x}^*))\overline{\theta}(1 - q^* - (1 - r^* + r^*\overline{x})(N - 1)q^*) \\ = & \frac{\overline{x}}{1 - r^* + r^*\overline{x}}(c + q^*). \end{aligned}$$

Consider for all $j \neq i$, $q_i = q_j = q^*$ and $x_j = \overline{x}$. The expected profit of firm i is

$$\begin{aligned} E(\pi_i) &= q^*[s(r^*, x_i)\frac{\overline{x}}{1 - r^* + r^*\overline{x}}(c + q^*) - cx_i] \\ &= q^*[\overline{x}\frac{1 - r^* + r^*x_i}{1 - r^* + r^*\overline{x}}(c + q^*) - cx_i], \end{aligned}$$

where x_i is firm i's process. Since $\widetilde{x}^* = 0$, this single firm does not deviate from the symmetric equilibrium $x_i = \overline{x}$, if and only if

$$\frac{r^*\overline{x}}{1 - r^* + r^*\overline{x}}(c + q^*) \geq c \text{ or } r^* \geq \widetilde{r} = \frac{c}{c + \overline{x}q^*} < 1.$$

In other words, high inspection rates $r^* \geq \widetilde{r}$ encourage firms to adopt $\overline{x}$ voluntarily, even if they attempt to fraud with L products.

However, $\widetilde{r} \leq r^* < 1$ cannot be included in the optimal policy in this setting. *PI* is an approach of quality signaling by preventing some L products from consumption. Intuitively, higher *PI* rate r generates high credibility of H market b and consumers' belief $\widetilde{b}$. The former enhances the consumer surplus (see formula (3.1)), because high b suggests a large probability that consumers can get high quality food. The latter increases the firms' expected price (see formula (3.8)), because consumers are willing to pay more if they have more faith in the quality signal. Thus, it also increases the firms' expected profits. This is only one factor of social gain (sum of consumer surplus and firms' profits) from *PI*. Another factor is the quantity. Note that $q^* = (\overline{\theta} - c)/[\overline{\theta}(2 + (1 - r + r\overline{x})(N - 1))]$ increases with r. The reason behind this relationship between q^* and r is straightforward. With high rate r, each firm expects a high possibility that his competitors are prevented from selling. Therefore all firms prepare to take others' share in market by enhancing their own quantities. And in a Cournot setting, a larger total quantity usually leads to a larger social welfare. These two factors timing each other causes the social welfare function is convex in r. Thus, r^* equals to 0 or 1 in equilibrium.

Proposition 3.3. *Under credence goods assumption and A3.1, A3.2, the optimal regulation adopts*

1. *only PI, i.e.,* $r^* = 1$ *and* $\widetilde{x}^* = 0$*, if* $K < \widetilde{K}$*,*

2. *only PC, i.e.,* $r^* = 0$ *and* $\widetilde{x}^* = \overline{x}$*, if* $K \geq \widetilde{K}$*,*

where

$$\widetilde{K} = \frac{\overline{x}(\overline{\theta} - c)^2}{2\overline{\theta}}[\frac{\overline{x}(N-1) + 3}{(\overline{x}(N-1) + 2)^2} - \frac{N+2}{(N+1)^2}].$$

.

Proof. Taking the equilibrium value of process and quantity into formula (3.10), the regulator's problem is

$$\max_{r,\widetilde{x}} E(SW) = \frac{N\overline{x}(\overline{\theta} - c)^2}{2\overline{\theta}} \frac{3 + (1 - r + r\overline{x})(N-1)}{[2 + (1 - r + r\overline{x})(N-1)]^2} - rNK.$$

Since $\widetilde{x}$ never affects the equilibrium social welfare, the regulator's problem is equivalent to

$$\max_{q^*} E(SW) = \frac{\overline{\theta}N\overline{x}}{2}[(\overline{\theta}-c)q^* + q^{*2}] - \frac{q^*(N+1)-(1-c/\overline{\theta})}{q^*(N-1)(1-\overline{x})}NK,$$

where $q^* = (\overline{\theta}-c)/[\overline{\theta}(2+(1-r+r\overline{x})(N-1))]$ is firms' best response to $(r,\widetilde{x})$. It is a convex problem in q^* due to

$$\frac{\partial E(SW/N)^2}{\partial^2 q^*} = \overline{x}q^* + 2\frac{(1-c/\overline{\theta})K}{(N-1)(1-\overline{x})q^{*3}} > 0.$$

So the solution is $q^* = (1-c/\overline{\theta})/(2+(N-1)\overline{x})$ and $r^* = 1$, if

$$K < \widetilde{K} = \frac{\overline{x}(\overline{\theta}-c)^2}{2\overline{\theta}}[\frac{\overline{x}(N-1)+3}{(\overline{x}(N-1)+2)^2} - \frac{N+2}{(N+1)^2}];$$

otherwise, the solution is $q^* = (1-c/\overline{\theta})/(N+1)$ and $r^* = 0$. From the above analysis, $r^* \geq \frac{c}{c+\overline{x}q^*}(<1)$ encourages firms to adopt the safest process. Since there is a slight cost for monitoring the process, setting $\widetilde{x}^* = 0$ and abandoning *PC* is the best policy for the regulator in this case. If $r^* = 0 < \frac{c}{c+\overline{x}q^*}$, the safest process is required in the optimal regulation. Then the equilibrium policy satisfies

$$\widetilde{x}^* = \overline{x}, \text{ if } r^* = 0\text{or } 0, \text{ if } r^* = 1.$$

□

Proposition 3.3 suggests that *PC* may not be compatible with *PI*. When the optimal policy includes *PI*, it should provide a perfect signal. With high frequency $r^* = 1$, *PI* solely incentivizes firms to adopt the safest process $\overline{x}$. Then *PC* is unnecessary, but *PI* requires sampling and testing costs. The regulator may provide a noisy quality signal through monitoring and certifying firms' processes, when *PI* cost is relatively large. Note that the credibility of the *H* market is concave in *PI* rate, i.e.,

$$\frac{\partial^2 b(r,\widetilde{x},x,I_f)}{\partial r^2} = \partial^2(\frac{x}{1-r+rx})/\partial r^2 < 0.$$

It seems that *PI* is a convex signaling technology, which introduces concavity in the expected social welfare given in formula (3.4). It could be optimal to combine these two convex technologies, *PC* and *PI*, under some circumstances. But *PI* also changes the firms' quantity choices in equilibrium, which distorts the concave property. Consequently, an additional signal through *PI* cannot be efficient when the final signal is still noisy.

Moreover, $r^* = 1$ is optimal if PI consists of the optimal policy. This result relies on the assumption of *endogenous* limited liability, which exempts firms from any punishment if their fraudulent behavior is caught by PI. Suppose that the regulator introduces some *exogenous* penalty rule, which specifies a relatively small punishment with respect to consumers' benefit of purchasing H product and the PI cost K. The regulator will not commit to any relatively low PI frequency that would induce firms to choose fraud if PI is implemented. It is because that the income from punishments can neither exceed the consumers' loss from fraud, nor cover the PI cost. Consequently, when PI is included in the policy, it provides a perfect signal. Using the same logic of above analysis, PI should be abandoned whenever PC is optimally adopted. The main result of proposition 3.3 cannot be changed.[22]

Note that the threshold $\widetilde{K}$ in proposition 3.3 increases with $\overline{\theta}/2$, which is the consumers' average benefit from eating H products when compared to L products. When the quality is serious safety concern and low quality leads to deadly diseases, consuming safe food matters substantially, i.e., $\overline{\theta}/2$ is sufficiently large. Then $\widetilde{K}$ is likely to exceed the PI cost, which makes PI the optimal choice of food quality signaling.

Corollary 3.1 *For any K and $\overline{x} > 0$, there exists large enough $\overline{\theta}$ for every $N \geq 2$, such that optimal regulation only signals quality perfectly through PI.*

When China's Sanlu incident was caused by deadly additives, more strict regulations on the Chinese diary industry became imperative. PI, which was abandoned by the label of National Inspection Exemption, is now re-activated. It seems that the government has not found a cheap method for testing the deadly additives. To date, China's PI frequency is believed to be incapable of eliminating fraud. However, Chinese have realized that in dealing with food safety, the old-fashioned approach PI is more desirable for consumers than PC.

Furthermore, $\widetilde{K}$ in proposition 3.3 tends to 0 when N tends to infinity. It implies that when competition is perfect (i.e., $N = \infty$), the advantage of PC is most obvious. In this case, any inspection cost $K > 0 = \widetilde{K}$ leads to $r^* = 0$ and $\widetilde{x}^* = \overline{x}$.

Corollary 3.2 *With perfect competition, optimal regulation only signals quality imperfectly through process certification.*

As mentioned above, firms increase their quantities with hope that PI can drive some competitors out of the market. They expect their market share will be extended after

[22] When the punishment is relatively large, policies combining PI and PC may be optimal. Specifically, the regulator conducts PI with relatively low frequency. By doing that, firms commit fraud with quality L. Then the regulator expects to collect punishments covering the PI cost or more. This low PI rate may not induce firms to adopt the safest process $\overline{x}$. So the regulator also sets $\overline{x}$ as the process standard. However, the enforcement of heavy punishment may be a problem, because profit from fraud has not been generated.

the inspection. But in equilibrium, the marginal affect of PI over the quantity can be diluted by intensive competition, since with large enough N, the following inequality holds:

$$\frac{\partial^2 q}{\partial r \partial N} = \frac{(\overline{\theta} - c)^2}{\overline{\theta}}(1 - \overline{x})\frac{2 - (1 - r + r\overline{x})(N-1)}{[2 + (1 - r + r\overline{x})(N-1)]^3} < 0.$$

In words, PI cannot affect firms' decision over quantity significantly. Then its impact on the social gain from quality signaling is diminutive. Another reason that PI cannot be involved in optimal regulation is that if N is large, then its expected cost is rNK. Note that the social benefit from PI is bounded. If too many sampling and testing should be done due to the large N, the expenditure will be an unlimited burden for the regulator. Therefore PC is more favorable in this situation.

Nevertheless, corollary 3.2 also suggests a paradox of PC: PC is more effective than its alternative before it becomes the policy, but after that it is not. The reason is that PC, which is more efficient with intensive competitiveness, may undermine the competition due to its relatively large set-up cost. Some companies may not be able to afford that cost and will then leave the market. To see this paradox, consider the following example.

Illustration Example:

$$\text{Let } K = 0.04 \times 0.4 \times \frac{(\overline{\theta} - c)^2}{\overline{\theta}}, \ \overline{x} = 0.8 \text{ and } N = 4.$$

The regulator optimally stops PI and starts to monitor these 4 firms' processes, since

$$\widetilde{K}(N = 4) \approx 0.0389 \times 0.4 \times \frac{(\overline{\theta} - c)^2}{\overline{\theta}} < K.$$

Now consider one firm cannot pay for the set-up cost for PC. And there is only $N' = 3$ firms left in the market due to the implementation of PC. Then the optimal policy changes to PI, since

$$\widetilde{K}(N' = 3) \approx 0.0425 \times 0.4 \times \frac{(\overline{\theta} - c)^2}{\overline{\theta}} > K.$$

The above example suggests that policy makers should estimate and adopt PC with discretion. To overcome this paradox, governments may need to subsidize firms whose budgets are exceeded by the set-up cost, or they may support these firms with financial plans. After all, an effective and credible food quality regulatory system should be consistent without hampering the competitiveness of the industry or leaving little marginal benefit to the consumers.

3.6 Conclusion and Further Developments

Food quality is a widespread concern in today's public policies. In developed countries, as incomes rise, consumers are prepared to pay for a regulatory regime that provides higher quality and minimizes risk. Public regulation might not be the only approach. The others include economic incentives, voluntary practice, private standards and labeling, etc. However, according to Strauz (2005), a certification agency without any competition should be needed to provide a consistent and credible signal for credence goods in a long period. And the food quality is highly related to public health and safety. Therefore, governments are usually considered as the best candidates to monitor and control the food quality provision. In developing countries, where average incomes are relatively low, food companies are less likely to recover the costs of quality provision from marketing.[23] Governments have to respond by setting strict regulations to ensure food quality.

This chapter investigates the efficiency of food quality regulatory regimes. Particular attention is paid to two approaches: the newer *process certification* and the more traditional *product inspection*. Using *PI*, regulations can provide a perfect signal of quality but relatively large monitoring costs must be incurred. On the contrary, *PC* is a less expensive method but leads to imperfect quality signals. Although most foods are credence goods, the case of search and experience goods are also studied, which suggests that neither *PC* nor *PI* is needed for signaling the quality, because the informational asymmetry between firms and consumers is solved by the market mechanism. In the case of credence goods, market failure may be inevitable without any regulation. If using *PC*, the regulator should set the safest process as the standard to mandate the maximum quality provision in production. If using *PI*, the regulator should inspect the products as frequently as possible, such that no business fraud exists in equilibrium. Moreover, proposition 3.3 suggests that combining *PC* and *PI* may not generate more cost-effective regulatory approaches. If the cost of *PI* is larger than some threshold, it should be stopped in optimal regulations. However, the threshold is relatively large when the quality is some serious safety attribute. Under this circumstance, *PI* is more efficient than *PC*. The threshold also depends on the competitiveness of the market. With perfect competition, *PC* is always a better choice for policy makers. But when the set-up costs for monitoring *PC* must be paid by firms themselves, small market powers imply little chances of firms recovering the set-up costs. Then some firms may withdraw from the market which may causes a paradox: *PC* is a more cost-effective approach than *PI* before *PC*'s implementation; but after that, it is not.

[23]For example, Wanga et al. (2008) investigated the price premium for Chinese milk products with safety labels. The premium in Beijing supermarkets is only about 5% over products without a label.

However, the pro-competitive effect of PI may be overestimated, because I assume the quality being independent among different firms.[24] Food contamination could sometimes drive from the utilization of common inputs, such as polluted river used for irrigation and feeding animals. The outbreaks of zoonotic diseases, such as bird flu, also can easily be transmitted in between farms. If the common hazard has been introduced, firms may not increase their quantity with the expectation that their rivals will withdraw their products due to PI. Considering the extreme case where the quality of firms is perfectly correlated. The detection of one firm's unsafe products suggests all products in the whole industry should be disposed. Instead of increasing their quantities, the firms may downsize their scales of production. Consequently, the social welfare can be concave in PI's frequency and achieve its maximized value when PC and PI are combined. Moreover, the optimal regulatory approach (only using PI, or PC, or both of them) will depend not only on the the cost of PI, but also the correlation between firms' quality. It obviously deserves more discussions, which will provide more complete and meaningful practical implications.

[24] Thanks to Sara Biancini for sharing this perspective in her report on this thesis.

Chapter 4

Patents and Common Values: Over-investment in Research and Development

4.1 Introduction

Scientists are continuously searching for the new innovations. However, the new innovations build on the current facts and theories. Without the dissemination of the previous findings, no subsequent results could be easily produced. Nowadays, much of the existing technical and industrial knowledge is protected by patent rights. On one hand, patent rights reveal the knowledge, and the whole scientific community can learn from it. On the other hand, patent rights protect benefits of previous knowledge's finders through legitimizing their position as innovation sellers. Like the production of other goods, the routine of innovation manufacture (known as research and development, or R&D) is affiliated with large scale investments. Patent protection provides scientists an opportunity to recover their R&D investments by allowing only the patent holder to profit from the discovery. In this sense, patent rights may have successfully promoted more research (see for instance, Arora et al. (2004)). Recent policies, such as the U.S. Bayh-Dole Patent and Trademark Amendments Act (known as Bayh-Dole Act),[1] encourage scientists to apply for patent protection over their innovations.

Heller and Eisenberg (1998), however, treat patent protection as a two-edged sword: it spurs scientific research by securing scientists the fruits of their labors, but it also gives patent holders potential power to restrict how others conduct research based on

[1] The U.S. Bayh–Dole Patent and Trademark Amendments Act of 1980 allowed universities to receive patents and grant licenses resulting from their researches. See more in the section of literature review.

the protected knowledge. When patent holders enforce these veto power, only they can search for the future subsequent innovation (or downstream products).[2] In practice, research cannot guarantee successful findings. Blocking other scientists from conducting R&D may delay new discoveries and reduce innovation manufacture in general. Heller and Eisenberg dubbed the veto rights "tragedy of the anticommons": patent holders may block others' research and restrict the usage of their owned innovations to a less than efficient level. Anticommons is rising in biomedical and other industries (see for instance, Lessig (2002); David (2004); Murray and Stern (2007); Huang and Murray (2009)).

Nevertheless, the extent of anticommons is highly debatable according to other scholars. They have two main reasons. First, some theoretical studies, such as Green and Scotchmer (1995), suggest that anticommons may be avoidable through licensing activities: it may be profitable for a patent holder to offer a license at a fee with respect to the licensee's R&D cost. Second, the number (and quality) of new patent-protected scientific innovations has continued to move upward in recent years (Mowery and Ziedonis (2002); Sampat et al. (2003)). This implies that the new patent holders are not simply motivated by their ability to block new innovations by others. Moreover, if new findings have decreased as Eisenberg and Heller suppose, then one would expect a decline in R&D investment. Yet there is little evidence to suggest that this has taken place (Refferty (2008); Noel and Schankerman (2006)). Therefore, the impact of anticommons appears limited.

This paper aims to analyze the controversy over patent rights. It considers a simple model with two scientists: one patentee and one potential licensee. Endowed with same insight, they are both capable of research based on the previous knowledge. The value of the improvement (new innovation) is random, but both scientists have prior knowledge about it. At best, a scientist is uncertain about the contribution his research will make to the world. However they can somehow predict the value using their advanced knowledge. For simplicity, there are two pieces of private information that scientists may have: the bad news τ_1 and the good news τ_2. The good news τ_2 leads to a higher expected value than τ_1. Moreover, a *common-value* situation is assumed: each scientist's private information is an augment of the other scientist's expectation of the value. If both scientists have τ_2, their common expectation is the highest; it is the lowest if they both have τ_1. Due to the scientists' distinguished personalities and backgrounds, they may have

[2] Anticommons may also take place when there are multiple patent holders who stop research. Any dispute among the patentees may delay the licensee's research. Such disputes may happen when a high license fee required by one patentee decreases anothers' profits. If the licensee has already reached an agreement with a patentee and paid his fees, the other patentees may take advantage of the licensee being anxious to recover his costs and therefore require very high payments. This probably creates a negative profit for the licensee, thus he would not bargain license contracts in the first place. See for instance, Llanes and Trento (2009).

ex ante different perspectives on the value of any given project. For instance, a scientist with an optimistic personality is inclined to focus on the positive aspects; while one with a pessimistic personality is more likely to see the negative ones. Through sharing their perspectives, scientists may know more about the issue, thus the efficiency of the patent rights mechanism may depend on its ability to disclose the private information.

One key feature of the mechanism of patent rights is that the potential licensee can only conduct his R&D if he and patentee achieve a license agreement.[3] To capture the veto rights of the patentee over the other's R&D, I assume that he proposes any license contract on a take-it-or-leave-it basis to the licensee. The proposed price of licensing should be increased with the patentee's expected value of the improvement, which might be enhanced significantly after the good news of the licensee is disclosed. Thus, the patentee may screen the licensee's information using a license menu. Furthermore, a blockade could be introduced in the menu: the patentee incentivizes the licensee to report good news by blocking him if he reports bad news. Then anticommons (e.g., fewer new innovations than is efficient) may arise as a "tragedy". Statistically speaking, a single R&D activity leads to fewer findings than multiple ones.

I also assume that an R&D programme requires sunk investment and that, on average, a higher investment leads to more innovations.[4] This assumption immediately combines the observations of "many innovations" and "many R&D investments" into one. Scientists balance their R&D costs with the expected income from their innovations. The patentee may signal his information through his R&D investment in the proposed contract as observable and verifiable actions.[5] Now neither "consecutively increasing new innovations" nor "increasing R&D investments" can prove the efficiency of patent rights. The patentee may conceal his bad news through signaling: he invests heavily to pretend that he has the good news. By doing that, he increases the total chance of finding the new innovation. He may also enhance the licensee's valuation of the improvement, who is then willing to pay for higher license fee. The patentee may not expect to recover the heavy investment from the future income. He may recover it through licensing, when the patentee's good news significantly enhances the licensee's valuation. Such a heavy investment in the patentee's R&D may lead to an over-investment situation: the equilibrium total investment level is strictly larger than the socially optimal one. Hence

[3] For pronominal clarity, I take these two scientists as men. Moreover, because of patent protection, licensing incomes matter substantially to patentees. Patents themselves are expensive to acquire due to R&D investments. To recover the investments, patent holders only make money from licensing the protected innovation, or developing and selling (or licensing) some new downstream innovation, or both. In fact, 71% respondents to Thursby et al. (2003)'s survey claimed that licensing fees generated by patents were extremely important in determining the success of their scientific institutions' management.

[4] For example, compared to a low-tech laboratory, a high-tech one facilitates scientific researches but increases their sunk costs.

[5] See Myerson (1983): when a party designing a contract has private information, the contract may reveal his information to other parties.

the controversy over patent rights may be generated: anticommons and over-investment (and more innovations than efficiency) may *ex ante* co-exist, when some good news leads to a large increase in the common valuation.

The rest of this paper is organized as follows. Section 4.2 is the literature review. I build the model in section 4.3, then solve the equilibrium and present the main results in section 4.4. And section 4.5 is the conclusion.

4.2 Literature Review

After the passage of 1980 Bayh–Dole Act, the extent of anticommons has been empirically studied. That Act created a U.S. federal patent policy that allows universities to retain rights to any patents resulting from government funded research and to license these patents on an exclusive or non-exclusive basis. Although universities continue to publish their findings in peer-reviewed journals, their patent rights may block subsequent research. From an anticommons perspective, the citation rate for a scientific publication should fall after the formal patent rights associated with that publication are granted. However, by using a long stream of patent citations data, Mowery and Ziedonis (2002) and Sampat et al. (2003) suggest that citations did not decline dramatically. Additionally, patents are often granted years after the knowledge is introduced in a paper. Murray and Stern (2007) focus on the changes in the citation rate before and after the patent rights have established. They support anticommons by only finding a modest effect. To identify the Bayh–Dole effect on incentives of R&D investments, Refferty (2008) categorizes universities' R&D expenditure data into three groups: basic research, applied research and developmental research. Research in the first group is less patentable than the other two. One might expect that in order to generate more licensing revenue, universities would shifte funding from basic research to applied and developmental research. Alternately, one could anticipate the opposite as an implication of anticommons: universities de-emphasizing application and development but investing heavily in basic research. According to Refferty (2008), however, no change has taken place saliently: R&D investments in all three groups have generally continued to increase. In all, the negative effect of patent rights (anticommons) on scientific researches does not seem to be significant. The contradiction here is too apparent to be overlooked. This paper attempts to give an explanation.

The mechanism of patent rights has been studied extensively, including theoretical literature on sequential innovation (Scotchmer (1991), Green and Scotchmer (1995), Chang (1995), Scotchmer (1996)). In these models, there are two scientists who research sequentially (the licensee cannot do research until the patentee stops). The goal is to find the patent policy that maximizes the incentives to invest in both scientists' R&D programme. However, the above papers exclude the possibility that the two scientists are competing in their search for new innovation. Note that any patent that creats a blockade may only benefit the patentee, who also pursues new innovations. Otherwise, the revenue from licensing would be preferred over waiting for patent expiration. So in the above papers, any blockade is rare. An exception is Bessen and Maskin (2009). In their setting, both patentee and licensee can conduct R&D based on previous and patented knowledge. Then, a blockade can exterminate the competition in a future market of new innovation. As a result, anticommons may arise: R&D may be under-invested and the new innovation production may be downsized.[6] Furthermore, the patentee also needs the new innovations to develop even newer innovations from which he can make a profit in the distant future. In the long-run, scientists may prefer an environment without patent rights, where the expected number of new innovations is large. So patent rights may not protect the patentee at all and the "tragedy" may become even worse. But neither Bessen and Maskin (2009) nor others, as far as I know, have considered the case where the patentee's information directly affects the licensee's valuation. In this paper, I extend their ideas by introducing the *common-value* assumption. The results suggest that anticommons (under-investment in R&D) is not the only negative impact of patent rights; the effect of over-investment must also be considered.

From the common-value case of Cremer and Mclean (1985), we know that if the joint distribution of parties' information satisfies some appropriate rank conditions, some mechanism discloses all information with zero cost (the socially optimal outcome). It is the *ex ante* correlation between information of parties that pulls the trigger: given the correlation and the assumption that others reveal their information truthfully, it is optimal for any individual to reveal his information under the provided mechanism. However, Auriol and Laffont (1992) point out that Cremer and Mclean's rank conditions can easily be evaded: it is more than a conventional setting to simplify computation or ensure a solution. Without that condition, full informational disclosure requires positive social cost. Thus, the first best results in a common-value model may not be equilibrium. For example, in the Spence (1974) education model, the productivity of a worker affects the employer's payoff, thus has the common-value property. The result is that a highly productive worker may invest in wasteful education to avoid being considered less capable. Note that the over-investment in Spence (1974) is different from that which is studied in this paper, which can be expressed as a low productive worker over-investing in education to be mistaken as a more capable one. Moreover, Maskin and Tirole (1992)

[6] They assume that, to manufacture the same amount of innovation, an efficient licensee invests less in R&D than an inefficient one does. The licensee holds private information about his own efficiency. Thus, a blockade may incentivize an efficient licensee to reveal his type: a patentee may block the inefficient licensee in order to get higher license payment from him if he was efficient.

extend the principal-agent theory with an informed principal under common-value assumption. They focus on the set of equilibrium contracts where the principal truthfully reveals his information. Like Spence (1974), the informational disclosure relies on the assumption of infinite signaling choices of the principal, which belong to some compact and convex set. Then, under some conventional hypothesis, there exists an equilibrium satisfying the assumption that the principal reveals his information and the agent believes he does. In this paper, the principal (patentee) only has limited choices for signaling: finite levels of R&D's fixed investments. Thus, the incentive compatible condition may not be applicable for the principal. Consequently, I also focus on the equilibrium where the patentee may successfully conceal his information. The lack of efficiency contrasts with the positive recognitions of patent rights.

4.3 The Model

I consider two risk-neutral scientists: A and B, whose reservation utilities are normalized to 0. One of them is the patent holder (denoted as $h = A$ or B), who owns exclusive rights over the previous knowledge. The other scientist is a potential licensee (denoted as $-h$), who also understands the knowledge because patent rights have disclosed. At the start of the game, the same insight upon the protected knowledge occurs to both A and B. The improvement, referred as "new innovation", could be one of any new scientific findings, one of downstream applications, or both.

h's veto power over $-h$'s R&D is established by patent protection. First, I assume that it is long enough: waiting for the expiration of the patent is not an option for $-h$. Second, it is wide enough. There exists no substitution of the previous knowledge that $-h$ can use to develop new innovation without infringing h's patent rights. Consequently, h can propose any licensing contract on a take-it-or-leave-it basis to the licensee $-h$.

Furthermore, the social value of the new innovation (denoted as v) remains uncertain until it is manufactured. The game under investigation takes place in two periods: one present period and one future period. Licensing activity and research is in the present. The realization of new innovation and its social value v occurs in the future.

4.3.1 The Previous Knowledge: R&D Technology

New innovation and its social value v can be found in any scientist's R&D programme, if it is successful. The previous knowledge is necessary for producing new innovation. It specifies two R&D technologies: an advanced one and a laggard one. Both of them require fixed costs being invested before research can start. Let $c_i \in \{0, c\}$ ($c > 0$) be scientist i's ($i \in \{A, B\}$) fixed investment in his R&D. If $c_i = 0$, i adopts the laggard technology; if $c_i = c > 0$, the advanced one is used.[7] This c_i is publicly observable when it is sunk and i's R&D's successful rate is

$$p_i = \underline{p} \text{ if } c_i = 0 \text{ or } p_i = \overline{p} = \underline{p} + \Delta p \text{ if } c_i = c.$$

Clearly, if scientist $i \in \{A.B\}$ performs no research, his probability of finding v is $p_i = 0$. The fixed cost of the laggard technology is normalized to 0. Compared to the laggard one, the advanced technology improves the success rate of R&D, but at a higher fixed cost. This setting captures the fact that more R&D costs usually lead to more discoveries. Note that it is more than just assuming that the success rates of R&D will increase along (and marginal decrease) with continued investment. In other literature, a patentee's decision on whether or not to sell his technology has been generally considered. In this setting, the patentee can also decide which technology to sell. Note that many patent policies only ask for some minimum disclosure requirements. The strategy of "which technology to sell" becomes applicable: the patentee only discloses the laggard technology when applying for patent rights, and is willing to disclose the advanced technology after receiving a high license payment. Moreover, it has been found that the success of a technical rival reduces a scientist's profitability from his own R&D.[8] When getting the license payment in the present, a patentee may sell some laggard technology to a licensee and limit his competition in the future.[9]

By having multiple R&D programmes, the probability of v (and the new innovation) being found should be larger than by having a single programme. For simplicity, I assume that the success rates of R&D projects are independent. The probability that only scientist i's ($i \in \{A, B\}$) R&D is successful (or i is the only finder of v) is $p_i(1-p_{-i})$. The probability that both A and B's R&D are successful (or both A and B are finders of v) is $p_A p_B$. Thus, the social rate of finding v is $p_A + p_B - p_A p_B$.[10]

[7] Since the R&D cost is fixed, any investment level above the fixed cost cannot enhance the successful rate of the R&D using either technology.

[8] See Noel and Schankerman (2006).

[9] An example is the Sino-Soviet Treaty of Friendship and Alliance in 1950, which includes technology "aid" from the USSR to China (but China had to pay for it). In order to maintain their military advantage and status of Big Brother in the Communism world, the USSR only provided China with their technology developed before the 1930s.

[10] I exclude the possibility that large R&D expenditures lead to by-products and additional returns. Here R&D is only conducted to find v, the social value v cannot be changed by R&D costs.

4.3.2 Scientists' Information

The social value of new innovation v is random. However, each scientist has some prior knowledge on v before he invests in his R&D. Let $\tau_i \in \{\tau_1, \tau_2\}$ $(i = A, B)$ be i's private information about the future. τ_2 is the good news about the future and τ_1 is the bad one, i.e., for any $i = A, B$, $E(v|\tau_i = \tau_2) > E(v|\tau_i = \tau_1)$.

A and B probably have disparate information about future. But τ_A and τ_B are positively correlated.[11] Without losing generality, A is assumed to be *pessimistic* about the future, while B is *optimistic*: A observes the bad news τ_1 more often than B. (τ_A, τ_B) is drawn from the distribution satisfying Assumption 4.1:

$$
\begin{aligned}
&\textit{A4.1}: \qquad \Pr(\tau_A = \tau_1, \tau_B = \tau_1) = \Pr(\tau_B = \tau_1) = 1 - q_B, \\
\Pr(\tau_A &= \tau_1, \tau_B = \tau_2) = q_B - q_A, \ \Pr(\tau_A = \tau_2, \tau_B = \tau_1) = 0 \text{ and} \\
\Pr(\tau_A &= \tau_2, \tau_B = \tau_2) = \Pr(\tau_A = \tau_2) = q_A, \text{ where } 0 < q_A < q_B < 1.
\end{aligned}
$$

For simplicity, nature *never* selects $(\tau_A = \tau_2, \tau_B = \tau_1)$. Under *A4.1*, A holds complete information if $\tau_A = \tau_2$, but his information is partial if $\tau_A = \tau_1$. Similarly, B's information is complete if he observes τ_1, but it is partial otherwise. Moreover, $q_B - q_A = \Pr(\tau_A \neq \tau_B)$ measures the correlation between two scientists' *ex ante* perspectives. They are less correlated with a larger $q_B - q_A$ than with a smaller one.

A more general setting is for all $j, k \in \{1, 2\}$,

$$\frac{q_{12}}{q_{11}} < \frac{q_{22}}{q_{21}}, \text{ where } q_{j,k} = \Pr(\tau_{-h} = \tau_k | \tau_h = \tau_j).$$

A4.1 describes this general setting with an additional coordinate $h \in \{A, B\}$ and only excludes some cases where the patentee holds the partial information.[12] I argue that *A4.1* is sufficiently generalized. Since h designs the contracts, he can induce $-h$ to truthfully reveal τ_{-h}. Therefore, the general setting cannot quantitatively change the results .

[11] Because scientists have the common idea of improvement, their expectations of the value should be correlated.

[12] The general setting assumes that when $\tau_h = \tau_j$, h may not certainly know τ_{-h}, for any $j \in \{1, 2\}$. For the licensee's knowledge of τ_h after observing τ_{-h}, *A4.1* includes every possibility: i) licensee knows $\tau_{-h} = \tau_1$, but does not know τ_h, i.e., $-h = B$; ii) licensee knows $\tau_{-h} = \tau_2$, but does not know τ_h, i.e., $-h = A$; iii) licensee knows $\tau_{-h} = \tau_1$, then he certainly knows $\tau_h = \tau_1$, i.e, $-h = B$; iv) licensee knows $\tau_{-h} = \tau_2$, then he certainly knows $\tau_h = \tau_2$, i.e, $-h = A$. Since τ_A and τ_B are positively correlated, there is no fifth case where licensee knows $\tau_{-h} = \tau_j$, then he certainly knows $\tau_h = \tau_{-j}$, for any $j \in \{1, 2\}$.

4.3.3 Value of New Innovation and Socially Optimal R&D Costs

Nature selects $v \in (-V, +\infty)$ according to a publicly known conditional cumulative distribution $F(v|\tau_A, \tau_B)$, where V is a strictly positive number.[13] For any τ_A and τ_B, the density function $f(v|\tau_A, \tau_B)$ is strictly positive for all $v \in (-V, +\infty)$. Thus, v cannot be used as *ex post* evidence for (τ_A, τ_B). A scientist immediately commercializes the new innovation, if he succeeds in his R&D programme, and he will get the whole social value of new innovation v, if he is its unique finder. If both scientists find v, they will share v equally.[14]

A *common-value* situation is considered. The *socially optimal R&D costs* (or R&D technology usage) are assumed as follows:

i) if $\tau_A = \tau_B = \tau_2$, the common valuation of the innovation $E(v|\tau_A, \tau_B)$ is the highest, then socially optimal investments are $c_A^o = c_B^o = c$, i.e., both A and B use the advanced technology;

ii) if $\tau_A = \tau_1$ and $\tau_B = \tau_2$, $E(v|\tau_A, \tau_B)$ is modest, then the first-best outcome is $c_i^o = 0$ and $c_{-i}^o = c$, for any $i = \{A, B\}$, i.e, one scientist uses the advanced technology and the other uses the laggard one;

iii) if $\tau_A = \tau_B = \tau_1$, $E(v|\tau_A, \tau_B)$ is lowest, then social optimal outcome requires $c_A^o \neq c$ and $c_B^o \neq c$, i.e., neither A nor B adopts the advanced technology.[15]

Consider that τ_A and τ_B are observable to a risk-neutral social planner (she). Remember that $(p_A + p_B - p_A p_B)$ is the social rate of finding v by investing c_A and c_B. Her problem is

$$\max_{\substack{(c_A, c_B) \\ \in \{0,c\}\times\{0,c\}}} SW = (p_A + p_B - p_A p_B)E(v|\tau_A, \tau_B) - (c_A + c_B).$$

The solution to the above problem is $c_A = c_B = c$ only when

$$\begin{aligned} (2\overline{p} - \overline{p}^2)E(v|\tau_A, \tau_B) - 2c &\geq (\overline{p} + \underline{p} - \overline{p}\underline{p})E(v|\tau_A, \tau_B) - c \\ \text{or } E(v|\tau_A, \tau_B) &\geq \frac{c}{(1-\overline{p})\Delta p}, \end{aligned}$$

[13] It could be a purely bad influence.

[14] Since scientists are risk neutral, it is equivalent to assume that scientists apply for patent protection over the new innovation in the future. Note that the strong protection also allows the patent winner to get the whole social value v. If v is found by both scientists, they may have equal oppotunity to get the new patent.

[15] This paper is centered on the fact that scientists may efficiently adjust their R&D investment strategies according to their information.

where $(1-\overline{p})\Delta p E(v|\tau_A, \tau_B)$ is the marginal social benefit by using advanced technology for a second time; it is $c_i = 0$ and $c_{-i} = c$, $i = A$ or B only when

$$\frac{c}{(1-\underline{p})\Delta p} \leq E(v|\tau_A, \tau_B) < \frac{c}{(1-\overline{p})\Delta p},$$ [16]

where $(1-\underline{p})\Delta p E(v|\tau_A, \tau_B)$ is the marginal social benefit by using one advanced technology, given the laggard technology being used once; and it is characterized by $c_A \neq c$ and $c_B \neq c$ only when $E(v|\tau_A, \tau_B)$ is small enough. Thus, I assume that $E(v|\tau_A, \tau_B)$ satisfying Assumption 4.2:

$$\begin{aligned} \mathit{A4.2} \quad &: \quad E(v|\tau_1, \tau_1) = 0, \\ v_1 \quad &= \quad E(v|\tau_1, \tau_2) \in \left(\frac{c}{(1-\underline{p})\Delta p}, \frac{c}{(1-\overline{p})\Delta p}\right) \text{ and} \\ v_2 \quad &= \quad E(v|\tau_2, \tau_2) \geq \frac{c}{(1-\overline{p})\Delta p}. \end{aligned}$$

Here $E(v|\tau_1, \tau_1)$ is normalized to 0. With $\tau_A = \tau_B = \tau_1$, it is the first best that only patentee h conducts R&D (and uses the laggard technology). However, if at least one scientist observes the good news τ_2, *A4.2* implies that allowing multiple R&D projects is more efficient than a single one. As a matter of fact, anticommons (e.g., licensee's R&D being rejected) may not be a tragedy at all, if the improvement upon previous knowledge is insignificant.

To be precise, *anticommons* refers to events where the previous knowledge is used less than efficiently. In this paper, if R&D is under-invested (the total investment in R&D being strictly less than that required for a socially optimal outcome), or if the number of project is 1 when that required by efficiency is 2, anticommons occurs.

4.3.4 Licensing Proposal and Timing

The patentee h proposes licensing contracts to the licensee $-h$. $-h$ either accepts one contract or rejects any of them. If $-h$ accepts one contract, both scientists' R&D programs are conducted. Otherwise, $-h$'s R&D is blocked and only h researches.

[16] Since $\underline{p} < \overline{p}$, we have $\frac{c}{(1-\underline{p})\Delta p} < \frac{c}{(1-\overline{p})\Delta p}$.

4.3.4.1 Signaling and Interim Belief

After observing τ_h, the patentee h first invests $c_h(\tau_h) \in \{0, c\}$ (or pre-commits to invest $c_h(\tau_h)$). $c_h(\tau_h)$ is observable, so it might be useful message about τ_h for the licensee.[17] Let $b(\tau_{-h}, c_h)$ be $-h$'s interim belief about τ_h conditional on c_h and his private information τ_{-h}. His estimation of the future (valuation of the new innovation) is $E_{-h}(v|\tau_{-h}, b(\tau_{-h}, c_h))$. If $b(\tau_{-h}, c_h) = \{\tau_j\}$, $j = 1, 2$, this estimation is

$$E_{-h}(v|\tau_{-h}, b(\tau_{-h}, c_h)) = E(v|\tau_{-h}, \tau_j).$$

It is also possible that $-h$ does not update his belief according to the signal c_h. I use $b(\tau_{-h}, c_h) = \{\tau_1, \tau_2\}$ to denote this case when $-h$'s estimation of the future is

$$\begin{aligned} E_{-h}(v|\tau_{-h}, \{\tau_1, \tau_2\}) &= \Pr(\tau_h = \tau_1|\tau_{-h})E(v|\tau_1, \tau_{-h}) \\ + \Pr(\tau_h &= \tau_2|\tau_{-h})E(v|\tau_2, \tau_{-h}). \end{aligned}$$

In equilibrium, h knows $b(\tau_{-h}, c_h)$ and $-h$ knows h knowing that.

4.3.4.2 Licensing Menu

After signaling τ_h with c_h, h proposes license contracts. A license contract specifies which technology the licensee can use. It also specifies a fee that $-h$ should pay. Intuitively, a higher licensee's estimation of future $E_{-h}(v|\tau_{-h}, b(\tau_{-h}, c_h))$ leads to a higher license fee. I consider a license menu $\{(f_j, c_{-h,j}), j = 1, 2\}$, which allows h to incentive $-h$ to reveal his private information τ_{-h}, where f_j is the license fee and $c_{-h,j} \in \{0, c\}$ indicates the technology that $-h$ is authorized to adopt. The subscript j allows the strategy of selling information-dependent technology to be considered.

4.3.4.3 Timing

The timing of this game is as follows:

In the present period:

stage 1, Nature selects τ_h, τ_{-h} and v.

stage 2, h invests $c_h(\tau_h)$ in his R&D and sets a license menu $\{(f_j, c_{-h,j}),\ j = 1, 2\}$.

[17]Note that τ_h is soft information. Any saying about τ_h might be a cheap talk without pre-investment.

stage 3, $-h$ decides whether to take one license contract in the menu and which one to take. If he takes $(f_j, c_{-h,j})$, he pays license fee f_j to h and invests $c_{-h,j}$ in his own R&D (or uses the technology which requires $c_{-h,j}$ as fixed cost).

In the future period:

stage 4, a scientist who finds new innovation through his R&D, commercializes it. If he is the only finder, he gets v; if the other scientist also finds it, he gets $\frac{1}{2}v$.

In equilibrium, the patentee's strategy for signaling and licensing must be compatible with each other. Thus, the equilibrium interim belief $b(\tau_{-h}, c_h)$ about the patentee's private information (conditional on the signaling c_h) cannot be changed by the proposed license menu.

4.3.5 Scientists' Profits

If licensee $-h$ rejects the menu, his profit is $\pi_{-h} = 0$. Now suppose $-h$ takes the contract $(f_j, c_{-h,j})$ at *stage 3*. He becomes the only finder and secures the whole social value v with probability $p_{-h}(1-p_h)$. The probability that he shares v equally with patentee h is $p_h p_{-h}$. Since v is unknown, $-h$ relies on his estimation of future $E_{-h}(v|\tau_{-h}, b(\tau_{-h}, c_h))$. Given τ_{-h}, c_h, $c_{-h,j}$ and f_j, the licensee's expected profit at *stage 3* is

$$\begin{aligned} & \pi_{-h}(\tau_{-h}, c_h, c_{-h,j}, f_j) \\ = \; & (p_{-h}(1-p_h) + \frac{1}{2}p_h p_{-h})E_{-h}(v|\tau_{-h}, b(\tau_{-h}, c_h)) - c_{-h,j} - f_j \\ = \; & p_{-h}(1 - \frac{1}{2}p_h)E_{-h}(v|\tau_{-h}, b(\tau_{-h}, c_h)) - c_{-h,j} - f_j, \end{aligned} \tag{4.1}$$

where for any $i \in \{h, -h\}$, p_i is R&D's successful rate of scientist i ($i = h, -h$) and satisfies

$$p_i = \overline{p} \text{ if } c_i = c \text{ or } c_{i,j} = c \text{ and } p_i = \underline{p} \text{ if } c_i = 0 \text{ or } c_{i,j} = 0. \tag{4.2}$$

$c_{-h,j} + f_j$ ($j = 1, 2$) is $-h$'s total cost. Truthtelling is a licensee's dominate strategy when the *incentive compatibility constraint* is satisfied:

$$(\text{IC}_j)\text{: } \pi_{-h}(\tau_j, c_h, c_{-h,j}, f_j) \geq \pi_{-h}(\tau_j, c_h, c_{-h,-j}, f_{-j}). \tag{4.3}$$

Moreover, $-h$'s R&D may be blocked by some license menu, when h sets f_j as unaffordable for $-h$. Equivalently, the following *individual rationality constraint* is *not* satisfied:

$$(\text{IR}_j)\text{: } \pi_{-h}(\tau_j, c_h, c_{-h,j}, f_j) \geq 0. \tag{4.4}$$

Three blocking strategies of patentee h should be considered:

1. *full blockade*. h absolutely hinders $-h$'s R&D (i.e., neither (IR_1) nor (IR_2) holds). In this case, h will be the only one who does research. But he can neither know $-h$'s information, nor collect the license fee. Thus, h's expected profit at *stage 2* is his estimation of future timing his R&D success rate minus his R&D cost:

$$\pi_h^K(\tau_h, c_h, c_{-h,1}, f_1, c_{-h,2}, f_2) = p_h E(v|\tau_h) - c_h, \tag{4.5}$$

where function p_h is given by formula (4.2) and superscript K indicates full blocking. Since h only knows τ_h, $E(v|\tau_h)$ is his estimation of the future valuation of the new innovation.

2. *contingent blockade*. $-h$'s research activity is blocked only when he receives the bad news τ_1 (i.e., (IC_2) and (IR_2) hold, but (IR_1) does not).[18] In equilibrium, $-h$ with good news τ_2 will not be blocked, if h does not create a *full blockade*. Note that $\tau_{-h} = \tau_2$ leads to a higher $-h$'s estimation of future than $\tau_{-h} = \tau_1$. $-h$ with higher estimation of future can be charged for a higher license fee. When *full blockade* is off the table, patentee h seizes a good chance of collecting license incomes. And a higher license fee with $\tau_{-h} = \tau_2$ is better than a lower fee with $\tau_{-h} = \tau_1$. As the result, a *contingent blockade* will only block $-h$'s R&D with $\tau_{-h} = \tau_1$ in equilibrium. After $-h$ is licensed or not, h knows τ_{-h} and collects license fee f_2 if $\tau_{-h} = \tau_2$. Hence that h's expected profit at *stage 2* is

$$\begin{aligned} \pi_h^k(\tau_h, c_h, c_{-h,1}, f_1, c_{-h,2}, f_2) &= \Pr(\tau_{-h} = \tau_2|\tau_h)[p_h(1 - \frac{1}{2}p_{-h})E(v|\tau_h, \tau_2) \\ +f_2] + \Pr(\tau_{-h} &= \tau_1|\tau_h)p_h E(v|\tau_h, \tau_1) - c_h, \end{aligned} \tag{4.6}$$

where functions p_h and p_{-h} are given by formula (4.2) and superscript k stands for this contingent blocking. The patentee correctly revises his belief about τ_{-h} after observing the licensee action. Then $p_h E(v|\tau_h, \tau_1) - c_h$ is h's expected profit at *stage 3* when $-h$ is blocked (i.e., $\tau_{-h} = \tau_1$). $p_h(1 - \frac{1}{2}p_{-h})E(v|\tau_h, \tau_2) + f_2 - c_h$ is that when $-h$ pays license fee f_2 and invests $c_{-h,2}$ as his R&D cost (i.e., $\tau_{-h} = \tau_2$).

3. *non blockade*. $-h$ takes the contract $(f_j, c_{-h,j})$ if his information $\tau_{-h} = \tau_j$ (i.e., (IC_j) and (IR_j) hold for all $j \in \{1, 2\}$). Then h's expected profit at *stage 2* is

$$\begin{aligned} \pi_h(\tau_h, c_h, c_{-h,1}, f_1, c_{-h,2}, f_2) &= \Pr(\tau_{-h} = \tau_1|\tau_h)[p_h(1 - \frac{1}{2}p_{-h})E(v|\tau_h, \tau_1) + f_1] \\ + \Pr(\tau_{-h} &= \tau_2|\tau_h)[p_h(1 - \frac{1}{2}p_{-h})E(v|\tau_h, \tau_2) + f_2] - c_h, \end{aligned} \tag{4.7}$$

[18] Holding of (IC_1) is not necessary here, if $\pi_{-h}(\tau_1, c_h, c_{-h,2}, f_2)$ is strictly negative. $-h$ will not take contract $(f_2, c_{-h,2})$ given $\tau_{-h} = \tau_1$, even when $\pi_{-h}(\tau_1, c_h, c_{-h,2}, f_2) > \pi_{-h}(\tau_1, c_h, c_{-h,1}, f_1)$.

where functions p_h and p_{-h} is given by formula (4.2) and $p_h(1-\frac{1}{2}p_{-h})\ E(v|\tau_h,\tau_j)+f_j-c_h$ is h's expected profit at *stage 3* given $-h$ accepts $(f_j, c_{-h,j})$.

4.4 Equilibrium

In this section, I first identify the equilibrium when the *pessimistic* scientist A is the patentee, i.e., $h = A$. Then I solve the equilibrium of the other case: $h = B$. Nevertheless, it is convenient to introduce some general results here. For any patentee $h \in \{A, B\}$, the following 2 lemmas hold.

Lemma 4.1 *Under A4.2, if license menu $\{(f_j, c_{-h,j}),\ j = 1, 2\}$ satisfies the incentive compatibility constraints (IC_1) and (IC_2) and $f_1 < f_2$, then it satisfies $c_{-h,1} = 0$ and $c_{-h,2} = c$.*

Proof. From formula (4.1), the licensee's expected profit $\pi_{-h}(\tau_{-h}, c_h, c_{-h,j}, f_j)$ $(j = 1, 2)$ equals to

$$p_{-h}(1-\frac{1}{2}p_h)E_{-h}(v|\tau_{-h}, b(\tau_{-h}, c_h)) - c_{-h,j} - f_j.$$

Suppose $c_{-h,1} = c_{-h,2}$. Note that $-h$'s expected profit is decreasing with f_j. For any $c_h \in \{0, c\}$, the *incentive compatibility constraint* (IC_j) cannot hold (i.e., $\pi_{-h}(\tau_j, c_h, c_{-h,-j}, f_{-j}) > \pi_{-h}(\tau_j, c_h, c_{-h,j}, f_j)$) given $f_j > f_{-j}$. In other words, the licensee prefers to pay the lower license fee provided that his contracted R&D cost is fixed.

Now suppose $c_{-h,1} = c$ and $c_{-h,2} = 0$. (IC_1) and (IC_2) can be rewritten as

$$\begin{aligned} f_2 - f_1 &\geq c - \Delta p(1-\frac{1}{2}p_h)E_{-h}(v|\tau_1, b(\tau_1, c_h)) \\ f_2 - f_1 &\leq c - \Delta p(1-\frac{1}{2}p_h)E_{-h}(v|\tau_2, b(\tau_2, c_h)). \end{aligned}$$

Note that the licensee $-h$'s estimation of the future should be higher when he receives the good news τ_2 than when his information is τ_1, i.e., for any $c_h \in \{0, c\}$,

$$E_{-h}(v|\tau_2, b(\tau_2, c_h)) > E_{-h}(v|\tau_1, b(\tau_1, c_h)).$$

The reason behind the above formula is that when compared to the bad news, the licensee's good news leads to 1) higher expectation of future, since *A4.2* implies $E_{-h}(v|\tau_2, b) > E_{-h}(v|\tau_1, b)$ for any $b \subset \{\tau_1, \tau_2\}$; 2) higher probability that the patentee also receives a good news, since the positive correlation between τ_A and τ_B implies $\Pr(b(\tau_2, c_h) = \{\tau_2\}) \geq \Pr(b(\tau_1, c_h) = \{\tau_1\})$, for any $c_h \in \{0, c\}$. As the result, at least one of (IC_1) and (IC_2) cannot hold. □

Lemma 4.2 *Under A4.2, h invests $c_h = c$ in equilibrium whenever $\tau_h = \tau_2$.*

Proof. Under *A4.2*, $\tau_h = \tau_2$ and h's estimation of future is large enough that it is more profitable to invest heavily in his R&D programme for each blocking strategy. Specifically, *A4.1* implies i) $\overline{p}v_1 - c > \underline{p}v_1$, ii) $\overline{p}E(v|\tau_2) - c > \underline{p}E(v|\tau_2)$, iii) $\overline{p}(1-\frac{1}{2}\underline{p})v_1 - c > \underline{p}(1-\frac{1}{2}\underline{p})v_1$ and iv) $\overline{p}(1-\frac{1}{2}\overline{p})v_2 - c > \underline{p}(1-\frac{1}{2}\overline{p})v_2$.

If h chooses a *full blockade*, investing c takes priority over investing 0 according to his profit function (formula (4.5)) and the above inequality ii).

Consider the case of *contingent blockade*. Formula (4.6) together with inequality i), iii) and iv) can prove that for any $c_{-h,2} \in \{0, c\}$ and $f_2 \in R^+$, $c_h = c$ leads to a strictly higher $\pi_h^k(\tau_2, c_h, c_{-h,1}, f_1, c_{-h,2}, f_2)$ than $c_h = 0$.

Now suppose that the license menu satisfies (IC_1), (IR_1), (IC_2) and (IR_2), and there is *non blockade*. Formula (4.7), inequality iii) and iv) as well as lemma 4.1 implies that h achieves more profit by $c_h = c$ than $c_h = 0$, for any f_1, $f_2 \in R^+$. □

Remember that holding of both (IC_1) and (IC_2) is only required by a license menu with purpose of *non blockade* (see subsection 4.3.3). Summing up lemma 4.1 and 4.2, they suggest that without any blockade, each scientist invests heavily in his R&D (or uses the advanced technology) when he receives the good news from nature, i.e., in equilibrium $c_i = c$ when $\tau_i = \tau_2$, for any $i = h$ or $-h$). Under *A4.2*, the socially optimal R&D investments are i) $c_h = c_{-h} = c$, if $\tau_h = \tau_{-h} = \tau_2$; ii) $c_i = 0$ and $c_{-i} = c$, $i = h$ or $-h$, if $\tau_h \neq \tau_{-h}$; iii) neither h nor $-h$ invests c in their R&Ds, if $\tau_h = \tau_{-h} = \tau_1$ (see subsection 3.2). As a result, total investment in R&Ds is at least as large as the socially optimal level, if there is *non blockade*. Consequently, only when the patentee h creates a *full blockade* or a *contingent blockade*, there are circumstances with under-investment (i.e., total investment level being strictly less than efficient level). Furthermore, lemma 2 does not reject the possibility that h may invest c in his R&D when his information is τ_1. Thus, there may be cases with over-investment (i.e., total investment level being strictly larger than efficient level).

4.4.1 Equilibrium with Pessimistic Patentee

Consider $h = A$ and $-h = B$. The patentee A can encourage B to reveal τ_B through licensing proposal. But the licensee B makes his belief $b(\tau_B, c_A)$ about τ_A based on the message c_A. The positive correlation between these two scientists' information also impacts B's belief. Specifically, I have the following lemma.

Lemma 4.3 *Given A4.1 and A4.2, B's equilibrium interim belief about τ_A satisfies*

(i) $b(\tau_B, 0) = \{\tau_1\}$ *for any* $\tau_B \in \{\tau_1, \tau_2\}$*;*
(ii) $b(\tau_1, c_A) = \{\tau_1\}$ *for any* $c_A \in \{0, c\}$*;*
(iii) $b(\tau_2, c) \in \{\{\tau_1, \tau_2\}, \{\tau_2\}\}$*.*

Proof. Lemma 4.2 states that A invests $c_A = c$ in equilibrium whenever $\tau_A = \tau_2$. Thus, B must believe $\tau_A = \tau_1$ in equilibrium whenever he observes $c_A = 0$, i.e., $b(\tau_B, 0) = \{\tau_1\}$ for any $\tau_B \in \{\tau_1, \tau_2\}$.

Note that *A4.1* implies that B knows A receives the bad news τ_1 if $\tau_B = \tau_1$, i.e., $\Pr(\tau_A = \tau_1|\tau_B = \tau_1) = 1$. Thus, B's belief in this circumstance is $b(\tau_1, c_A) = \{\tau_1\}$ for any $c_A \in \{0, c\}$.

According to lemma 4.2, B knows that A always invests heavily with $\tau_A = \tau_2$ in equilibrium. Given $\tau_B = \tau_2$ and $c_A = c$, B cannot exclude the possibility of $\tau_A = \tau_2$. Thus, $b(\tau_2, c)$ should not be $\{\tau_1\}$. □

If nature selects $\tau_A = \tau_2$, A uses the advanced technology (i.e., $c_A = c$) and will succeed in his R&D with probability $\overline{p}$ (see lemma 4.2). Remember that A is the pessimistic scientist. $\tau_A = \tau_2$ implies that B also receives the good news τ_2, i.e., $\Pr((\tau_B = \tau_2|\tau_A = \tau_2) = 1$. Allowing B to research as if $\tau_B = \tau_1$ cannot be in equilibrium. So A does not offer a *non blockade* license menu to B. Instead, A may create a *full blockade* by setting license fee $f_1 = f_2 = \infty$ (or large enough). In this case, his expected profit (formula (4.5)) can be rewritten as

$$\pi_A^K = \overline{p}E(v|\tau_A = \tau_2) - c = \overline{p}v_2 - c, \tag{4.8}$$

since $E(v|\tau_A = \tau_2) = E(v|\tau_2, \tau_2) = v_2$. Or A may create a *contingent blockade*: he sets f_2 such that the *individual rationality constraint* (IR_2) holds (i.e., $\pi_B(\tau_2, c_A, c_{B,2}, f_2) \geq 0$) and sets $f_1 = \infty$ (or large enough) such that the *incentive compatibility constraint* (IC_2) (i.e., $\pi_B(\tau_2, c_A, c_{B,2}, f_2) \geq \pi_B(\tau_2, c_A, c_{B,1}, f_1)$) holds. His expected profit (formula (4.6)) is

$$\pi_A^k = \overline{p}(1 - \frac{1}{2}p_B)v_2 + f_2 - c. \tag{4.9}$$

It is optimal for him to leave B with 0 profit and make (IR$_2$) binding (i.e., $\pi_B(\tau_2, c_A, c_{B,2}, f_2) = 0$). Thus, A sets the license price

$$\begin{aligned} f_2 &= \pi_B(\tau_2, c_A, c_{B,2}, f_2) + f_2 \\ &= p_B(1 - \frac{1}{2}\overline{p})E_B(v|\tau_2, b(\tau_2, c)) - c_{B,2}, \end{aligned} \tag{4.10}$$

where $\pi_B(\tau_2, c_A, c_{B,2}, f_2)$ is given in formula (4.1). Under *A4.2*, the expected social welfare is maximized if B also conducts his R&D with $c_{B,2} = c$. However, the equilibrium may not be efficient. Whether B's R&D is fully blocked or conducted with the laggard (i.e., $c_{B,2} = 0$) or advanced (i.e., $c_{B,2} = c$) technology depends on $E_B(v|\tau_2, b(\tau_2, c))$. Specifically, when selling B a permission for conducting R&D in the present period, A has to face the possibility that he shares v with B in the future. This leads to a loss on A's expected profit. This loss must be exceeded by the license fee, if the license is issued. By setting f_2 as in formula (4.10), the license fee decreases with $c_{B,2}$, if $E_B(v|\tau_2, b(\tau_2, c))$ is relatively small. Thus proposing $c_{B,2} = c$ may not be optimal for A. Through combining formula (4.8), (4.9) and (4.10), one can calculate that in equilibrium, B's R&D is

$$\begin{aligned} &\text{i) fully blocked if } (1 - \tfrac{1}{2}\overline{p})E_B(v|\tau_2, b(\tau_2, c)) < \tfrac{1}{2}\overline{p}v_2; \\ &\text{ii) with } c_{B,2} = 0 \text{ if } (1 - \tfrac{1}{2}\overline{p})E_B(v|\tau_2, b(\tau_2, c)) \in [\tfrac{1}{2}\overline{p}v_2, c/\Delta p + \tfrac{1}{2}\overline{p}v_2); \\ &\text{iii) with } c_{B,2} = c \text{ if } (1 - \tfrac{1}{2}\overline{p})E_B(v|\tau_2, b(\tau_2, c)) \geq c/\Delta p + \tfrac{1}{2}\overline{p}v_2. \end{aligned} \tag{4.11}$$

From lemma 4.3, it must be $b(\tau_2, c) \in \{\{\tau_1, \tau_2\}, \{\tau_2\}\}$. If $b(\tau_2, c) = \{\tau_2\}$, B's estimation is $E_B(v|\tau_2, b(\tau_2, c)) = v_2$. Under *A4.2*, the following inequality holds:

$$(1 - \frac{1}{2}\overline{p})v_2 \geq c/\Delta p + \frac{1}{2}\overline{p}v_2.$$

From formula (4.11), B is allowed to invest $c_{B,2} = c$ in his R&D. If $b(\tau_2, c) = \{\tau_1, \tau_2\}$, B's estimation of future is

$$\begin{aligned} E_B(v|\tau_2, \{\tau_1, \tau_2\}) &= \Pr(\tau_A = \tau_2|\tau_A = \tau_2)v_2 + \Pr(\tau_A = \tau_1|\tau_B = \tau_2)v_1 \\ &= \frac{q_A}{q_B}v_2 + \frac{q_B - q_A}{q_B}v_1, \end{aligned}$$

as defined in subsection 3.4. The equilibrium license menu depends on the *ex ante* correlation between τ_A and τ_B. I will address this later in this subsection.

In equilibrium, $b(\tau_2, c) = \{\tau_2\}$ or $\{\tau_1, \tau_2\}$ depends on $c_A(\tau_1)$, the patentee's R&D investment strategy when he receives the bad news. It is $b(\tau_2, c) = \{\tau_2\}$ only when given $b(\tau_2, c) = \{\tau_2\}$, A prefers $c_A(\tau_1) = 0$ to the heavy investment c. Oterwise it is $b(\tau_2, c) = \{\tau_1, \tau_2\}$. Suppose nature selects $\tau_A = \tau_1$. According to lemma 4.3, the licensee

B's belief is $b(\tau_B, 0) = \{\tau_1\}$ for any $\tau_B \in \{\tau_1, \tau_2\}$. Thus, A reveals his bad news by using the laggard technology and investing $c_A = 0$. If $c_A(\tau_1) = 0$ is in equilibrium, the license menu should create a *contingent blockade*. When offered with such a license menu, B truthfully discloses his information through taking the license contract $(f_2, c_{B,2})$ only if $\tau_B = \tau_2$. If $\tau_A = \tau_1$ and $\tau_B = \tau_2$ are disclosed at *stage 2* and *3* respectively, both scientists update their beliefs. The expected social welfare, as defined in subsection 3.3, is the sum of A and B's expected profit at *stage 3*. Given $\tau_A = \tau_1$, $\tau_B = \tau_2$ and $c_A = 0$, I have $SW = \pi_A + \pi_B$, where

$$\begin{aligned}
\pi_B &= p_B(1 - \frac{1}{2}p_A)E_B(v|\tau_2, b(\tau_B, 0)) - c_{B,2} - f_2 \\
&= p_B(1 - \frac{1}{2}\underline{p})v_1 - c_{B,2} - f_2 \text{ and} \\
\pi_A &= p_A(1 - \frac{1}{2}p_B)E(v|\tau_1, \tau_2) - c_A + f_2 \\
&= \underline{p}(1 - \frac{1}{2}p_B)v_1 + f_2
\end{aligned}$$

Remember that with $\tau_A = \tau_1$ and $\tau_B = \tau_2$, the social welfare is maximized by $c_A = 0$ and $c_{B,2} = c$. A's expected profit at *stage 3* can equalize the maximum social welfare, if he sets $c_{B,2} = c$ and makes the above $\pi_B = 0$ with license fee

$$f_2 = \overline{p}(1 - \frac{1}{2}\underline{p})v_1 - c. \tag{4.12}$$

Moreover, *A4.2* assumes $E(v|\tau_1, \tau_1) = 0$. Thus, blocking B's research does not harm the social welfare if $\tau_A = \tau_B = \tau_1$. Consequently, if A reveals his bad information by investing $c_A(\tau_1) = 0$, a *contingent blockade* at *stage 2* provides A an expected profit equal to the maximum expected social welfare conditional on $\tau_A = \tau_1$. This expected profit (formula (4.6)) is

$$\pi_A^k = \frac{q_B - q_A}{1 - q_A}[(\overline{p} + \underline{p} - \overline{p}\underline{p})v_1 - c], \tag{4.13}$$

because of $c_{B,1} \in \{0, c\}$, $f_1 = \infty$ (or large enough), $c_{B,2} = c$, $\frac{q_B - q_A}{1 - q_A} = \Pr(\tau_B = \tau_2 | \tau_A = \tau_1)$ and f_2 being given by formula (4.12). Any *full* or *non blockade* license menu cannot do better.

However given $b(\tau_2, c) = \{\tau_2\}$, A may invest $c_A(\tau_1) = c$ and conceal his bad information at *stage 2*. Consider the case where B naively believes $\tau_A = \tau_2$ whenever $\tau_B = \tau_2$ and $c_A = c$. When nature reports B the good news and then A heavily invests in his R&D, B's estimation of the future is $E_B(v|\tau_2, b(\tau_2, c)) = v_2$. To mimic $\tau_A = \tau_2$, A also needs to set $c_{B,2} = c$ and license fee f'_2 satisfying formula (4.10), which is

$$\begin{aligned} f'_2 &= \overline{p}(1-\frac{1}{2}\overline{p})E_B(v|\tau_2,\{\tau_2\}) - c \\ &= \overline{p}(1-\frac{1}{2}\overline{p})v_2 - c. \end{aligned} \tag{4.14}$$

This license fee f'_2 is only collectable by A when $\tau_B = \tau_2$. But $c_A(\tau_1) = c$, as the cost of concealing $\tau_A = \tau_1$, is always sunk. A's expected profit at *stage 2* is

$$\begin{aligned} \pi'_A &= \frac{q_B - q_A}{1-q_A}[\overline{p}(1-\frac{1}{2}\overline{p})E(v|\tau_A = \tau_1, \tau_B = \tau_2) + f'_2] - c \\ &= \frac{q_B - q_A}{1-q_A}[\overline{p}(1-\frac{1}{2}\overline{p})(v_1+v_2) - c] - c. \end{aligned}$$

Thus, $c_A(\tau_1) = 0$ and $b(\tau_2, c) = \{\tau_2\}$ are in equilibrium if and only if π^k_A in formula (4.13) is larger than π'_A, or

$$q_B - q_A \leq s_1 = \frac{(1-q_A)c}{\Delta p(1-\overline{p})v_1 + \overline{p}(1-\frac{1}{2}\overline{p})(v_2 - v_1)}. \tag{4.15}$$

Proposition 4.1 summarizes the above analysis.

Proposition 4.1. *Let s_1 be given in formula (4.15). Given A4.1, A4.2, $q_B - q_A \leq s_1$ and A being the patentee, the equilibrium is characterized as follows:*

1. *A invests $c_A = c$ if and only if $\tau_A = \tau_2$.*
2. *B believes $\tau_A = \tau_2$ if $c_A = c$ and believes $\tau_A = \tau_1$ if otherwise.*
3. *B researches only when $\tau_B = \tau_2$; he invests $c_B = c$ when he researches.*

Note that the equilibrium satisfies i) $c_A = c_{B,2} = c$, if $\tau_A = \tau_B = \tau_2$; ii) $c_A = 0$ and $c_{B,2} = c$, if $\tau_A = \tau_1$ and $\tau_B = \tau_2$; iii) $c_A = 0$ and B does not conduct R&D, if $\tau_A = \tau_B = \tau_1$. Thus, the socially optimal outcome is implemented in this case.

remark 4.1 Under the assumptions of proposition 4.1, the socially optimal outcome is implemented.

If $q_B - q_A \leq s_1$, the correlation between A and B's information about future is relatively large (i.e., $\tau_A = \tau_B$ occurs frequently enough). It is difficult for scientists to conceal private information from each other. Not surprisingly, this game leads to a full disclosure of information under these circumstances. Here a *contingent blockade* does not harm the efficiency, because *A4.1* and *A4.2* imply that given $\tau_B = \tau_1$, one only expects a zero value generated in the future, i.e., $E(v|\tau_B = \tau_1) = E(v|\tau_1, \tau_1) = 0$. There is no welfare loss if B's R&D is blocked only when he receives the bad news. Moreover, $q_B - q_A \leq s_1$

implies

$$\frac{q_B - q_A}{1 - q_A} \leq \frac{c}{\Delta p(1-\overline{p})v_1 + \overline{p}(1-\frac{1}{2}\overline{p})(v_2 - v_1)}.$$

Under *A4.2*, the fixed cost of advanced technology c is strictly larger than $\Delta p(1-\overline{p})v_1$. If v_2 is close enough to v_1 (or both v_2 and v_1 are close enough to $\frac{c}{\Delta p(1-\overline{p})}$), $q_B - q_A \leq 1 < s_1$ always holds. Consequently, the equilibrium always implements the efficient investment level. Intuitively, v_2 represents the best future and v_1 stands for the modest one. If the best future is not significantly better than the modest one, the license fee $f_2' = \overline{p}(1-\frac{1}{2}\overline{p})v_2 - c$ (formula (4.14)) is not sufficient enough for A to conceal bad news (i.e., $c_A(\tau_1) = c$).

corollary 4.1 *Given A being the patentee, patent rights always lead to the socially optimal outcome, if v_2 is close enough to v_1.*

In other words, if a second good news does not significantly enhance the common valuation, the mechanism of patent rights may be efficient. Common-value models usually fail in efficiency because of the social cost of informational disclosure.[19] However, when the scientists' valuation of the future depends little on the others' information, neither the licensee nor the patentee is very interested in concealing information. The cost of informational disclosure is no longer a problem.

Now suppose $q_B - q_A > s_1$ (and $s_1 < 1$). Given $b(\tau_2, c) = \{\tau_2\}$, it is more profitable for A to invest $c_A(\tau_1) = c$ than $c_A(\tau_1) = 0$ in this case. Now in equilibrium, $c_A = c$ cannot be sufficient for B to believe $\tau_A = \tau_2$, i.e., $b(\tau_2, c) = \{\tau_1, \tau_2\}$. Given $c_A = c$, $\tau_B = \tau_2$ and $b(\tau_2, c) = \{\tau_1, \tau_2\}$, B's estimation of future is $E_B(v|\tau_2, \{\tau_1, \tau_2\}) = \frac{q_A}{q_B}v_2 + \frac{q_B - q_A}{q_B}v_1$. From the above analysis (especially formula (4.11)), B's R&D is fully blocked if

$$q_B - q_A > s_2 = \frac{(1-\overline{p})v_2}{(1-\frac{1}{2}\overline{p})(v_2 - v_1)} > s_1. \tag{4.16}$$

It must be $s_2 > s_1$, because of *A4.1*. B's R&D should be restricted with laggard technology and $c_{B,2} = 0$ if and only if

$$\max\{s_1, s_3\} < q_B - q_A \leq s_2 \text{ and } s_3 = \frac{(1-\overline{p})v_2 - c/\Delta p}{(1-\frac{1}{2}\overline{p})(v_2 - v_1)} < s_2. \tag{4.17}$$

And according to formula (4.10), B pays a license fee amounting to

$$f_2 = \underline{p}(1 - \frac{1}{2}\overline{p})(\frac{q_A}{q_B}v_2 + \frac{q_B - q_A}{q_B}v_1). \tag{4.18}$$

f_2 in formula (4.18) decreases with $q_B - q_A$. When $q_B - q_A$ is relatively large, B's estimation is close to v_1. Thus, f_2 may be too little for A compared to what he thinks

[19] See Auriol and Laffont (1992) and Maskin and Tirole (1992).

he deserves since he knows $E(v|\tau_A = \tau_2) = E(v|\tau_2, \tau_2) = v_2$. From formula (6), a patentee's expected profit decreases with the successful rate of a licensee's R&D. To compensate himself, A is willing to limit p_B to 0 and block B from research, or limit p_B to $\underline{p}$ and set $c_{B,2} = 0$. Thus, *full blockade* or inefficient *contingent blockade* (i.e., $c_{B,2} = 0$) can be equilibrium. Only when $q_B - q_A$ is relatively small, i.e.,

$$q_B - q_A \in (s_1, s_3] \text{ and } s_1 < s_3,$$

B's estimation of future is close enough to v_2 such that A will collect a license fee up to

$$f_2 = \overline{p}(1 - \frac{1}{2}\overline{p})(\frac{q_A}{q_B}v_2 + \frac{q_B - q_A}{q_B}v_1) - c. \tag{4.19}$$

At the same time, B can uses the advanced technology and invest c. Then this game leads to an efficient outcome (i.e., $c_A = c_{B,2} = c$).

Consider $\tau_A = \tau_1$. Note that given $q_B - q_A > s_1$, A will choose $c_A(\tau_1) = c$ over 0 if $b(\tau_2, c) = \{\tau_2\}$. Here given $b(\tau_2, c) = \{\tau_1, \tau_2\}$, he may still invest $c_A(\tau_1) = 0$. If he does that in equilibrium, I have already proved that he sets a *contingent blockade* with $c_{B,2} = c$ and $f_2 = \overline{p}(1-\frac{1}{2}\underline{p})v_1 - c$ (formula (4.12)). He also takes the whole and maximizes the expected social welfare. If A invests $c_A(\tau_1) = c$, he has a tendency to conceal his information τ_1 from B. Clearly, B is not easily misled. If $\tau_B = \tau_1$, B knows $\tau_A = \tau_1$ according to lemma 4.3. However, if $\tau_B = \tau_2$, B's interim belief is $b(\tau_2, c) = \{\tau_1, \tau_2\}$. Now A may conceal his bad news successfully by setting the license menu as if $\tau_A = \tau_2$. That is a *contingent blockade* with $c_{B,2} = 0$ and f_2 as in formula (4.18) if $\max\{s_1, s_3\} < q_B - q_A \leq s_2$; or with $c_{B,2} = c$ and f_2 as in formula (4.19) if $q_B - q_A \leq s_3$ and $s_1 < s_3$. The benefit of concealing the bad news is to trick B in to paying a high license fee. So A will not fully block B's R&D programme. Whether $c_A(\tau_1) = 0$ or $c_A(\tau_1) = c$ being equilibrium depends on A's expected profit. Assume $\max\{s_1, s_3\} < q_B - q_A \leq s_2$. By choosing $c_A(\tau_1) = c$, $c_{B,2} = 0$ and $f_2 = \underline{p}(1 - \frac{1}{2}\overline{p})(\frac{q_A}{q_B}v_2 + \frac{q_B - q_A}{q_B}v_1)$, A's expected profit (formula (4.6)) is

$$\begin{aligned}
\pi'_A &= \Pr(\tau_B = \tau_2|\tau_A = \tau_1)[p_A(1 - \frac{1}{2}p_B)E(v|\tau_1, \tau_2) + f_2] - c \\
&= \frac{q_B - q_A}{1 - q_A}[\overline{p}(1 - \frac{1}{2}\underline{p})v_1 + \underline{p}(1 - \frac{1}{2}\overline{p})(\frac{q_A}{q_B}v_2 + \frac{q_B - q_A}{q_B}v_1)] - c. \qquad (4.20)
\end{aligned}$$

Comparing formula (4.20) with (4.13), $c_A(\tau_1) = c$ and $c_{B,2} = 0$ are in equilibrium if and only if

$$q_B - q_A > s_4 = \frac{(1 - q_B)c}{\underline{p}(1 - \frac{1}{2}\overline{p})q_A(v_2 - v_1)} \text{ and } s_4 < s_2. \tag{4.21}$$

Now assume $q_B - q_A \leq s_3$ and $s_1 < s_3$. By setting $c_A(\tau_1) = c$, $c_{B,2} = c$ and $f_2 = \overline{p}(1 - \frac{1}{2}\overline{p})(\frac{q_A}{q_B}v_2 + \frac{q_B - q_A}{q_B}v_1) - c$, A can achieve an expected profit up to

$$\pi'_A = \frac{q_B - q_A}{1 - q_A}[\overline{p}(1 - \frac{1}{2}\overline{p})(v_1 + \frac{q_A}{q_B}v_2 + \frac{q_B - q_A}{q_B}v_1) - c] - c. \tag{4.22}$$

Comparing formula (4.22) with (4.13), $c_A(\tau_1) = c$ and $c_{B,2} = c$ are in equilibrium if and only if

$$q_B - q_A > s_5 = \frac{(1 - q_A)q_B c}{q_B \Delta p(1 - \overline{p})v_1 + q_A \overline{p}(1 - \frac{1}{2}\overline{p})(v_2 - v_1)} \text{ and } s_5 < s_3, \tag{4.23}$$

where $s_5 > s_1$ because of $q_B > q_A$. Proposition 4.2 summarizes the results for $q_B - q_A > s_1$.

Proposition 4.2. *Let s_1, s_2, s_3, s_4 and s_5 be given by formula (4.15), (16), (17), (21) and (23) respectively. Given A4.1, A4.2 and $q_B - q_A > s_1$, the equilibrium is characterized as follows:*

1. licensee B's interim belief is $b(\tau_2, c) = \{\tau_1, \tau_2\}$, thus $c_A = c$ cannot be sufficient for B to believe $\tau_A = \tau_2$.

2. if $q_B - q_A \in (\max\{s_3, s_4\}, s_2] \cup (s_5, s_3] \neq \varnothing$, then patentee A always invests $c_A = c$; otherwise, he invests $c_A = c$ only when $\tau_A = \tau_2$.

3. when $c_A = c$, B cannot research if $q_B - q_A > s_2$ or $\tau_B = \tau_1$; B invests $c_B = 0$ if $\tau_B = \tau_2$ and $q_B - q_A \in (s_3, s_2]$; he invests $c_B = c$ if $\tau_B = \tau_2$, and $q_B - q_A \in (s_1, s_3] \neq \varnothing$.

4. when $c_A = 0$, B researches only when $\tau_B = \tau_2$ and he invests $c_B = c$ when he researches.

Now patent rights may lead to inefficiency, because *ex ante* scientists' perspectives about the future are less correlated (i.e., $q_B - q_A > s_1$). The problem is that licensee B cannot trust the signal $c_A = c$, because A may heavily invest in his R&D (i.e., $c_A = c$) to conceal his bad news. So B must question the motivation of the pessimistic patent holder, who suddenly acts positively and uses the advanced technology. If A really received the good news, he may punish B for being skeptical by blocking him from research or denying his access to the advanced technology. Specifically, if $q_B - q_A > \max\{s_1, s_3\}$, R&D is conducted less than efficiency, i.e., total investments in R&D or the number of conducted activities falls below the socially optimal level. This result fits the prediction of anticommons: a patent owner may inefficiently reduce the usage of previous knowledge. If A receives the bad news, using the advanced technology and investing heavily, it is misleading to say that the new innovation will be most valuable.

A chooses to lie when $q_B - q_A \in (\max\{s_3, s_4\}, s_2] \cup (s_5, s_3] \neq \varnothing$. However, a socially optimal outcome requires that A not set $c_A = c$ in this situation. This lie may harm the social welfare by inducing over-investment outcomes (i.e., total investments are strictly less than the social optimal level). It contradicts anticommons, but does not prove the efficiency of patent rights.

Remark 4.2 Under the assumptions of proposition 4.2, R&D is used less efficiently, if $\tau_A = \tau_2$ and $q_B - q_A > \max\{s_1, s_3\}$; R&Ds are over-invested, if $\tau_A = \tau_1$ and $q_B - q_A \in (\max\{s_3, s_4\},\ s_2] \cup (s_5, s_3] \neq \varnothing$.

Note that s_3 touches its lowest boundary $\frac{1-\bar{p}}{1-\frac{1}{2}\bar{p}}$ (see formula (4.17)), when v_2 tends to infinity. At the same time, s_1 (see formula (4.15)) is arbitrarily close to 0, so $\max\{s_1, s_3\}$ is s_3. Then $q_B - q_A > \frac{1-\bar{p}}{1-\frac{1}{2}\bar{p}}$ is the sufficient condition that anticommons occurs when the patentee A has the good news. Using the continuity of s_3 as a function of v_2, one can easily deduce the following corollary.

corollary 4.2 *Given A is the patentee, there exists some large enough v_2 for any $q_B - q_A > \frac{1-\bar{p}}{1-\frac{1}{2}\bar{p}}$, such that R&D is always used less efficiently in equilibrium if $\tau_A = \tau_2$.*

Here patent rights may lead to the anticommons if a second good news significantly affects the common valuation largely. Under the same circumstances, patent rights may also lead to over-investment. Observing A's heavy investment, B cannot exclude the most profitable future (i.e., the common valuation is v_2), thus is willing to pay a relatively high licensing fee. Consequently, $c_A(\tau_1) = c$ occurs more often as v_2 grows. Proposition 4.2 suggests that $c_A(\tau_1) = c$ appears in equilibrium only when $q_B - q_A \in (\max\{s_3, s_4\}, s_2] \cup (s_5, s_3] \neq \varnothing$. From formula (4.15), (4.16), (4.17), (4.21) and (23), one can easily calculate that for any $0 < q_A < q_B < 1$, $c > 0$ and v_1 is bounded (as *A4.2*), $(\max\{s_3, s_4\}, s_2] \cup (s_5, s_3]$ contains $[0, 1]$ when v_2 is large enough. So for all $q_B - q_A > \lim_{v_2 \to \infty} s_1 = 0$, over-investment is a sure thing provided $\tau_A = \tau_1$.

Corollary 4.3 *Given A is the patentee, there exists some large enough v_2 such that R&D is always over-invested in equilibrium if $\tau_A = \tau_1$.*

After the passage of the U.S. Bayh–Dole Act of 1980, concerns arose about the potential impact of anticommons. Those criticisms quieted when universities' trend of increasing investment in patentable researches became more clear.[20] However, corollary 2 and 3 suggest that the impact of anticommons was concealed, but not insignificant, as it may be hidden behind the co-existing over-investment. The results depend on some strong common-value assumption: both scientists observing the good news τ_2 results in a large enough common valuation. Consider that scientists' information is less correlated (e.g.,

[20] In fact, this trend began in the 1970s. The passage of Bayh–Dole Act in 1980 legitimized their activities.

$q_B - q_A > \frac{1-\overline{p}}{1-\frac{1}{2}\overline{p}}$). Patent rights can never lead to an efficient outcome. On one hand, anticommons happens with *ex ante* probability $q_A = \Pr(\tau_A = \tau_B = \tau_2)$. If the patentee observes good news, R&D is used less efficiently because the licensee cannot use the advanced technology required for social optimality. On the other hand, over-investment occurs with *ex ante* probability $1 - q_A = \Pr(\tau_A = \tau_1)$. It is the patentee who invests more efficiently in R&D when he observes the bad news. The analysis only focuses on a single research project. In general, the *ex ante* co-existence of under-usage of protected knowledge and over-investment may not be negligible. In fact, universities view revenues as the most important objective of licensing activity.[21] They may over-invest in their own R&D with the hope of high license income, and in turn cause anticommons by denying licensing requests with lower fees.

4.4.2 Equilibrium with Optimistic Patentee

Consider $h = B$ and $-h = A$. As before, the patent holder B reveals his bad news τ_1 if he invests 0 in his R&D. The key question is whether the licensee A can trust on $c_B = c$ as a sufficient message of $\tau_B = \tau_2$. Suppose that A believes $\tau_B = \tau_2$ whenever B invests $c_B = c$, that is $b(\tau_A, c) = \{\tau_2\}$ for any $\tau_A \in \{\tau_1, \tau_2\}$. If $\tau_B = \tau_1$, B knows $\tau_A = \tau_1$ because *A4.1* implies $\Pr(\tau_A = \tau_1|\tau_B = \tau_2) = 1$. Additionally, *A4.2* implies that $c_B(\tau_1) = c$ cannot be recovered in the future, i.e., $E(v|\tau_1, \tau_1) = 0$. If B invests $c_B(\tau_1) = c$, B should recover his investment from the license fees. Thus, B has to use a *non blockade* license menu and collect the license fee f_1. Consequently, there are two sufficient conditions for $b(\tau_A, c) = \{\tau_2\}$: i) $c_B(\tau_1) = c$ cannot be recovered by f_1 given $b(\tau_A, c) = \{\tau_2\}$, or $f_1 < c$; ii) B is more profitable by creating a *contingent blockade* or *full blockade* than a *non blockade* given $b(\tau_A, c) = \{\tau_2\}$ and $\tau_B = \tau_2$.

According to lemma 4.1, a *non blockade* license menu should be characterized by $c_{A,1} = 0$. Given $c_B = c$, $\tau_A = \tau_1$, $b(\tau_A, c) = \{\tau_2\}$ and $c_{A,1} = 0$, A's expected profit (formula (4.1)) is

$$\begin{aligned}\pi_A(\tau_1, c_B, c_{A,1}, f_1) &= p_A(1 - \frac{1}{2}p_B)E_A(v|\tau_A, b(\tau_A, c_B)) - c_{A,1} - f_1 \\ &= \underline{p}(1 - \frac{1}{2}\overline{p})v_1 - f_1. \qquad (4.24)\end{aligned}$$

It is optimal for B to leave A with 0 profit, or the license fee is $f_1 = \underline{p}(1 - \frac{1}{2}\overline{p})v_1$. If $\underline{p}(1 - \frac{1}{2}\overline{p})v_1 < c$, investing $c_B(\tau_1) = c$ results in a negative expected profit of B. Thus, *holding of* $\underline{p}(1 - \frac{1}{2}\overline{p})v_1 < c$ *is one sufficient condition for* A *to believe* $\tau_B = \tau_2$ *provided with* $c_B = c$.

[21] See for instance, Jensen and Thursby (2001).

Now suppose that B receives the good news. B invests $c_B = c$ according to lemma 4.2. Similar to the above subsection, B does not fully block A's R&D in equilibrium, given $b(\tau_A, c) = \{\tau_2\}$. By adopting the *full blockade*, B only gets (formula (4.5))

$$\begin{aligned} \pi_B^K &= p_B E(v|\tau_B = \tau_2) - c_B \\ &= \frac{q_A}{q_B}(\overline{p}v_2 - c) + \frac{q_B - q_A}{q_B}(\overline{p}v_1 - c), \end{aligned} \tag{4.25}$$

where $\frac{q_B - q_A}{q_B} = \Pr(\tau_A = \tau_1|\tau_B = \tau_2)$, $\frac{q_A}{q_B} = \Pr(\tau_A = \tau_2|\tau_B = \tau_2)$. Using a *contingent blockade* license menu, B allows A to research and extracts his whole expected profit by f_2 if $\tau_A = \tau_2$. And A's R&D is forbidden if $\tau_A = \tau_1$. By the same logic used in subsection 4.1 (particularly formula (4.10) and (4.12)), B should gain the maximum expected social welfare if $\tau_A = \tau_B = \tau_2$: B sets $c_{A,2} = c$ and $f_2 = \overline{p}(1 - \frac{1}{2}\overline{p})v_2 - c$. And he can get the same profit as using the *full blockade* if $\tau_A = \tau_1$. Hence that B's expected profit (formula (4.6)) is

$$\pi_B^k = \frac{q_A}{q_B}((2\overline{p} - \overline{p}^2)v_2 - 2c) + \frac{q_B - q_A}{q_B}(\overline{p}v_1 - c), \tag{4.26}$$

where $(2\overline{p} - \overline{p}^2)v_2 - 2c$ is the maximum expected social welfare when $\tau_A = \tau_B = \tau_2$. Clearly, π_B^k in formula (4.26) is strictly larger than π_B^K in formula (25). Hence that *full blockade* cannot appear in any equilibrium. Consider B offers a *non blockade* license menu to A, provided with $\tau_B = \tau_2$ and $c_B = c$. In this case, B should motivate A to reveal his good news τ_2 to collect a higher license fee f_2 than f_1. (IC_1) and (IC_2) (formula (4.3)) should be satisfied by the license menu. As stated in lemma 4.1, the license menu satisfies $c_{A,1} = 0$ and $c_{A,2} = c$. (IC_1) and (IC_2) imply

$$\begin{aligned} \underline{p}(1 - \frac{1}{2}\overline{p})v_2 - f_1 &\leq \overline{p}(1 - \frac{1}{2}\overline{p})v_2 - c - f_2 \text{ or} \\ f_2 - f_1 &\leq \Delta p(1 - \frac{1}{2}\overline{p})v_2 - c \end{aligned}$$

Moreover, *non blockade* requires that (IR_1) (formula (4)) should also hold, i.e., $f_1 \leq \underline{p}(1 - \frac{1}{2}\overline{p})v_1$ (see formula (4.23)). To maximize his expected profit, B sets

$$f_1 = \underline{p}(1 - \frac{1}{2}\overline{p})v_1 \text{ and } f_2 = (1 - \frac{1}{2}\overline{p})(\underline{p}v_1 + \Delta p v_2) - c.$$

He achieves an expected profit (formula (4.7)) up to

$$\begin{aligned} \pi_B = & \frac{q_B - q_A}{q_B}((\overline{p} + \underline{p} - \overline{p}\underline{p})v_1 - c) + \\ & \frac{q_A}{q_B}((2\overline{p} - \overline{p}^2)v_2 - 2c - \underline{p}(1 - \frac{1}{2}\overline{p})(v_2 - v_1)), \end{aligned} \tag{4.27}$$

where $(\overline{p} + \underline{p} - \overline{p}\underline{p})v_1 - c$ is the maximum expected social welfare if $\tau_A = \tau_1$, $\tau_B = \tau_2$

and $\underline{p}(1-\frac{1}{2}\overline{p})(v_2-v_1)$ is the minimum rent for A to disclose his good news τ_2. If π_B in formula (4.26) is strictly less than π_B^k in formula (4.25), or

$$q_B - q_A < s_6 = q_A \frac{(1-\frac{1}{2}\overline{p})(v_2-v_1)}{(1-\overline{p})\, v_1}, \tag{4.28}$$

B creates a *contingent blockade* in equilibrium if $\tau_B = \tau_2$. The *non blockade* strategy is adopted if otherwise. From the above analysis, $q_B - q_A < s_6$ is another sufficient condition for $b(\tau_A, c) = \{\tau_2\}$.

In summary, if $\underline{p}(1-\frac{1}{2}\overline{p})v_1 < c$ or $q_B - q_A < s_6$ holds, B invests c only when he observes the good news given $b(\tau_A, c) = \{\tau_2\}$. If $\underline{p}(1-\frac{1}{2}\overline{p})v_1 < c$ but $q_B - q_A \geq s_6$, A invests c if $\tau_A = \tau_B = \tau_2$ and 0 if $\tau_A = \tau_1 \neq \tau_B$. This game generates the socially optimal outcome. If $\underline{p}(1-\frac{1}{2}\overline{p})v_1 < c$ and $q_B - q_A < s_6$, A's R&D is blocked whenever $\tau_A = \tau_1$.

Proposition 4.3. *Let s_6 be given in formula (4.27). Given A4.1, A4.2, B being the patentee and holding of at least one of $\underline{p}(1-\frac{1}{2}\overline{p})v_1 < c$ and $q_B - q_A < s_6$, the equilibrium is characterized as follows:*

1. *B invests $c_B = c$ only when $\tau_B = \tau_2$.*
2. *A believes $\tau_B = \tau_2$ whenever $c_B = c$; he believes $\tau_B = \tau_1$ when $c_B = 0$.*
3. *if $q_B - q_A < s_6$, B blocks A from research whenever $\tau_A = \tau_1$; if $q_B - q_A \geq s_6$, B blocks A from research only when $\tau_A = \tau_B = \tau_1$.*
4. *A is allowed to invest $c_A = c$ whenever $\tau_A = \tau_2$.*

Note that the social optimal outcome is characterized by $c_B = c$ and $c_A = 0$, if $\tau_A = \tau_1$ and $\tau_B = \tau_2$. But when $q_B - q_A < s_6$, the optimistic patentee blocks the other scientist from research according to proposition 4.3. The equilibrium leads to an inefficient outcome where the previous knowledge is not fully utilized and anticommons takes place. Here B reveals τ_B truthfully and knows τ_A at the end of *stage 3*. According to proposition 4.1, full disclosure leads to a socially optimal outcome, when the pessimistic scientist A owns the previous knowledge. Now A is the licensee, who may observe τ_1 given $\tau_B = \tau_2$. So he may honestly disagree with a high estimation of future value. However, any licensee may intentionally under-estimate the value of the new innovation to avoid a high license fee. To encourage a truth telling to A, B may offer a rent to A if he receives the good news. From above, the expected rent is

$$\frac{q_A}{q_B}\underline{p}(1-\frac{1}{2}\overline{p})(v_2-v_1).$$

When $q_B - q_A$ is relatively small ($q_B - q_A < s_6$) and thus, $\frac{q_A}{q_B}$ is large, this rent is large. Then B chooses an alternative approach to induce truthtelling: he threatens to disagree with any low estimation of licensee and block whoever under-estimates it. This threat is credible since he has the veto power over A's R&D. When A really receives some bad news from nature, A is blocked.

Corollary 4.4 *Given B is the patentee, there exists some large a enough v_2 such that R&D is used less efficiently if $\tau_A = \tau_1$ and $\tau_B = \tau_2$.*

If a second good news significantly enhances the common valuation, the above rent that increases with $(v_2 - v_1)$ is not affordable for the patentee B. Then anticommons may take place. Nevertheless, if v_2 is close enough to v_1, the rent is relatively small for patentee B. Then the socially optimal outcome is implemented.

Corollary 4.5 *Given B is the patentee, if v_2 is close enough to v_1 and $q_B - q_A < s_6$ holds, patent rights lead to the socially optimal outcome.*

Contrary to corollary 4.1, corollary 4.5 is not sufficient to suggest that the patent is efficient under the control of the optimistic patentee if a second good news has little influence on the common valuation. The correlation also matters. Specifically, the following analysis focuses on the case of $\underline{p}(1 - \frac{1}{2}\overline{p})v_1 \geq c$ and $q_B - q_A \geq s_6$.

Proposition 4.4. *Let s_6 is given in formula (4.28) and*

$$s_7 = \frac{(1 - q_A)c}{\underline{p}(1 - \frac{1}{2}\overline{p})v_1} \text{ and } s_8 = q_A \frac{(1 - \frac{1}{2}\overline{p})(v_2 - \frac{1-q_B}{1-q_A}v_1)}{(1 - \overline{p})\, v_1}.$$

Given A4.1, A4.2, $\underline{p}(1-\frac{1}{2}\overline{p})v_1 \geq c$, $q_B - q_A \geq s_6$ and B being the patentee, the equilibrium is characterized as follows:

1. B invests $c_B = c$ only when $\tau_B = \tau_2$ if $q_B - q_A < \max\{s_7, s_8\}$; B always invests $c_B = c$ for any $\tau_B \in \{\tau_1, \tau_2\}$ if otherwise.

2. A believes $\tau_B = \tau_2$ only when $\tau_A = \tau_2$; he believes $\tau_B = \tau_1$ whenever $c_B = 0$; or his belief is $b(\tau_A, c) = \{\tau_1, \tau_2\}$ when $\tau_A = \tau_1$.

3. if $q_B - q_A < s_8$, B blocks A from research whenever $\tau_A = \tau_1$; if $q_B - q_A \geq s_8$, B blocks A from research only when $\tau_A = \tau_B = \tau_1$.

4. A is allowed to invest $c_A = c$ whenever $\tau_A = \tau_2$; A is allowed to invest $c_A = 0$ if $q_B - q_A \geq s_8$ and $\tau_A = \tau_1$.

Proof. Provided with $\underline{p}(1 - \frac{1}{2}\overline{p})v_1 \geq c$ and $q_B - q_A \geq s_6$, A's belief $b(\tau_1, c) = \varnothing$ is in equilibrium. Taking this belief to formula (4.23), one can easily equalize the license fee when $c_B = c$ and $\tau_A = \tau_1$ to

$$f_1 = \underline{p}(1 - \frac{1}{2}\overline{p})\frac{q_B - q_A}{1 - q_A}v_1.$$

Only if $f_1 \geq c$ or equivalently $q_B - q_A \geq s_7$, B may invest $c_B = c$ when he observes $\tau_B = \tau_1$. Since $\frac{q_B - q_A}{1-q_A} < 1$ must hold, $q_B - q_A \geq s_7$ is only possible when $\underline{p}(1 - \frac{1}{2}\overline{p})v_1 \geq c$. Moreover, B also does not invest heavily given $\tau_B = \tau_1$, if he contingently blocks A's R&D given $\tau_B = \tau_2$. Using the same logic when deducing s_6, $c_B(\tau_1) = c$ may be in equilibrium only when $q_B - q_A \geq s_8$. Since $s_6 < s_8$, $q_B - q_A \geq s_8$ is only possible if $q_B - q_A \geq s_6$. Thus, the equilibrium is the same as that in proposition 3 if $q_B - q_A < \max\{s_7, s_8\}$, when A can trust on $c_B = c$ as sufficient message of $\tau_B = \tau_2$. If $q_B - q_A \geq \max\{s_7, s_8\}$, B always uses a *non blockade* license menu and invests $c_B = c$ for any $\tau_B \in \{\tau_1, \tau_2\}$. □

Remark 4.3 Under the assumptions of proposition 4.4, if $q_B - q_A \geq \max\{s_7, s_8\}$ and $\tau_B = \tau_1$, R&D is over-invested; if $q_B - q_A < s_8$ and $\tau_B = \tau_2$, R&D is used less efficiently; if $s_8 < s_7$ and $q_B - q_A \in [s_8, s_7)$, the socially optimal outcome is implemented.

Proposition 4.4 and remark 4.3 suggest that only if correlation between τ_A and τ_B is very small (i.e., $q_B - q_A$ is very large), the patent hold B may successfully conceal his bad news through investing heavily. This over-investment outcome is different to that of A being patent holder. According to proposition 4.2, A reveals his bad news by a light investment, if $q_B - q_A$ is large. When A conceals his bad information, he attempts to misguide the licensee who receives the good news. Knowing $\tau_B = \tau_2$ and $q_B - q_A$ being large, B is unlikely to believe that A also gets a positive signal from nature (i.e., $\tau_A = \tau_2$). Consequently, it is too difficult for A to conceal $\tau_A = \tau_1$ from B under this circumstance. In the next scenario, B attempts to conceal the bad news. When $\tau_A = \tau_1$ and $q_B - q_A$ are very large, A has propensity to believe $\tau_B = \tau_2$ and that a relatively large licensing fee can be recovered in the future now. This result, together with that of proposition 4.2, clearly implies that over-investment may occur quite often. Moreover, remark 4.2 suggests that anticommons occurs if scientists' information is less correlated (i.e., $q_B - q_A > \max\{s_1, s_3\}$). On the contrary, here (and in corollary 4) anticommons may also occur when scientists' information about future is relative highly correlated (i.e., $s_6 \leq q_B - q_A < s_8$). Thus, anticommons is a serious issue when the patentee and potential licensee are in a common-value situation.

Note that s_6 increases with $(v_2 - v_1)$ (see formula (4.28)). The assumptions of proposition 4.4 cannot hold (i.e., $q_B - q_A \in (0, 1)$ cannot exceed s_6), if v_2 is large enough. Consequently, a large v_2 is not the only variable causing inefficiency in patent rights. Moreover, remark 4.3 also implies that the first-best result cannot be ensured , when v_2

is close enough to v_1More subtly, let v_2 get arbitrarily close to v_1. Then s_8 is close to, but always larger than

$$\frac{q_A(q_B - q_A)(1 - \frac{1}{2}\overline{p})}{(1 - q_A)\,(1 - \overline{p})}.$$

$q_B - q_A < s_8$ always holds if

$$q_A(1 - \frac{1}{2}\overline{p}) > (1 - q_A)\,(1 - \overline{p})\,. \tag{4.29}$$

According to remark 4.3, if formula (4.29) holds and $\tau_B = \tau_2$, the equilibrium leads to anticommons. Now consider that A is highly pessimistic and the fixed cost of the advanced technology is relatively low, i.e., q_A and c are small enough. Under this circumstance, neither s_7 nor s_8 can be large enough provided v_2 is close enough to v_1. Additionally, if B is highly optimistic (i.e., q_B is large enough), $q_B - q_A \geq \max\{s_7, s_8\}$ holds. Then the over-investment may appear if $\tau_B = \tau_1$.

Corollary 4.6 *Given B is the patentee, v_2 being close enough to v_1 cannot ensure the socially optimal outcome being implemented by patent rights.*

Corollary 4.1 and 4.5 may give one the impression that as long as a second good news cannot enhance the common valuation significantly, the mechanism of patent rights is efficient. However, corollary 4.6 indicates that it cannot provide the efficiency of the mechanism. Conditions around the correlation between scientists' information are also needed. If the correlation is too small, i.e., $q_B - q_A$ is large enough, scientists' *ex ante* valuations about the future is significantly disparate. As suggested by Heller and Eisenberg (1998) (also see Barton (1995) and Gallini (2002)), it may cause anticommons: when the patent holder expects a higher value than the licensee does, he may ask for a license fee higher than that affordable for the licensee. Moreover, the above deduction implies that given some small correlation between scientist' information, R&Ds could also be over-invested. This is due to the possible informational asymmetry between patentee and licensee. When the patentee correctly figures out the licensee's information using the correlation and his own information, the licensee may not be able to do the same thing. Then the patentee, using his informational advantage, falsely signals his bad news as good. The licensee believes the signal because of the small correlation and his own bad news. As the result, a "winner's curse" of the licensee may arise in equilibrium: he pays a higher license fee than he can recover from future earnings.

4.5 Conclusion

There is much controversy over the efficiency of patent rights. On one hand, the proponents of anticommons suggest that patentees of previous knowledge may block their competitors' research for new findings when they execute their veto power granted by their patent rights. The implication is that new findings and investments in R&D could be reduced by patent protection. On the other hand, some empirical studies suggest that policies favoring patenting have not caused apparent anticommons; the number of new findings and investments in R&D has not dropped dramatically. One possible explanation for this apparent controdiction is the co-existence of anticommons and over-investment in R&D. It is the patentee and licensee's common valuation of the future finding that pulls the trigger. When the patentee's valuation largely relies on licensee's information, he may consider a license menu involving some blockade to induce truthtelling from the licensee. When the licensee's valuation largely depends on patentee's information, the patentee may over-invest in his R&D to signal a high valuation. By doing that, he may trick the licensee into paying a high license fee. Furthermore, the correlation between scientists' perspectives may also cause inefficiency. When the correlation is relatively small, scientists' *ex ante* valuations about the future is significantly disparate. Anticommons arises because the patentee places a higher value on future earnings than the licensee does, and then asks for higher fee than the licensee can afford. Over-investment may also occur, if the patentee has more complete information than the licensee. The licensee then pays a higher fee than he can recover.

Although this paper suggests patent rights are inefficient, it does not imply that society will be better off by adopting the old-fashion approach of knowledge dissemination: *open science*. In the mechanism of open science, R&D is funded by governments and knowledge is publicly owned. Scientists are usually paid by a fixed salary and motivated by the status (the psychological benefit), which will eventually be cashed out (see Auriol and Renault (2008)). Generally, with more investments in their research, scientists' studies become more successful and they gain a higher status more easily. They may corporately (or independently) report only the good news and ask for more funding. Then over-investment in R&D, as predicted by "tragedy of commons", may arise. However, patent rights may lead to a socially optimal outcome, depending on the extent of a scientist valuation, impact of the other's perspective, and the correlation between their perspectives. In this sense, patent rights may be more efficient than open science as mechanism for dissemination of scientific findings.

Appendix A

Appendix for Chapter 2

A.1 Sub-game between Policyholder and Health Plan

Policyholders are equally assigned to providers in the same platform can be explained as that the platforms are prone to reveal only the number of each type providers contracted but no more. Assume that a patient's utility that a provider could bring is $v = v_L + (v_H - v_L)I_H$. I_H is index function: $I_H = 1$, if the patient is treated by a H provider or, $I_H = 0$, if he or she is treated by a L provider. Thus all policyholders are willing to meet H providers to get a high utility.

Consider the plans disclose the quality of each provider, then all policyholders will firstly attempt to get access to H providers. Since we exclude capacity limitation, plans will only contract with H providers. However, because there are some L type providers located much closer to the plans than some H providers, hence much cheaper for the plans to contract with than some H providers. Thus the plans should optimally choose to reveal n^L and n^H without specifying each provider's type.

In front of the above plans' strategy, the policyholders' best response is to match a provider that they believe is H. Formally, assume that the probability that a provider, who is believed as H by a policyholder, is actually H is z and the probability that a provider, who is believed as H by a policyholder, is actually L is $1 - z$, where $z \geq \frac{1}{2}$. Let $v_L = \lambda$ and $v_H = 1 + \lambda$, the expected utility gain from participating the health plan of a policyholder will be

$$\frac{z(\lambda + 1)n^H + (1 - z)\lambda n^L}{(1 - z)n^L + zn^H}(n^L + n^H),$$

where $\frac{z(\lambda+1)n^H+(1-z)\lambda n^L}{(1-z)n^L+zn^H}$ is the expected quality externality from interacting with one physician of the health plan. When $z=\frac{1}{2}$, the policyholders have no more information about providers' quality other than that disclosed by the health plan. Consequently they are randomly assigned to providers. Then we have

$$\frac{z(\lambda+1)n^H+(1-z)\lambda n^L}{(1-z)n^L+zn^H}(n^L+n^H)=(\lambda+1)n^H+\lambda n^L.$$

Moreover, as long as $z<1$, our results can not be changed quantitatively.

A.2 Quality and Welfare Comparison under Fixed Salary Scheme (Proof of Proposition 2.3 and 2.4)

Proof. To prove multi-homing enhance quality levels of plans, it is sufficient to prove that $\frac{\widetilde{x}_A^H}{\widetilde{x}_A^L}$ and $\frac{1-\widetilde{x}_B^H}{1-\widetilde{x}_B^L}$ are always higher in multi-homing case than single-homing. With multi-homing, $\frac{\widetilde{x}_A^H}{\widetilde{x}_A^L}=\frac{1-\widetilde{x}_B^H}{1-\widetilde{x}_B^L}=\frac{1+\lambda}{\lambda}$; while with single-homing,

$$\frac{\widetilde{x}_A^H}{\widetilde{x}_A^L}=\begin{cases}\frac{t-\frac{1+\lambda}{3}\widehat{\theta}(1-2F(\widehat{\theta}))}{\lambda\widehat{\theta}F(\widehat{\theta})} & if \quad t\in(\underline{t},\frac{1+\lambda}{3}\widehat{\theta})\neq\varnothing\\ \frac{2t}{\lambda\widehat{\theta}} & if \quad t\in[\frac{1+\lambda}{3}\widehat{\theta},\frac{1+\lambda}{2}\widehat{\theta}]\\ \frac{1+\lambda}{\lambda} & if \quad t\in(\frac{1+\lambda}{2}\widehat{\theta},\infty)\end{cases}\leq\frac{1+\lambda}{\lambda}$$

$$\frac{1-\widetilde{x}_B^H}{1-\widetilde{x}_B^L}=\begin{cases}\frac{t+\frac{1+\lambda}{3}\widehat{\theta}(1-2F(\widehat{\theta}))}{\lambda\widehat{\theta}(1-F(\widehat{\theta}))} & if \quad t\in(\underline{t},\frac{1+\lambda}{3}\widehat{\theta})\neq\varnothing\\ \frac{2t}{\lambda\widehat{\theta}} & if \quad t\in[\frac{1+\lambda}{3}\widehat{\theta},\frac{1+\lambda}{2}\widehat{\theta}]\\ \frac{1+\lambda}{\lambda} & if \quad t\in(\frac{1+\lambda}{2}\widehat{\theta},\infty)\end{cases}\leq\frac{1+\lambda}{\lambda}$$

Social welfare is utilitarian, i.e.,

$$\begin{aligned}SW &= (\lambda n_A+n_Aq_A)\int_0^{\widehat{\theta}}\theta f(\theta)d\theta+(\lambda n_B+n_Bq_B)\int_{\widehat{\theta}}^1\theta f(\theta)d\theta\\ &\quad -cE(\theta)-\frac{t}{2}[(1-q)(\widetilde{x}_A^{L2}+(1-\widetilde{x}_B^L)^2)+q(\widetilde{x}_A^{H2}+(1-\widetilde{x}_B^H)^2)\end{aligned}$$

In order to compare the welfare under two different constraints, we need to do it one by one according to the variations of the transportation cost t. To make it simple, we use the following separate form $SW=SW^L+SW^H-cE(\theta)$, where SW^j net social gain from interacting between policyholders and j type doctors. More precisely,

$$\begin{aligned} SW^L &= \lambda(1-q)\left[\widetilde{x}_A^L \int_0^{\widehat{\theta}} \theta f(\theta)d\theta + (1-\widetilde{x}_B^L)\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta\right] \\ & -\frac{t}{2}(1-q)\left[\widetilde{x}_A^{L2} + (1-\widetilde{x}_B^L)^2\right] \end{aligned}$$

$$\begin{aligned} SW^H &= (1+\lambda)q\left[\widetilde{x}_A^H \int_0^{\widehat{\theta}} \theta f(\theta)d\theta + (1-\widetilde{x}_B^H)\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta\right] \\ & -\frac{t}{2}q\left[\widetilde{x}_A^{H2} + (1-\widetilde{x}_B^H)^2\right]. \end{aligned}$$

Since, under fixed salary scheme and A1, $\widetilde{x}_A^{L,S} = \widetilde{x}_A^{L,M}$ and $1-\widetilde{x}_B^{L,S} = 1-\widetilde{x}_B^{H,M}$ (see proposition 2.1 and 2.2), we have $SW^{L,S} = SW^{L,M}$.

With multi-homing, SW^H is

$$\begin{aligned} SW^{H,M} &= (1+\lambda)q\left[\frac{1+\lambda}{2t}\widehat{\theta}F(\widehat{\theta})\int_0^{\widehat{\theta}} \theta f(\theta)d\theta + (\frac{1+\lambda}{2t}\widehat{\theta}(1-F(\widehat{\theta})))\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta\right] \\ & -\frac{t}{2}q((\frac{1+\lambda}{2t}\widehat{\theta}F(\widehat{\theta}))^2 + (\frac{1+\lambda}{2t}\widehat{\theta}(1-F(\widehat{\theta})))^2). \end{aligned}$$

With single-homing, if $t \in [\frac{1+\lambda}{3}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}]$,

$$\begin{aligned} SW_1^{H,S} &= (1+\lambda)q\left[F(\widehat{\theta})\int_0^{\widehat{\theta}} \theta f(\theta)d\theta + (1-F(\widehat{\theta}))\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta\right] \\ & -\frac{t}{2}q(F(\widehat{\theta})^2 + (1-F(\widehat{\theta}))^2). \end{aligned}$$

If $t \in (\underline{t}, \frac{1+\lambda}{3}\widehat{\theta}) \neq \varnothing$,

$$\begin{gathered} SW_2^{H,S} = (1+\lambda)q\left[\frac{1}{2} - \frac{1+\lambda}{6t}\widehat{\theta}(1-2F(\widehat{\theta}))\right] \\ \int_0^{\widehat{\theta}} \theta f(\theta)d\theta + (1+\lambda)q\left[\frac{1}{2} + \frac{1+\lambda}{6t}\widehat{\theta}(1-2F(\widehat{\theta}))\right]\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta \\ -\frac{t}{2}q((\frac{1}{2} - \frac{1+\lambda}{6t}\widehat{\theta}(1-2F(\widehat{\theta})))^2 + (\frac{1}{2} + \frac{1+\lambda}{6t}\widehat{\theta}(1-2F(\widehat{\theta})))^2). \end{gathered}$$

Thus to prove multi-homing improving social welfare, it is sufficient to show that $SW^{H,M} \geq max\left\{SW_1^{H,S}, SW_2^{H,S}\right\}$. And this is proven by following Lemmas. □

Lemma A.1. *$m_{\widehat{\theta}} = F(\widehat{\theta})\int_0^{\widehat{\theta}} \theta f(\theta)d\theta + (1-F(\widehat{\theta}))\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta$, $n_{\widehat{\theta}} = F(\widehat{\theta})^2 + (1-F(\widehat{\theta}))^2$, we have $\frac{m_{\widehat{\theta}}}{n_{\widehat{\theta}}} > \frac{\widehat{\theta}}{2}$.*

Proof.

$$\begin{aligned} m_{\widehat{\theta}} &> (1-F(\widehat{\theta}))\int_{\widehat{\theta}}^{1}\theta f(\theta)d\theta > \widehat{\theta}(1-F(\widehat{\theta}))^2 \\ &\Rightarrow \frac{m_{\widehat{\theta}}}{n_{\widehat{\theta}}} > \frac{\widehat{\theta}(1-F(\widehat{\theta}))^2}{F(\widehat{\theta})^2+(1-F(\widehat{\theta}))^2} \end{aligned}$$

since since $1-F(\widehat{\theta}) > \frac{1}{2} > F(\widehat{\theta})$, $\frac{\widehat{\theta}(1-F(\widehat{\theta}))^2}{F(\widehat{\theta})^2+(1-F(\widehat{\theta}))^2} > \frac{\widehat{\theta}}{2}$,thus, $\frac{m_{\widehat{\theta}}}{n_{\widehat{\theta}}} > \frac{\widehat{\theta}}{2}$. □

Lemma A.2. *For all $t \in [\frac{1+\lambda}{3}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}]$, $SW^{H,M} \geq SW_1^{H,S}$.*

Proof.

$$SW^{H,M} = \frac{q(1+\lambda)^2}{2t}\widehat{\theta}m_{\widehat{\theta}} - \frac{tq}{2}(\frac{(1+\lambda)\widehat{\theta}}{2t})^2\widetilde{\theta}n_{\widehat{\theta}}$$

$$SW_1^{H,S} = q(1+\lambda)m_{\widehat{\theta}} - \frac{tq}{2}\widehat{\theta}n_{\widehat{\theta}}$$

$$\begin{aligned} SW^{H,M} - SW_1^{H,S} &= q[(\frac{(1+\lambda)\widehat{\theta}}{2t}-1)(1+\lambda)m_{\widehat{\theta}} - \frac{t}{2}(\widehat{\theta}^2(\frac{1+\lambda}{2t})^2-1)n_{\widehat{\theta}}] \\ &= qn_{\widehat{\theta}}[(\frac{(1+\lambda)\widehat{\theta}}{2t}-1)(\frac{(1+\lambda)m_{\widehat{\theta}}}{n_{\widehat{\theta}}} - \frac{t}{2}(\frac{(1+\lambda)\widehat{\theta}}{2t}+1))] \end{aligned}$$

Since $t \in [\frac{1+\lambda}{3}\widehat{\theta}, \frac{1+\lambda}{2}\widehat{\theta}]$, we have $\frac{(1+\lambda)\widehat{\theta}}{2t} \geq 1$. Moreover, from lemma A.1, $(1+\lambda)\frac{m_{\widehat{\theta}}}{n_{\widehat{\theta}}} > (1+\lambda)\frac{\widehat{\theta}}{2} \geq (1+\lambda))\frac{\widehat{\theta}}{4} + \frac{t}{2}$, thus, $SW^{H,M} \geq SW_1^{H,S}$. □

Lemma A.3. *For all $t \in (\underline{t}, \frac{1+\lambda}{3}\widehat{\theta}) \neq \varnothing$, $SW^{H,M} \geq SW_2^{H,S}$.*

Proof.

$$\begin{aligned} SW^{H,M} - SW_2^{H,S} &= (1+\lambda)q[(\frac{1+\lambda}{3t}\widehat{\theta}F(\widehat{\theta}) + \frac{1+\lambda}{6t}\widehat{\theta}(1-F(\widehat{\theta})) - \frac{1}{2})\int_0^{\widehat{\theta}}\theta f(\theta)d\theta \\ &\quad +(\frac{1+\lambda}{6t}\widehat{\theta}F(\widehat{\theta}) + \frac{1+\lambda}{3t}\widehat{\theta}(1-F(\widehat{\theta})) - \frac{1}{2})\int_{\widehat{\theta}}^1\theta f(\theta)d\theta] \\ &\quad -\frac{t}{2}q[((\frac{1+\lambda}{2t}\widehat{\theta}F(\widehat{\theta}))^2 + (\frac{1+\lambda}{2t}\widehat{\theta}(1-F(\widehat{\theta})))^2) \\ &\quad -((\frac{1}{2}-\frac{1+\lambda}{6t}\widehat{\theta}(1-2F(\widehat{\theta})))^2 + (\frac{1}{2}+\frac{1+\lambda}{6t}\widehat{\theta}(1-2F(\widehat{\theta})))^2)] \\ &= q[(\frac{1+\lambda}{6t}\widehat{\theta}-\frac{1}{2})(1+\lambda)E(\theta) + \frac{t}{4} + \frac{(1+\lambda)^2}{6t}\widehat{\theta}m_{\widehat{\theta}} \\ &\quad -\frac{1}{72t}(1+\lambda)^2\widehat{\theta}^2(5n_{\widehat{\theta}}+2). \end{aligned}$$

Note that if $t = \frac{1+\lambda}{3}\widehat{\theta}$, $SW_1^{H,S} = SW_2^{H,S}$ and $SW^{H,M} \geq SW_1^{H,S}$, thus to prove this lemma, it is sufficient to show that $\frac{d(SW^{H,M}-SW_2^{H,S})}{dt} < 0$.

$$\frac{d(SW_1^H - SW_3^H)}{dt} = y(\frac{1}{4} + \frac{1}{72t^2}(1+\lambda)^2\widehat{\theta}^2(5n_{\widehat{\theta}} + 2) - \frac{(1+\lambda)^2}{6t^2}\widehat{\theta}(m_{\widehat{\theta}} + E(\theta)))$$

.

Similar to Lemma A.2, we get $\frac{d(SW^{H,M} - SW_2^{H,S})}{dt} < 0$. □

A.3 An Example of Distribution for Proposition 2.5

Consider a continuously differentiable distribution

$$F(\theta) = \frac{3}{8}\theta, \text{ if } 0 < \theta < \frac{8}{9} \text{ and} F(\theta) = a + (\theta - b)^c, \text{ if } \frac{8}{9} \leq \theta \leq 1,$$

where a, b, c satisfy that (1) $a + (\frac{8}{9} - b)^c = \frac{1}{3}$, (2) $a + (1-b)^c = 1$ and (3) $c(\frac{8}{9} - b)^{c-1} = \frac{3}{8}$. Because the system of (1), (2) and (3) is not over-identified, such a distribution function exists. We will prove that

$$F(\widehat{\theta})^2[\widehat{\theta} - E(\theta|\theta < \widehat{\theta})] > [1 - F(\widehat{\theta})]^2[E(\theta|\theta \geq \widehat{\theta}) - \widehat{\theta}],$$

$$\text{or } \frac{E(\theta|\theta \geq \widehat{\theta}) - \widehat{\theta}}{\widehat{\theta} - E(\theta|\theta \leq \widehat{\theta})} < [\frac{F(\widehat{\theta})}{1 - F(\widehat{\theta})}]^2$$

With this distribution, we have $1 - 2F(\frac{8}{9}) = f(\frac{8}{9})\frac{8}{9}$. Indeed, $\frac{8}{9}$ is the equilibrium risk segmentation $\widehat{\theta}$, the only value in $(0, 1]$ such that $1 - 2F(\theta) = f(\theta)\theta$.

First, for all $\theta \in (0, \frac{8}{9})$, $1 - 2F(\theta) - f(\theta)\theta = 1 - \frac{9}{8}\theta > 0$.

Second, we need to prove that for all $\theta \in (\frac{8}{9}, 1]$, $1 - 2F(\theta) < f(\theta)\theta$. For all $\theta \in (\frac{8}{9}, 1]$, $F(\theta) = a + (\theta - b)^c$. Since $(1-b)^c - (\frac{8}{9} - b)^c = \frac{2}{3} > \frac{1}{9}$, we must have $c > 1$. Subsequently, because $c(\frac{8}{9} - b)^{c-1} = \frac{3}{8} > 0$, we have $b < \frac{8}{9}$. Thus for all $\theta \in (\frac{8}{9}, 1]$, $d(1 - 2F(\theta))/d\theta < 0$ and $d(f(\theta)\theta)/d\theta = c(\theta - b)^{c-2}(c - b) > 0$.

Thus, $\widehat{\theta} = \frac{8}{9}$.

Now let us prove that $\frac{E(\theta|\theta \geq \widehat{\theta}) - \widehat{\theta}}{\widehat{\theta} - E(\theta|\theta \leq \widehat{\theta})} < [\frac{F(\widehat{\theta})}{1 - F(\widehat{\theta})}]^2$. Because $\widehat{\theta} = \frac{8}{9}$, we have $F(\widehat{\theta}) = \frac{1}{3}$, $[\frac{F(\widehat{\theta})}{1 - F(\widehat{\theta})}]^2 = \frac{1}{4}$, $\widehat{\theta} - E(\theta|\theta \leq \widehat{\theta}) = \frac{8}{9} - \frac{4}{9} = \frac{4}{9}$. Furthermore, since $E(\theta|\theta \geq \widehat{\theta}) < 1$, we have $\frac{E(\theta|\theta \geq \widehat{\theta}) - \widehat{\theta}}{\widehat{\theta} - E(\theta|\theta \leq \widehat{\theta})} < \frac{1 - \frac{8}{9}}{\frac{4}{9}} = \frac{1}{4} = [\frac{F(\widehat{\theta})}{1 - F(\widehat{\theta})}]^2$.

A.4 Equilibrium Under FFS (Proof of Proposition 2.7)

Let $\widehat{R}_A^j = \frac{R_A^j - c}{(1-y)\widetilde{x}_A^L + y\widetilde{x}_A^H}$ and $\widehat{R}_B^j = \frac{R_B^j - c}{(1-y)(1-\widetilde{x_B^L}) + y(1-\widetilde{x_B^H})}$ denote a type j provider's expected payment from each activity (treatment) in plan A and B respectively. Then we have for all $j = H, L$,

$$T_A^j = \widehat{R}_A^j \int_0^{\tilde{\theta}} \theta f(\theta) d\theta \text{ and } T_B^j = \widehat{R}_B^j \int_{\tilde{\theta}}^1 \theta f(\theta) d\theta.$$

Apparently, for every $(R_i^j, \widetilde{x}_i^j)$ there is only one corresponding $(\widehat{R}_i^j, \widetilde{x}_i^i)$. Thus for any plan, maximizing its profit with respect to $(R_i^j, \widetilde{x}_i^i)$ is equivalent to doing that with respect to $(\widehat{R}_i^j, \widetilde{x}_i^j)$. The problem of plan A is

$$\begin{aligned}
\underbrace{\max}_{P_A, \widehat{R}_A^L, \widetilde{x}_A^L, \widehat{R}_A^H, \widetilde{x}_A^H} \quad \pi_A &= P_A F(\widetilde{\theta}) - (1-q) t (\widetilde{x}_A^L)^2 \\
&\quad - \left[c + q \widehat{R}_A^H \widetilde{x}_A^H \right] \int_0^{\tilde{\theta}} \theta f(\theta) d\theta \\
s.t. \ (IR_A^H) \quad &: \quad \widehat{R}_A^H \int_0^{\tilde{\theta}} \theta f(\theta) d\theta - t \widetilde{x}_A^H \geq 0 \\
(IC_A^H) \quad &: \quad \widehat{R}_A^H \int_0^{\tilde{\theta}} \theta f(\theta) d\theta - t \widetilde{x}_A^H \geq \widehat{R}_B^H \int_{\tilde{\theta}}^1 \theta f(\theta) d\theta - t(1 - \widetilde{x}_A^H),
\end{aligned}$$

and that of plan B is

$$\begin{aligned}
\underbrace{\max}_{P_B, \widehat{R}_B^L, \widetilde{x}_B^L, \widehat{R}_B^H, \widetilde{x}_B^H} \quad \pi_B &= P_B(1 - F(\tilde{\theta})) - (1-q)(1 - \widetilde{x}_B^L)^2 \\
&\quad - \left[c + q \widehat{R}_B^H (1 - \widetilde{x}_B^H) \right] \int_{\tilde{\theta}}^1 \theta f(\theta) d\theta \\
s.t. \ (IR_B^H) \quad &: \quad \widehat{R}_B^H \int_{\tilde{\theta}}^1 \theta f(\theta) d\theta - t(1 - \widetilde{x}_B^H) \geq 0 \\
(IC_B^H) \quad &: \quad \widehat{R}_B^H \int_{\tilde{\theta}}^1 \theta f(\theta) d\theta - t(1 - \widetilde{x}_B^H) \geq \widehat{R}_A^H \int_0^{\tilde{\theta}} \theta f(\theta) d\theta - t \widetilde{x}_B^H.
\end{aligned}$$

he first order conditions[1] for P_A and P_B are

$$\begin{aligned}
\frac{\partial \pi_A}{\partial P_A} &= F(\tilde{\theta}) - \frac{P_A f(\tilde{\theta}) \tilde{\theta}}{\Delta P} + c \frac{\tilde{\theta}^2 f(\tilde{\theta})}{\Delta P} - \xi_2 \widehat{R}_B^H \frac{\tilde{\theta}^2 f(\tilde{\theta})}{\Delta P} = 0, \\
\frac{\partial \pi_B}{\partial P_B} &= 1 - F(\tilde{\theta}) - \frac{P_B f(\tilde{\theta}) \tilde{\theta}}{\Delta P} + c \frac{\tilde{\theta}^2 f(\tilde{\theta})}{\Delta P} - \xi_4 \widehat{R}_A^H \frac{\tilde{\theta}^2 f(\tilde{\theta})}{\Delta P} = 0,
\end{aligned}$$

[1] Assume that these are two concave problems.

where $\xi_2 > 0$ and $\xi_4 > 0$ are Lagrangian parameters associated with IC_A^H and IC_B^H respectively (ξ_1 and ξ_3 are those with IR_A^H and IR_B^H respectively). Rearrange the first-order conditions for P_A and P_B, then equilibrium risk segmentation $\widetilde{\theta}$ is given by

$$1 - 2F(\tilde{\theta}) = f(\tilde{\theta})\tilde{\theta}(1 + \frac{\tilde{\theta}}{\Delta P}(\xi_2 \widehat{R}_B^H - \xi_4 \widehat{R}_A^H)) \tag{A.1}$$

The first order conditions for $\widetilde{x}_A^L$ and $(1 - \widetilde{x}_B^L)$ are

$$\begin{aligned} \frac{\partial \pi_A}{\partial \widetilde{x}_A^L} &= (P_A - (c - \xi_2 \widehat{R}_B^H)\tilde{\theta})\frac{\widetilde{\theta}^2 f(\tilde{\theta})\lambda(1-q)}{\Delta P} - 2t(1-q)\widetilde{x}_A^L = 0, &\text{(A.2)}\\ \frac{\partial \pi_B}{\partial (1 - \widetilde{x}_B^L)} &= (P_B - (c - \xi_4 \widehat{R}_A^H)\tilde{\theta})\frac{\widetilde{\theta}^2 f(\tilde{\theta})\lambda(1-q)}{\Delta P} - 2t(1-q)(1 - \widetilde{x}_B^L) = 0. &\text{(A.3)} \end{aligned}$$

Combining these first order conditions and replacing $\widetilde{\theta}$ with $\widetilde{\theta}(t)$, we have

$$\widetilde{x}_A^L = \frac{\lambda}{2t}\widetilde{\theta}(t)F(\widetilde{\theta}(t)) \text{ and } 1 - \widetilde{x}_B^L = \frac{\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t))).$$

The first order conditions for $\widetilde{x}_A^H$ and $1 - \widetilde{x}_B^H$ are

$$\begin{aligned} \frac{\partial \pi_A}{\partial \widetilde{x}_A^L} &= (P_A - (c - \xi_2 \widehat{R}_B^H)\tilde{\theta})\frac{\widetilde{\theta}^2 f(\tilde{\theta})(1+\lambda)q}{\Delta P} - 2tq\widetilde{x}_A^L = 0,\\ \frac{\partial \pi_B}{\partial (1 - \widetilde{x}_B^L)} &= (P_B - (c - \xi_4 \widehat{R}_A^H)\tilde{\theta})\frac{\widetilde{\theta}^2 f(\tilde{\theta})(1+\lambda)q}{\Delta P} - 2tq(1 - \widetilde{x}_B^L) = 0. \end{aligned}$$

Now let us prove that $\widetilde{x}_A^H = 1 - \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$ and $1 - \widetilde{x}_B^H = \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$ should consist of an equilibrium H provider configuration.

Lemma A.4. *When $t \in (\underline{t}\ , \frac{1+\lambda}{2}\widehat{\theta})$, single-homing equilibrium risk segmentation $\widetilde{\theta}(t)$ satisfies $t < \frac{1+\lambda}{2}\widetilde{\theta}(t)$ and IC_i^H must be binding.*

Proof. If $t \geq \frac{1+\lambda}{2}\widetilde{\theta}(t)$, plans will optimally choose $\widetilde{x}_A^H = \frac{1+\lambda}{2t}\widetilde{\theta}(t)F(\widetilde{\theta}(t))$ and $1 - \widetilde{x}_B^H = \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$, because $1 - \widetilde{x}_B^H + \widetilde{x}_A^H = \frac{1+\lambda}{2t}\widetilde{\theta}(t) \leq 1$ and plans' IC_i^H are automatically satisfied. However, in this case,

$$\begin{aligned} \pi_A &= P_A F(\widetilde{\theta}) - (1-q)t(\widetilde{x}_A^L)^2 - qt(\widetilde{x}_A^H)^2 - c\int_0^{\tilde{\theta}} \theta f(\theta)d\theta \text{ and}\\ \pi_B &= P_B(1 - F(\widetilde{\theta})) - (1-q)t(1 - \widetilde{x}_B^L)^2 - qt(1 - \widetilde{x}_B^H)^2 - c\int_{\tilde{\theta}}^1 \theta f(\theta)d\theta, \end{aligned}$$

which means that $\widetilde{\theta}(t) = \widehat{\theta}$ as proved in Proposition 2.6. However, $t < \frac{1+\lambda}{2}\widehat{\theta}$ generates the contradiction. So $t < \frac{1+\lambda}{2}\widetilde{\theta}(t)$ and IC_i^H must be binding. $\square$

Lemma A.5. *With single-homing, when $t \in (\underline{t}\ , \frac{1+\lambda}{2}\widehat{\theta})$ and IC_i^H and IR_i^H are all binding, $\widetilde{\theta}(t)$ is the equilibrium risk segmentation, then $\widetilde{x}_A^H = 1 - \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$ and $1 - \widetilde{x}_B^H = \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$ consist of an equilibrium H provider configuration.*

Proof. The equilibrium risk segmentation $\widetilde{\theta} = (P_B - P_A)/((1-q)\lambda(1 - \widetilde{x}_B^L - \widetilde{x}_A^L) + q(1 + \lambda)(1 - \widetilde{x}_B^H - \widetilde{x}_A^H))$, then plan A's profit is

$$\begin{aligned}\pi_A &= (P_B - \widetilde{\theta}((1-q)\lambda(1 - \widetilde{x}_B^L - \widetilde{x}_A^L) + q(1+\lambda)(1 - \widetilde{x}_B^H - \widetilde{x}_A^H)))F(\widetilde{\theta}) \\ &\quad -(1-q)t(\widetilde{x}_A^L)^2 - qt(\widetilde{x}_A^H)^2 - c\int_0^{\tilde{\theta}} \theta f(\theta)d\theta\end{aligned}$$

The first order condition for $\widetilde{x}_A^H$ is

$$\frac{\partial \pi_A}{\partial \widetilde{x}_A^H} = q(1+\lambda)\widetilde{\theta}F(\widetilde{\theta}) - 2qt\widetilde{x}_A^H,$$

which is larger than 0, if IC_A^H is binding. Otherwise, it equals to 0. Thus we have $\frac{\partial \pi_A}{\partial \widetilde{x}_A^H} > 0$ for all $\widetilde{x}_A^H \le \frac{1+\lambda}{2t}\widetilde{\theta}(t)F(\widetilde{\theta}(t))$. Given $1 - \widetilde{x}_B^H = \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$, $\widetilde{x}_A^H \le 1 - \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$. From Lemma A.1, we know that $t < \frac{1+\lambda}{2}\widetilde{\theta}(t)$. We must have

$$\widetilde{x}_A^H \le 1 - \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t))) < F(\widetilde{\theta}(t)) < \frac{1+\lambda}{2t}\widetilde{\theta}(t)F(\widetilde{\theta}(t))$$

As the result, the best response of plan A given plan B choosing $1 - \widetilde{x}_B^H = \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$ is $1 - \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$. Using the same logic to plan B, we can conclude that the best response of plan B given plan A choosing $\widetilde{x}_A^H = 1 - \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$ is $1 - \widetilde{x}_B^H = \frac{1+\lambda}{2t}\widetilde{\theta}(t)(1 - F(\widetilde{\theta}(t)))$. □

Finally, we solve the system, when IC_i^H and IR_i^H $(i = A, B)$ are all binding. The first order condition for $\widetilde{x}_A^H$ and $1 - \widetilde{x}_B^H$ are

$$\begin{aligned}0 = \frac{\partial \pi_A}{\partial \widetilde{x}_A^H} &= (P_A - (c - \xi_2 \widehat{R}_B^H)\tilde{\theta})\frac{\tilde{\theta}^2 f(\tilde{\theta})(1+\lambda)q}{\Delta P} - \\ &\quad q\widehat{R}_A^H \int_0^{\tilde{\theta}} \theta f(\theta)d\theta - 2t\xi_2 - t\xi_1 \qquad \text{(A.4)} \\ 0 = \frac{\partial \pi_B}{\partial (1 - \widetilde{x}_B^H)} &= (P_B - (c - \xi_4 \widehat{R}_A^H)\tilde{\theta})\frac{\tilde{\theta}^2 f(\tilde{\theta})(1+\lambda)q}{\Delta P} - \\ &\quad q\widehat{R}_A^H \int_{\tilde{\theta}}^1 \theta f(\theta)d\theta - 2t\xi_4 - t\xi_3 \qquad \text{(A.5)}\end{aligned}$$

and those for $\widehat{R}_A^H$ and $\widehat{R}_B^H$ are

$$\begin{aligned}\frac{\partial \pi_A}{\partial \widehat{R}_A^H} &= -q\widetilde{x}_A^H \int_0^{\tilde{\theta}} \theta f(\theta)d\theta + (\xi_2+\xi_1)\int_0^{\tilde{\theta}} \theta f(\theta)d\theta = 0 \quad (A.6)\\ \frac{\partial \pi_B}{\partial \widehat{R}_B^H} &= -q(1-\widetilde{x}_B^H)\int_{\tilde{\theta}}^1 \theta f(\theta)d\theta + (\xi_4+\xi_3)\int_{\tilde{\theta}}^1 \theta f(\theta)d\theta = 0. \quad (A.7)\end{aligned}$$

Combing (A.2), (A.3), (A.4), (A.5) together with $\widehat{R}_A^H \int_0^{\tilde{\theta}} \theta f(\theta)d\theta = t\widetilde{x}_A^H = t(1-\frac{1+\lambda}{2t}\widetilde{\theta}(1-F(\widetilde{\theta})))$ and $\widehat{R}_B^H \int_{\tilde{\theta}}^1 \theta f(\theta)d\theta = t(1-\widetilde{x}_B^j) = \frac{1+\lambda}{2}\widetilde{\theta}(1-F(\widetilde{\theta}))$ (since IR_i^H are binding), the following two equations hold:

$$\begin{aligned}(1+\lambda)qF(\tilde{\theta})\tilde{\theta} &= qt(1-\frac{1+\lambda}{2t}\widetilde{\theta}(t)(1-F(\widetilde{\theta}))) + 2t\xi_2 + t\xi_1 \quad (A.8)\\ (1+\lambda)q(1-F(\tilde{\theta}))\tilde{\theta} &= q\frac{1+\lambda}{2}\widetilde{\theta}(1-F(\tilde{\theta})) + 2t\xi_4 + t\xi_3. \quad (A.9)\end{aligned}$$

From (A.6) and (A.7), we have

$$\begin{aligned}\xi_2+\xi_1 &= q\widetilde{x}_A^H = q(1-\frac{1+\lambda}{2t}\widetilde{\theta}(1-F(\widetilde{\theta})))\\ \xi_4+\xi_3 &= q(1-\widetilde{x}_B^H) = q\frac{1+\lambda}{2t}\widetilde{\theta}(1-F(\widetilde{\theta})),\end{aligned}$$

which (together with (A.8) and (A.9)) implies that $\xi_2 = 2q(\frac{1+\lambda}{2t}\tilde{\theta}-1) > 0$ and $\xi_4 = 0$. Formula (17) can be rewritten as

$$1-2F(\tilde{\theta}) = f(\tilde{\theta})\tilde{\theta}(1+\xi_2\widehat{R}_A^H\frac{\widetilde{\theta}}{\Delta P}) \quad (A.10)$$

Lemma A.6. *Risk segmentation $\tilde{\theta}(t)$ defined by formula (A.10) satisfies that $\tilde{\theta}(t) < \widehat{\theta} = \frac{1-2F(\widehat{\theta})}{f(\widehat{\theta})}$.*

Proof. We have $1-2F(\tilde{\theta}) > f(\tilde{\theta})\tilde{\theta}$ from (A.10) since $\xi_2 = 2q(\frac{1+\lambda}{2t}\tilde{\theta}-1) > 0$. Remind that $\frac{1-F(\theta)}{f(\theta)}$ decreases with θ. It must be $\tilde{\theta}(t) < \widehat{\theta} = \frac{1-2F(\widehat{\theta})}{f(\widehat{\theta})}$. □

Note that the equilibrium L provider configuration is $\widetilde{x}_A^L = \frac{\lambda}{2t}\widetilde{\theta}(t)F(\widetilde{\theta}(t))$ and $1-\widetilde{x}_B^L = \frac{\lambda}{2t}\widetilde{\theta}(t)(1-F(\widetilde{\theta}(t)))$, thus with $\tilde{\theta}(t) < \widehat{\theta}$ and A1, $1-\widetilde{x}_B^L + \widetilde{x}_A^L = \frac{\lambda}{2t}\widetilde{\theta}(t) < \frac{\lambda}{2t}\widehat{\theta} < 1$, or in words, L providers are not fully covered by health plans in equilibrium.

A.5 Efficiency of Single-homing under Fee-for-Service (Proof of Proposition 2.8)

Here we continue to separate the utilitarian social welfare function:

$$SW^k(t,\widetilde{\theta}) = SW^{H,k}(t,\widetilde{\theta}) + SW^{L,k}(t,\widetilde{\theta}) - cE(\theta)$$

where $\widetilde{\theta}$ is the equilibrium risk segmentation and $k = M, S$ indicates multi-homing and single-homing respectively. Because in the multi-homing outcome $\widetilde{\theta} = \widehat{\theta}$ for all t, taking the value of provider configuration from proposition 2.6, we have

$$\begin{aligned}
SW^{H,M}(t,\widetilde{\theta}) &= q\{\frac{(1+\lambda)^2}{2t}[\widehat{\theta}F(\widehat{\theta})\int_0^{\widehat{\theta}}\theta f(\theta)d\theta + \widehat{\theta}(1-F(\widehat{\theta}))\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta] \\
&\quad -\frac{(1+\lambda)^2}{8t}\widehat{\theta}^2[F(\widehat{\theta})^2 + (1-F(\widehat{\theta}))^2]\} \\
SW^{L,M}(t,\widetilde{\theta}) &= (1-q)\{\frac{\lambda^2}{2t}[\widehat{\theta}F(\widehat{\theta})\int_0^{\widehat{\theta}}\theta f(\theta)d\theta + \widehat{\theta}(1-F(\widehat{\theta}))\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta] \\
&\quad -\frac{\lambda^2}{8t}\widehat{\theta}^2[F(\widehat{\theta})^2 + (1-F(\widehat{\theta}))^2]\}.
\end{aligned}$$

In the single-homing outcome given by proposition 2.7,

$$\begin{aligned}
SW^{H,S}(t,\widetilde{\theta}) &= q\{(1+\lambda)[(1-\frac{1+\lambda}{2t}\widetilde{\theta}(1-F(\widetilde{\theta})))\int_0^{\widetilde{\theta}}\theta f(\theta)d\theta \\
&\quad +\frac{1+\lambda}{2t}\widetilde{\theta}(1-F(\widetilde{\theta}))\int_{\widetilde{\theta}}^1 \theta f(\theta)d\theta] \\
&\quad -\frac{t}{2}(1-\frac{1+\lambda}{2t}\widetilde{\theta}(1-F(\widetilde{\theta})))^2 - \frac{(1+\lambda)^2}{8t}\widetilde{\theta}^2(1-F(\widetilde{\theta}))^2\} \\
SW^{L,S}(t,\widetilde{\theta}) &= (1-q)\{\frac{\lambda^2}{2t}[\widetilde{\theta}F(\widetilde{\theta})\int_0^{\widetilde{\theta}}\theta f(\theta)d\theta + \widetilde{\theta}(1-F(\widetilde{\theta}))\int_{\widetilde{\theta}}^1 \theta f(\theta)d\theta] \\
&\quad -\frac{\lambda^2}{8t}\widetilde{\theta}^2[F(\widetilde{\theta})^2 + (1-F(\widetilde{\theta}))^2]\}.
\end{aligned}$$

Note that when $t = \frac{1+\lambda}{2}\widehat{\theta}$, $SW^S = SW^M$, because the multi-homing outcome is equivalent to the single-homing one when there is measure 0 provider multi-homing (only the H provider locates at $\frac{1+\lambda^2}{2t}\widehat{\theta}F(\widehat{\theta})$ multihomes). To prove for some t less than $\frac{1+\lambda}{2}\widehat{\theta}$, multi-homing leads to less social welfare than single-homing does, it is sufficient to prove that

$$-\frac{dSW^S(t,\widetilde{\theta})}{dt}|_{t=\frac{1+\lambda}{2}\widehat{\theta}} > -\frac{dSW^M(t,\widetilde{\theta})}{dt}|_{t=\frac{1+\lambda}{2}\widehat{\theta}}.$$

Furthermore, we have

$$\begin{aligned}
\frac{dSW^S(t,\widetilde{\theta})}{dt} &= \frac{\partial SW^{H,S}(t,\widetilde{\theta})}{\partial t} + \frac{\partial SW^{H,S}(t,\widetilde{\theta})}{\partial \widetilde{\theta}}\frac{\partial \widetilde{\theta}}{\partial t} + \frac{\partial SW^{L,S}(t,\widetilde{\theta})}{\partial t} + \frac{\partial SW^{L,S}(t,\widetilde{\theta})}{\partial \widetilde{\theta}}\frac{\partial \widetilde{\theta}}{\partial t} \\
\frac{dSW^M(t,\widetilde{\theta})}{dt} &= \frac{\partial SW^{H,M}(t,\widetilde{\theta})}{\partial t} + \frac{\partial SW^{L,M}(t,\widetilde{\theta})}{\partial t}.
\end{aligned}$$

It is obvious that $\frac{\partial SW^{L,M}(t,\widetilde{\theta})}{\partial t} = \frac{\partial SW^{L,S}(t,\widetilde{\theta})}{\partial t}$. From proposition 2.7 we have $\frac{\partial \widetilde{\theta}}{\partial t}|_{t=\frac{1+\lambda}{2}\widehat{\theta}} > 0$. To prove the proposition, it is sufficient to show the following all three conditions should hold

$$\begin{aligned}
a. \quad -\frac{\partial SW^{H,S}(t,\widetilde{\theta})}{\partial t}|_{t=\frac{1+\lambda}{2}\widehat{\theta}} &\geq -\frac{\partial SW^{H,M}(t,\widetilde{\theta})}{\partial t}|_{t=\frac{1+\lambda}{2}\widehat{\theta}} \\
b. \quad -\frac{\partial SW^{H,S}(t,\widetilde{\theta})}{\partial \widetilde{\theta}}|_{t=\frac{1+\lambda}{2}\widehat{\theta}} &> 0 \\
c. \quad -\frac{\partial SW^{L,S}(t,\widetilde{\theta})}{\partial \widetilde{\theta}}|_{t=\frac{1+\lambda}{2}\widehat{\theta}} &\geq 0.
\end{aligned}$$

First, remind that $\widehat{\theta} = \frac{1-2F(\widehat{\theta})}{f(\widehat{\theta})}$, we have

$$\begin{aligned}
-\frac{\partial SW^{H,M}(t,\widetilde{\theta})}{\partial t}|_{t=\frac{1+\lambda}{2}\widehat{\theta}} &= q\{\frac{2}{\widehat{\theta}}[F(\widehat{\theta})\int_0^{\widehat{\theta}} \theta f(\theta)d\theta + (1-F(\widehat{\theta}))\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta] \\
&\quad -\frac{1}{2}[F(\widehat{\theta})^2 + (1-F(\widehat{\theta}))^2] \\
-\frac{\partial SW^{H,S}(t,\widetilde{\theta})}{\partial t}|_{t=\frac{1+\lambda}{2}\widehat{\theta}} &= q\{\frac{2}{\widehat{\theta}}(1-F(\widehat{\theta}))[\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta - \int_0^{\widehat{\theta}} \theta f(\theta)d\theta] \\
&\quad +\frac{1}{2} - (1-F(\widehat{\theta}))^2.
\end{aligned}$$

Thus condition a holds if and only if

$$E(\theta|\theta \leq \widehat{\theta}) \leq \frac{\widehat{\theta}}{2}.$$

Second, it can be shown that

$$\begin{aligned}
-\frac{\partial SW^{H,S}(t,\widetilde{\theta})}{\partial \widetilde{\theta}}|_{t=\frac{1+\lambda}{2}\widehat{\theta}} &= q\{-\frac{1+\lambda}{\widehat{\theta}}F(\widehat{\theta})[\int_{\widehat{\theta}}^1 \theta f(\theta)d\theta - \int_0^{\widehat{\theta}} \theta f(\theta)d\theta] \\
&\quad +(1+\lambda)(2-\frac{3}{2}F(\widehat{\theta}))-1](1-2F(\widehat{\theta})).
\end{aligned}$$

As a result, condition b holds if and only if

$$\widehat{\theta} > \frac{(1+\lambda)F(\widehat{\theta})(\int_{\widehat{\theta}}^1 \theta dF(\theta) - \int_0^{\widehat{\theta}} \theta dF(\theta))}{[(1+\lambda)(2-\frac{3}{2}F(\widehat{\theta}))-1](1-2F(\widehat{\theta}))}.$$

Third, we have

$$
\begin{aligned}
-\frac{\partial SW^{L,S}(t,\widetilde{\theta})}{\partial\widetilde{\theta}}|_{t=\frac{1+\lambda}{2}\widehat{\theta}} &= (1-q)\{\frac{\lambda^2}{2t}[\widehat{\theta}(1-2F(\widehat{\theta}))^2-[F(\widehat{\theta})\int_0^{\widehat{\theta}}\theta f(\theta)d\theta \\
+(1-F(\widehat{\theta}))\int_{\widehat{\theta}}^1\theta f(\theta)d\theta] &\quad +\frac{\lambda^2}{4t}\widehat{\theta}F(\widehat{\theta})(1-F(\widehat{\theta}))\}.
\end{aligned}
$$

So condition c holds if and only if

$$\widehat{\theta}\geq\frac{(1-F(\widehat{\theta}))F(\widehat{\theta})}{(1-2F(\widehat{\theta}))^2}E(\theta|\theta\geq\widehat{\theta}).$$

Now let us prove that

$$\frac{(1-F(\widehat{\theta}))F(\widehat{\theta})}{(1-2F(\widehat{\theta}))^2}E(\theta|\theta\geq\widehat{\theta})>\frac{(1+\lambda)F(\widehat{\theta})(\int_{\widehat{\theta}}^1\theta dF(\theta)-\int_0^{\widehat{\theta}}\theta dF(\theta))}{[(1+\lambda)(2-\frac{3}{2}F(\widehat{\theta}))-1](1-2F(\widehat{\theta}))}.$$

Thus, if c holds, then b holds. Since $E(\theta|\theta\geq\widehat{\theta})=\frac{\int_{\widehat{\theta}}^1\theta dF(\theta)}{1-F(\widehat{\theta})}$, to prove the above inequality holds, it is equivalent to prove that for all $\lambda\geq 0$, we have

$$(1+\lambda)(1-2F(\widehat{\theta}))<(1+\lambda)(2-\frac{3}{2}F(\widehat{\theta}))-1.$$

This inequality holds apparently.

Thus, when $F(\theta)$ satisfies $\widehat{\theta}\geq\max\{2E(\theta|\theta\leq\widehat{\theta}),\frac{(1-F(\widehat{\theta}))F(\widehat{\theta})}{(1-2F(\widehat{\theta}))^2}E(\theta|\theta\geq\widehat{\theta})\}$, under A1 and FFS scheme, single-homing may lead to more social wefare than multi-homing for some t less than $\frac{1+\lambda}{2}\widehat{\theta}$.

Furthermore, we give an example of such $F(\theta)$:

$$F(\theta)=\frac{9}{32}\theta,\text{ if }0<\theta<\frac{8}{9}\text{ and }F(\theta)=a+(\theta-b)^c,\ \text{if }\frac{8}{9}\leq\theta\leq 1,$$

where a,b,c satisfy that (1) $a+(\frac{8}{9}-b)^c=\frac{1}{4}$, (2) $a+(1-b)^c=1$, (3) $c(\frac{8}{9}-b)^{c-1}=\frac{9}{32}$. Thus $\widehat{\theta}=\frac{8}{9}$ and $F(\widehat{\theta})=\frac{1}{4}$. Moreover, we have $E(\theta|\theta\leq\widehat{\theta})=\frac{4}{9}$ and $E(\theta|\theta\geq\widehat{\theta})<1$. So $\widehat{\theta}=2E(\theta|\theta\leq\widehat{\theta})$ and since

$$\widehat{\theta}\frac{(1-2F(\widehat{\theta}))^2}{(1-F(\widehat{\theta}))F(\widehat{\theta})}=\frac{32}{27}>1>E(\theta|\theta\geq\widehat{\theta}),$$

we have $\widehat{\theta}>\frac{(1-F(\widehat{\theta}))F(\widehat{\theta})}{(1-2F(\widehat{\theta}))^2}E(\theta|\theta\geq\widehat{\theta})$. As a result, $\widehat{\theta}\geq\max\{2E(\theta|\theta\leq\widehat{\theta}),\frac{(1-F(\widehat{\theta}))F(\widehat{\theta})}{(1-2F(\widehat{\theta}))^2}E(\theta|\theta\geq\widehat{\theta})\}$.

Bibliography

[1] Albano, G. L., & Lizzeri, A. (2001). Strategic certification and provision of quality. *International Economic Review, 42*(1), 267-283.

[2] Alger, I., & Salanie F. (2006). A Theory of Fraud and Over-Treatment.

[3] Ambrus, A., & Argenziano, R. (2009). Asymmetric networks in two-sided markets. *American Economic Journal: Microeconomics, 1*(1), 17-52.

[4] Anania, G., & Nistico R. (2004). Public regulation as a substitute for trust in quality food markets: what if the trust substitute cannot be fully trusted?. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift fur die gesamte Staatswissenschaft*, 681-701.

[5] Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics, 37*(3), 668-691.

[6] Armstrong, M., & Wright, J. (2007). Two-sided markets, competitive bottlenecks and exclusive contracts. *Economic Theory, 32*(2), 353-380.

[7] Arora, A., & Ceccagnoli, M. (2006). Patent protection, complementary assets, and firms' incentives for technology licensing. *Management Science,52*(2), 293-308.

[8] Arora, A., Fosfuri, A., & Gambardella, A. (2004). Markets for technology: The economics of innovation and corporate strategy. *The MIT Press.*

[9] Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American economic review, 53*(5), 941-973.

[10] Auriol, E., & Laffont, J. J. (1992). Regulation by duopoly. *Journal of Economics & Management Strategy, 1*(3), 507-533.

[11] Auriol, E., & Renault, R. (2008). Status and incentives. *The RAND Journal of Economics, 39*(1), 305-326.

[12] Auriol, E., & Schilizzi, S. G. (2003). Quality signaling through certification. theory and an application to agricultural seed market. *IDEI* Working Paper,165.

[13] Baker, G. A. (1999). Consumer preferences for food safety attributes in fresh apples: Market segments, consumer characteristics, and marketing opportunities. *Journal of Agricultural and Resource Economics*, 80-97.

[14] Bardey, D., & Bourgeon, J. M. (2010). Health Care Network Formation and Policyholders' Welfare (No. 007457).

[15] Bardey, D., Cremer, H., & Lozachmeur, J. M. (2012). Doctors Remuneration Schemes and Hospital Competition in a Two-Sided Market. *The BE Journal of Economic Analysis & Policy, 12*(1).

[16] Bardey, D., & Rochet, J. C. (2010). Competition Among Health Plans: A TwoSided Market Approach. *Journal of Economics & Management Strategy, 19*(2), 435-451.

[17] Besley, T. J. (1988). Optimal reimbursement health insurance and the theory of Ramsey taxation. *Journal of Health Economics, 7*(4), 321-336.

[18] Bessen, J., & Maskin, E. (2009). Sequential innovation, patents, and imitation. *The RAND Journal of Economics, 40*(4), 611-635.

[19] Biglaiser, G., & Friedman, J. W. (1994). Middlemen as guarantors of quality. *International journal of industrial organization, 12*(4), 509-531.

[20] Brekke, K. R., Nuscheler, R., & Rune Straume, O. (2006). Quality and location choices under price regulation. *Journal of Economics & Management Strategy, 15*(1), 207-227.

[21] Caillaud, B., & Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers. *RAND journal of Economics,* 309-328.

[22] Caswell, J. A. (1998). Valuing the benefits and costs of improved food safety and nutrition. *Australian Journal of Agricultural and Resource Economics,42*(4), 409-424.

[23] Caswell, J.A. and Mojduszka, E.M. (1996). Using informational labeling to influence the market for quality in food products. *American Journal of Agricultural Economics, 78*(4), 1248-1253.

[24] Chang, H. F. (1995). Patent scope, antitrust policy, and cumulative innovation. *The RAND Journal of Economics,* 34-57.

[25] Cremer J., & McLean, R. P. (1985). Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands Are Interdependent. *Econometrica, 53*(2), 345–361.

[26] Cremer, J., & McLean, R. P. (1988). Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica: Journal of the Econometric Society,* 1247-1257.

[27] Crespi, J. M. (2001). How should food safety certification be financed?. *American Journal of Agricultural Economics, 83*(4), 852-861.

[28] Crutchfield, S.R., Buzby, J. C., Roberts,T., Ollinger, M. and Jordan Lin, C. T. (1997) An economic assessment of food safety regulations: the new approach to meat and poultry inspection. *Agricultural Economic Report No.* (AER-755)

[29] Damiano, E., & Li, H. (2007). Price discrimination and efficient matching. *Economic Theory, 30*(2), 243-263.

[30] Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *Journal of law and Economics, 16*(1), 67-88.

[31] David, P. A. (2004). Can" Open Science" be Protected from the Evolving Regime of IPR Protections?. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift fur die gesamte Staatswissenschaft*, 9-34.

[32] De, S., & Nabar, P. (1991). *Economic implications of imperfect quality certification. Economics letters, 37*(4), 333-337.

[33] Eggleston, K., Ling, L., Qingyue, M., Lindelow, M., & Wagstaff, A. (2008). Health service delivery in China: a literature review. *Health economics, 17*(2), 149-165.

[34] Ellis, R. P., & McGuire, T. G. (1993). Supply-side and demand-side cost sharing in health care. *The Journal of Economic Perspectives, 7*(4), 135-151.

[35] Emons, W. (1997). Credence goods and fraudulent experts. *The RAND Journal of Economics,* 107-119.

[36] Emons, W. (2001). Credence goods monopolists. *International Journal of Industrial Organization, 19*(3), 375-389.

[37] Eom, Y. S. (1994). Pesticide residue risk and food safety valuation: a random utility approach. *American Journal of Agricultural Economics, 76*(4), 760-771.

[38] Evans, D. S. (2003). Some empirical aspects of multi-sided platform industries. *Review of Network Economics, 2*(3).

[39] Gallini, N. T. (2002). The economics of patents: Lessons from recent US patent reform. *The Journal of Economic Perspectives, 16*(2), 131-154.

[40] Giannakas, K. (2002). Information asymmetries and consumption decisions in organic food product markets. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie, 50*(1), 35-50.

[41] Gravelle, H. (1999). Capitation contracts: access and quality. *Journal of Health Economics, 18*(3), 315-340.

[42] Gravelle, H., & Masiero, G. (2000). Quality incentives in a regulated market with imperfect information and switching costs: capitation in general practice. *Journal of Health Economics, 19*(6), 1067-1088.

[43] Green, J. R., & Scotchmer, S. (1995). On the division of profit in sequential innovation. *The RAND Journal of Economics*, 20-33.

[44] Hall, B. H., & Ziedonis, R. H. (2001). The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995. *RAND Journal of Economics,* 101-128.

[45] Hardin, G. (2009). The Tragedy of the Commons*. *Journal of Natural Resources Policy Research, 1*(3), 243-253.

[46] Heller, M. A. (1998). The tragedy of the anticommons: property in the transition from Marx to markets. *Harvard law review,* 621-688.

[47] Heller, M. A., & Eisenberg, R. S. (1998). Can patents deter innovation? The anticommons in biomedical research. Science, 280(5364), 698-701.

[48] Holland, D., & Wessells, C. R. (1998). Predicting consumer preferences for fresh salmon: the influence of safety inspection and production method attributes. *Agricultural and Resource Economics Review*, 27, 1-14.

[49] Huang, K. G., & Murray, F. E. (2009). Does patent strategy shape the long-run supply of public knowledge? Evidence from human genetics. *Academy of Management Journal, 52*(6), 1193-1221.

[50] Jeon, D. S., & Rochet, J. C. (2010). The pricing of academic journals: A two-sided market perspective. *American Economic Journal: Microeconomics, 2*(2), 222-255.

[51] Lessig, L. (2002). The future of ideas: The fate of the commons in a connected world. *Vintage.*

[52] Lizzeri, A. (1999). Information revelation and certification intermediaries. *The RAND Journal of Economics,* 214-231.

[53] Llanes, G., & Trento, S. (2009). Anticommons and optimal patent policy in a model of sequential innovation (No. 09-148). Harvard Business School.

[54] Lyon, T. P. (1999). Quality competition, insurance, and consumer choice in health care markets. *Journal of Economics & Management Strategy, 8*(4), 546-580.

[55] Ma, C. T. A. (1994). Health care payment systems: cost and quality incentives. *Journal of Economics & Management Strategy, 3*(1), 93-112.

[56] Ma, C. T. A., & Burgess Jr, J. F. (1993). Quality competition, welfare, and regulation. *Journal of Economics, 58*(2), 153-173.

[57] Ma, C. T. A., & McGuire, T. G. (1997). Optimal health insurance and provider payment. *The American Economic Review,* 685-704.

[58] Ma, C. T. A., & McGuire, T. G. (2002). Network incentives in managed health care. *Journal of Economics & Management Strategy, 11*(1), 1-35.

[59] Marette, S., Bureau, J. C., & Gozlan, E. (2000). Product safety provision and consumers' information. *Australian Economic Papers, 39*(4), 426-441.

[60] Maskin, E., & Tirole, J. (1992). The principal-agent relationship with an informed principal, II: Common values. *Econometrica: Journal of the Econometric Society,* 1-42.

[61] Matthews, S., & Postlewaite, A. (1985). Quality testing and disclosure. *The RAND Journal of Economics,* 328-340.

[62] McCluskey, J. (2000). A game theoretic approach to organic foods: An analysis of asymmetric information and policy. *Agricultural and Resource Economics Review, 29*(1), 1-9.

[63] Miceli, T. J., & Segerson, K. (1995). Defining efficient care: The role of income distribution. J. Legal Stud., 24, 189.

[64] Miller, R. H., & Luft, H. S. (1997). Does managed care lead to better or worse quality of care?. *Health affairs, 16*(5), 7-25.

[65] Mojduszka, E. M., & Caswell, J. A. (2000). A test of nutritional quality signaling in food market prior to implementation of mandatory labeling. *American Journal of Agricultural Economics, 82*(2), 298-309

[66] Mowery, D. C., Nelson, R. R., Sampat, B. N., & Ziedonis, A. A. (2001). The growth of patenting and licensing by US universities: an assessment of the effects of the Bayh–Dole act of 1980. *Research policy, 30*(1), 99-119.

[67] Mowery, D. C., Sampat, B. N., & Ziedonis, A. A. (2002). Learning to patent: Institutional experience, learning, and the characteristics of US university patents after the Bayh-Dole Act, 1981-1992. *Management Science,* 73-89.

[68] Mowery, D. C., & Ziedonis, A. A. (2002). Academic patent quality and quantity before and after the Bayh-Dole act in the United States. *Research Policy, 31*(3), 399-418.

[69] Mukherjee, A., & Stern, S. (2009). Disclosure or secrecy? The dynamics of open science. *International Journal of Industrial Organization, 27*(3), 449-462.

[70] Murray, F., & Stern, S. (2007). Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior & Organization, 63*(4), 648-687.

[71] Mysliveček, J. (2008). Comparing Certification and Self-regulation. *CERGE-EI* Working Paper Series, (361).

[72] Newhouse, J. P. (1996). Reimbursing health plans and health providers: efficiency in production versus selection. *Journal of economic literature, 34*(3), 1236-1263.

[73] Noel, M. D., & Schankerman, M. (2006). Strategic patenting and software innovation. *LSE STICERD Research Paper No. EI43.*

[74] Nuscheler, R. (2003). Physician reimbursement, time consistency, and the quality of care. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift fur die gesamte Staatswissenschaft, 159*(2), 302-322.

[75] Pezzino, M., & Pignataro, G. (2008). Competition in the health care market: a two-sided approach. Working paper.

[76] Rafferty, M. (2008). The Bayh–Dole Act and university research and development. *Research Policy, 37*(1), 29-40.

[77] Roberts, T., Buzby, J. C., & Ollinger, M. (1996). Using benefit and cost information to evaluate a food safety regulation: HACCP for meat and poultry. *American Journal of Agricultural Economics, 78*(5), 1297-1301.

[78] Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association, 1*(4), 990-1029.

[79] Rochet, J. C., & Tirole, J. (2006). Two-sided markets: a progress report. *The RAND Journal of Economics, 37*(3), 645-667.

[80] Roson, R. (2005). Two-sided markets: A tentative survey. *Review of Network Economics, 4*(2), 142-160.

[81] Sampat, B. N., Mowery, D. C., & Ziedonis, A. A. (2003). Changes in university patent quality after the Bayh–Dole act: a re-examination. *International Journal of Industrial Organization, 21*(9), 1371-1390.

[82] Scotchmer, S. (1991). Standing on the shoulders of giants: cumulative research and the patent law. *The Journal of Economic Perspectives, 5*(1), 29-41.

[83] Scotchmer, S. (1996). Protecting early innovators: should second-generation products be patentable?. *The Rand Journal of Economics,* 322-331.

[84] Segerson, K. (1999). Mandatory Versus Voluntary Approaches. *Agribusiness, 15*(1), 53-70.

[85] Shapiro, C. (1982). Consumer information, product quality, and seller reputation. *The Bell Journal of Economics,* 20-35.

[86] Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *The quarterly journal of economics, 98*(4), 659-679.

[87] Spence, M. (1974). *Market signalling.* Harvard University Press.

[88] Sporleder, T. L., & Goldsmith, P. D. (2001). Alternative firm strategies for signaling quality in the food system. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie, 49*(4), 591-604.

[89] Strausz, R. (2005). Honest certification and the threat of capture. *International Journal of Industrial Organization, 23*(1), 45-62.

[90] Thursby, J. G., & Thursby, M. C. (2003). University licensing and the Bayh-Dole act. *Science, 301*(5636), 1052-1052.

[91] Unnevehr, L. J., & Jensen, H. H. (1996). HACCP as a regulatory innovation to improve food safety in the meat industry. *American Journal of Agricultural Economics, 78*(3), 764-769.

[92] Unnevehr, L. J., & Jensen, H. H. (1999). The economic implications of using HACCP as a food safety regulatory standard. *Food policy, 24*(6), 625-635.

[93] Unnevehr, L. J., Miller, J. Y., & Gomez, M. I. (1999). Ensuring food safety and quality in farm-level production: emerging lessons from the pork industry. *American Journal of Agricultural Economics, 81*(5), 1096-1101.

[94] Viecens, M. F. (2006). Two-sided platforms with endogenous quality differentiation. *N.: UC3M Working Paper. Economics 2006-04.*

[95] Wang, Z., Mao, Y., & Gale, F. (2008). Chinese consumer demand for food safety attributes in milk products. *Food Policy, 33*(1), 27-36.

[96] Wessells, C. R., Johnston, R. J., & Donath, H. (1999). Assessing consumer preferences for ecolabeled seafood: the influence of species, certifier, and household attributes. *American Journal of Agricultural Economics, 81*(5), 1084-1089.

[97] Wolinsky, A. (1993). Competition in a market for informed experts' services. *The RAND Journal of Economics,* 380-398.

[98] Wolinsky, A. (1995). Competition in markets for credence goods. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift fur die gesamte Staatswissenschaft*, 117-131.

[99] Zago, A. M., & Pick, D. (2004). Labeling policies in food markets: Private incentives, public intervention, and welfare effects. *Journal of Agricultural and Resource Economics,* 150-165.

[100] Ziedonis, R. H. (2004). Don't fence me in: fragmented markets for technology and the patent acquisition strategies of firms. *Management Science, 50*(6), 804-820.

Résumé

De nombreux épisodes de l'histoire de la théorie économique ont suggéré l'effet positif des innovations sur la croissance économique et de qualité sur le bien-être des êtres humains. Les régulateurs ont tenté en permanence de concevoir des politiques publiques visant à encourager l'innovation et améliorer la prestation de qualité. Cette thèse se compose 3 chapitres indépendantes concernant les politiques publiques incentivizing innovations et prestations de qualité.

Dans le premier chapitre, nous modélisons la concurrence des plans de santé à deux faces avec externalité de qualité. Nous examinons ce qui se passe lorsque les fournisseurs de haute qualité sont autorisés à fréquenter les plans de santé multiples (multi-homing) et la comparer à une seule prise d'origine. Multi-homing des fournisseurs de haute qualité donne le plus la qualité des plans, mais seule-homing peut générer les meilleurs résultats pour le bien-être des assurés et de protection sociale.

Le deuxième chapitre compare deux approches réglementaires de qualité des aliments: échantillonnage et d'essai de produits (inspections sur les produits) et de contrôle vérifiables contrôle du processus de production (Certifications de processus). On peut se demander si la certification du processus est mieux utilisé comme substitut ou complément à l'inspection du produit. Ce chapitre analyse officiellement cette question dans le cadre Cournot et suggère que la combinaison de ces deux approches peut pas améliorer l'efficacité.

Le troisième chapitre est inspiré par l'hypothèse anticommuns de Heller et Eisenberg (1998), ce qui implique que les activités de recherche et développement (R & D) pourraient être entravées par la protection des brevets. Cependant, des études récentes suggèrent que cet effet est surestimée. Ce chapitre examine une situation courante valeur: les scientifiques de l'évaluation repose sur des uns et des autres points de vue. Il donne un résultat surinvestissement qui peut masquer la présence d'anti-communs, en particulier lorsque l'évaluation commune dépend largement des informations privées des deux scientifiques.