# Delivering universal health coverage for an aging population: an analysis of the Chinese rural health insurance program

**Wenhui Zhou**

*School of Business Administration, South China University of Technology,*
*Guangzhou 510640, Guangdong, China*
*whzhou@scut.edu.cn*

**Yanfang Zheng (Corresponding author)**

*Institute of Management Science and Engineering，School of Business, Henan University,*
*Kaifeng 475001, Henan, China*
*annie0605@foxmail.com*

**Alec Morton**

*Strathclyde Business School, University of Strathclyde,*
*Glasgow G1 1XQ, Scotland, United Kingdom*
*alec.morton@strath.ac.uk*

**Feng Luo**

*College of Business, City University of Hong Kong,*
*Kowloon Tong, Kowloon, Hong Kong*
*fengluo3-c@my.cityu.edu.hk*

# Delivering universal health coverage for an aging population: an analysis of the Chinese rural health insurance program

**Abstract**

There is now high level international commitment to the goal of universal health coverage. But how can countries make this a reality in the face of a limited budget and an aging population? Since 2008, China has been rolling out an ambitious reform program, which aims to achieve affordable health insurance coverage for all Chinese citizens. Under this reform program, Chinese living in rural areas are eligible to enroll in a subsidized scheme called the New Cooperative Medical System (NCMS). Using a three stage game model involving a government, a private fund manager and population, we explore the impact of population aging on NCMS. Our model highlights the role of government regulation and subsidy in ensuring operation efficiency of the system. We show that at optimality the government sets the operating framework for the fund manager to constrain the potential for monopoly profits. The Government subsidizes the scheme to prevent an adverse selection death spiral. However, the effectiveness of the subsidy in achieving this goal is moderated by the age structure of the population. Our model gives insights into the strengths of the NCMS framework and also can be used to support decisions about resource allocation and understand how scheme dynamics may unfold as the Chinese population ages further.

## 1 Introduction

Throughout history, people have lived in fear of the onset of disabling illness-not only because of the pain and suffering of the illness itself, but because the financial consequences of illness can be devastating. Since the beginning of the twentieth century, the advance of medical science has made the treatment and cure of many diseases possible, opening up the prospect of a future where everyone lives in good health to the natural limit of their lives.

This inspiring vision has not yet been realized. Medical technology is costly and often beyond the means of those who need it most. Perhaps the most heartening development of the early years of the twenty-first century has been the number of countries who have implemented, or set out on the path towards, Universal Health Coverage (UHC) - the principle that "all people can use the promotive, preventive, curative, rehabilitative and palliative health services they need, of sufficient quality to be effective, while also ensuring that the use of these services does not expose the user to financial hardship" (WHO, 2016). Many nations, such as Taiwan (Lu and Hisao 2003), India (Reddy et al. 2011) and Dutch (Rosenau and Lako 2008), are implementing different medical schemes to achieve universal health care.

The path to UHC is beset with challenges. In this paper we explore how population aging may threaten a country on its journey to UHC. Our focus will be on China - a country which has made a strong commitment to achieve UHC for its citizens, and which faces imminent and severe aging, as a legacy of the now-discontinued one-child policy. According to United Nation population projections, in 2030 for every Chinese senior aged 65 or over, there will be only five under 65s-as opposed to almost ten in 2015 (United Nations, 2015). Within China, we focus on the insurance system for the vulnerable rural population. As part of its commitment to social solidarity, the Chinese government provides financial support for this part of the population through its so-called New Cooperative Medical System (Wagstaff et al., 2009a; Yip et al., 2012).

As the purpose of this research is to explore the challenges faced by the NCMS in the next couple of decades, we take a mathematical modeling approach in this paper. Inevitably this involves assuming away certain complexities of the system in order to focus on strategic issues. Our model of the interactions between the different parties in the system is novel- as the organization of the Chinese system is distinctive, and has hitherto attracted little attention from theorists.

The rest of this paper is organized as follows. Section 2 describes the New Cooperative Medical System and places it an international context. Section 3 describes the assumptions, notation, and model. Section 4 discusses the solution procedure and the sensitivity of our results. Section 5 illustrates the core concepts of Section 4 with an empirical example. Section 6 discusses a generalization of the model wherein we explore a richer model of financing in which individuals contribute to the government pool through a universal tax. Section 7 concludes the paper.

## 2 Background of the New Cooperative Medical System

To finance the medical needs of its citizens for severe diseases or injuries, the Chinese central government proposed a primary medical security system (i.e., NCMS) in 2003. NCMS is a multi-channel fund-raising system comprising the government, collectives, and individuals. The goal is to protect the citizens from poverty due to the high medical costs. NCMS involves four parties, namely, the government, NCMS fund managers, designated hospitals, and villagers. Every September, the central government and the local government release a budget for NCMS and the local government issues a premium subsidy contract for the coming year. For instance, the NCMS budget for 2014 was 236.26 billion yuan, and the unit collected level was 410 yuan. The central government and the local government subsidized 320 yuan for each enrolled villager. Of course, the local government contains different level governments such as provincial governments, municipal governments and county-level governments. In this paper, to make the model analytically tractable, we combine each level government into a government.

After the collection of the premium and the subsidy, all of the collective funds are placed in a fund account named as the NCMS fund. The fund is used in reimbursing the medical fee of those enrolled villagers who access medical service from the designated hospitals. Hence, finding a

professional institution, also known as the NCMS fund manager to manage the fund, is necessary. The local government is responsible for selecting the fund manager. Most fund is operated by financial departments at the county level. Moreover, some insurance companies, such as China Life Insurance Company, China Pacific Insurance Company, and PICC Health Insurance Company Limited, are permitted to act as the NCMS fund managers. The fund managers maximize their profits by setting the reimbursement rate for the enrolled villagers. For instance, in 2014, the fund managers of Gansu and Sichuan provinces set the reimbursement rates at 60% and 70%, respectively. The designated hospitals are certified by the local health departments. Until 2004, nearly 310 villages had implemented the NCMS in China, and approximately 72.6% of the villagers had enrolled in the NCMS. The enrollment rate continued increasing and reached approximately 98.9% (almost 736 million villagers) by 2014.

In keeping with what is internationally recognized as a distinctive Chinese policy style (Heilmann 2009; Heilmann and Perry 2011), the roll-out of NCMS has been accompanied by the launch of local pilots and a concerted attempt by the central government to identify and diffuse good practice lessons. In keeping with this philosophy, several studies of local implementation have appeared in international journals (for reviews, see Eggleston et al., 2008 and Wagstaff et al, 2009b; for examples, see the studies in the Special Issue of Health Economics introduced by Yip et al., 2009). However, the existing research on NCMS were empirical and focused on the question of what worked and why it worked in a particular locality, e.g., Liu et al.(2011), Hu et al.(2012), Audibert et al.(2013).   As our intention in this paper is to explore broader questions, related to the behavior of the system as the population ages, we take a modeling approach, which is more theoretic in nature. There have been, to our knowledge, no modeling studies of NCMS (as well as the similar insurance schemes for the Chinese urban populations).

There is of course a long tradition of the modeling of health insurance systems. However, the preoccupations of this literature are different from the focus of our paper. In particular, in much of the international literature, the interest was in the competition between multiple insurance funds. In this paper we are interested in a setting where insurance funds do not compete directly for customers and the problem of (for example) funds seeking to divert high risk customers to other schemes do not arise. In the model we develop, the government offsets the tendency of firms to exploit their local monopoly position by setting contract terms in order to achieve efficiency.

This paper aims to answer the following questions: How does the government design the NCMS contract to maximize the number of enrollees? Does the age structure of the villagers matter when the government designs the contract? How does the fund manager respond to the contract issued by the government? To this end, we develop a three-stage game model involving three players, namely, a government, a fund manager, and villagers. We solve the game and derive the optimal contract of the government to demonstrate that (1) the optimal contract depends on the population and the age structure of the villagers, (2) the age structure of the villagers largely affects the implementation performance of the NCMS (the number of enrollees is larger with fewer or more aging citizens, and the enrollment rate reaches the least when the degree of aging among the citizens is moderate), and (3) the fund manager is forced/induced to choose a reimbursement rate such that the fund manager collects zero profit.

## 3 The Model

We model the players' interaction in NCMS of China as a three-stage game. The three players are the government, the fund manager and the villagers. The Chinese government's stated objective is to achieve universal coverage and so for the purposes of the model we assume that the government aims to maximize the number of enrollees.

In the first stage, the government offers a subsidy contract $(T, S, \theta)$, in which each villager who enrolls in the NCMS should pay a portion of the premium, $T(\geq 0)$, to the fund manager, and the government offers subsidy $S(\geq 0)$, which is turned over to the fund manager, for each enrollee as the remaining amount of the premium. Meanwhile, when the fund manager decides the reimbursement rate $\theta$, it has to meet a requirement on the reimbursement rate, $\theta \leq \underline{\theta}$ where $0 \leq \underline{\theta} \leq 1$. For instance, the central government announced in 2014 that the NCMS premium is 410 yuan, of which the government subsidizes 320 yuan for each enrollee and the enrollee pays for the remaining 90 yuan; meanwhile, the required reimbursement rate in Sichuan Province is $70\%$. Thus, the contract is T = 90 yuan, S = 320 yuan, and $\theta = 70\%$. For convenience, we refer to T and S as the premium and subsidy, respectively, throughout the paper. Furthermore, the central government is faced with a finance budget $B$; that is, the total amount of subsidy should be smaller than B. For instance, the NCMS budget in 2014 was 236.26 billion yuan. The goal of the government is to maximize the number of enrollees.

In the second stage of the game, we assume that the NCMS fund is charged by only one fund manager. The NCMS fund comprises the premium and subsidy of each enrollee. After a enrollee goes to the designated hospital for medical service, the fund manager should reimburse a portion $\theta$ of medical fee for her. The reimbursement rate $\theta$ should be no less than the rate required by the government $\underline{\theta}$, i.e., $\theta \geq \underline{\theta}$. For instance, in 2016, the fund managers of Shanxi Province and Inner Mongolia set $\theta = 75\%$; both rates are no lower than the governmental requirement $\underline{\theta} = 75\%$. The objective of the fund manager is to set the reimburse rate θ to maximize the profit.

In the final stage of the game, we assume that the total population of the rural residents is $\Lambda$ and each resident has a probability $p$ of seeking medical service from hospitals, i.e., visiting rate, where $p$ draws from $\mathscr{T}^k$ distribution with probability density function $f(p) = k(1-p)^{k-1}$ where $k > 0$ (see Debo, et al. 2005). Figure 1 plots the density function $f(p)$ for three different values of $k$. As $k$ increases, the mass of villagers shifts from the high probability region to the low probability region. Generally speaking, older villagers have a high probability of seeking medical service. In this sense, the smaller the value of $k$ is, the more elderly the population is. Hence, $k$ represents the degree of aging of the village population.
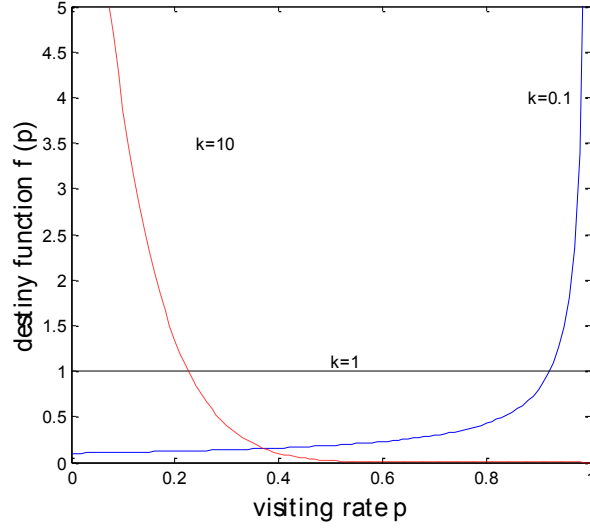
Figure 1: Examples of the $\mathscr{F}^k$ distribution: $k = 0.1,\ 1,\ 10$

We assume that the medical expenses of the villagers, denoted by $C$, are homogeneous. The villagers decide whether to enroll in NCMS based on the objective of minimizing their expected expenses. Denote by $E_1(p)(E_2(p))$ the expected expense of the villager with visiting rate $p$, when he/she enrolls (does not enroll) in NCMS. We assume that the villagers are risk neutral and rational. Hence, they are willing to enroll in NCMS if and only if $E_1(p) \le E_2(p)$. Intuitively, villagers will be more willing to enroll NCMS if they are risk averse. Thus, the assumption of risk neutrality allows us to identify other factors that affect villagers' behavior. Finally, the contract offered by the government and the distribution of visiting rate $p$ are common knowledge for the government, the fund manager, and all of the villagers, but a villager's visiting rate is her private information unobservable by other villagers, the fund manager, and the government.

# 4    Model Analysis

The government acts as a Stackelberg leader, and the fund manager and villagers act as the followers. To solve this problem, we use backward induction.

## 4.1 Villager behavior

In this section, we determine the choice of the villagers when they know the decisions of the government and the fund manager, i.e., the premium $T$, the subsidy $S$ and the reimbursement rate $\theta$. The net payoff of joining and not joining NCMS for a villager with visiting rate $p$ is respectively given by

$$E_1(p) = (1-\theta)pC + T,$$
$$E_2(p) = Cp$$

$\lambda$ denotes the expected number of enrollees and $\gamma$ denotes the expected number of enrollees who require medical service and receive reimbursement. The following theorem reports the equilibrium behavior of the villagers.

**Theorem 1.** *The equilibrium behavior of the villagers is given in the following cases:*

*1. When $T/C \le \theta \le 1$, a villager will enroll only when the visiting rate is high, i.e.,*

$p \ge \dfrac{T}{\theta C}$. *Correspondingly,* $\lambda = \Lambda(1-\dfrac{T}{\theta C})^k$ *and* $\gamma = \Lambda[\dfrac{T}{\theta C}(1-\dfrac{T}{\theta C})^k + \dfrac{(1-\dfrac{T}{\theta C})^{k+1}}{k+1}]$.

*2. When $0 \le \theta < T/C$, none of the villagers is willing to enroll. Correspondingly, $\lambda = 0$, and $\gamma = 0$.*

*Proof:* 1. When $\theta \ge T/C$, if $p \ge \dfrac{T}{\theta C}$, then $E_1(p) \le E_2(p)$, and these villagers will enroll in NCMS. Thus,

$$\lambda = \Lambda \int_{\frac{T}{\theta C}}^{1} f(p)dp = \Lambda(1-\frac{T}{\theta C})^k,$$

$$\gamma = \Lambda \int_{\frac{T}{\theta C}}^{1} pf(p)dp = \Lambda[\frac{T}{\theta C}(1-\frac{T}{\theta C})^k + \frac{(1-\frac{T}{\theta C})^{k+1}}{k+1}].$$

2. When $\theta < T/C$, or $\dfrac{T}{\theta C} > 1$, we obtain that for any $p$,

$$E_1(p) \ge Cp - Tp + T \ge Cp = E_2(p),$$

Such that on one wants to participate in NCMS. Thus, $\lambda = 0$, $\gamma = 0$.

According to Theorem 1, the enrolling decision of the villagers depends on the probability of seeking medical service. If the probability is high, the expected medical expense of the villagers is substantially high, thereby it is better off enrolling in NCMS which provides reimbursements. By contrast, if the villagers have a low probability of seeking medical service, the expected medical expense may be smaller than the premium, thus making enrolling in NCMS undesirable. This situation explains why elderly villagers have a stronger preference of enrolling in NCMS more than young villagers.

## 4.2 Decision of the fund manager

The goal of the fund manager is to maximize its expected profit by setting the reimbursement rate $\theta$ with the knowledge on the contract specified by the government. We denote $FM(\theta)$ as the expected profit of the fund manager, and the fund manager's problem can be formulated as

$$\max_{\theta} FM(\theta) = \lambda(T+S) - \gamma C\theta$$

$$s.t. \quad \underline{\theta} \le \theta \le 1$$

The following theorem presents the optimal decision of the fund manager.

**Theorem 2.** *The optimal reimbursement rate of the fund manager $\theta^*$ is given in the following cases:*

1. *When $0 \le T \le \max\{C\underline{\theta} - (k+1)S, 0\}$, the optimal strategy of the fund manager is to quit. In this case, the NCMS is not feasible. See region (I) in Figure 2.*

2. *When* $\max\{C\underline{\theta}-(k+1)S,0\}\le T<T_1(S)$, *the optimal strategy of the fund manager is to set the reimbursement rate as low as possible; that is,* $\theta^*=\underline{\theta}$. *See region* (*II*) *in Figure 2.*

3. *When* $T_1(S)\le T<T_2(S)$, *the optimal reimbursement rate is*

$$\theta^*=\frac{(1-k)T+\sqrt{(1+k)^2T^2+4kT(k+1)S}}{2C}.$$
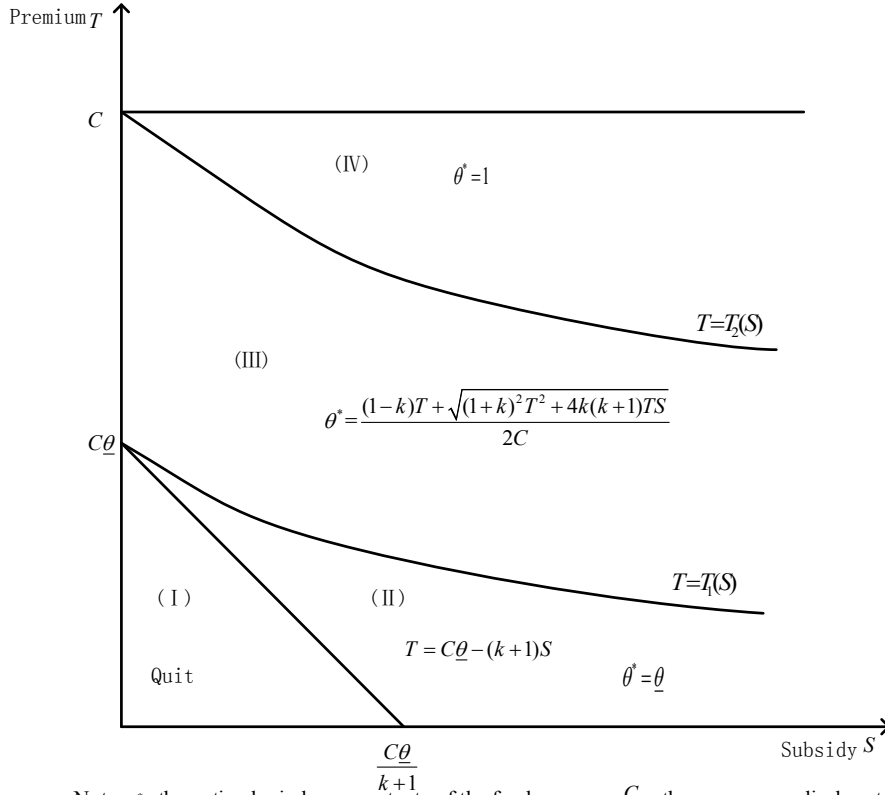
*See region* (*III*) *in Figure 2.*

4. *When* $T_2(S)\le T\le C$, *the optimal strategy of the fund manager is to set a maximum reimbursement rate, that is,* $\theta^*=1$. *See region* (*IV*) *in Figure 2.*

*In particular,*

$$T_1(S)=\frac{-k(k+1)S-C\underline{\theta}(1-k)+\sqrt{k^2(k+1)^2S^2+C^2\underline{\theta}^2(k+1)^2+2C\underline{\theta}Sk(1-k)(1+k)}}{2k},$$

$$T_2(S)=\frac{-k(k+1)S-C(1-k)+\sqrt{k^2(k+1)^2S^2+C^2(k+1)^2+2CSk(1-k)(1+k)}}{2k}.$$

*Proof.* See the supporting information in Appendix A.



Note: $\theta^*$ :the optimal reimbursement rate of the fund manager; $C$ : the average medical cost for villagers; $\underline{\theta}$ :the reimbursement rate required by the government; $T_1(S)$, $T_2(S)$ is given in page 11; $k$ : the degree of aging for population

Figure 2: Optimal strategy of the fund manager

Apparently, when both the premium and subsidy are small, the amount of the NCMS funding is small as well. In this case, the fund manager cannot afford to reimburse the medical service received by the enrolled residents. Hence, the fund manager quits the NCMS (because of the negative expected profit). Consider the scenario in which the fund manager does not quits the NCMS fund, i.e., $T\ge\max\{C\underline{\theta}-(k+1)S,0\}$. When the premium is small (i.e., $T\le T_1(S)$, the fund manager will limit the reimbursement rate at $\underline{\theta}$ to maximize its profit. When the premium is

between $T_1(S)$ and $T_2(S)$, the fund manager sets a reimbursement rate larger than the requirement of the government; that is, $\theta^* = \dfrac{(1-k)T + \sqrt{(1+k)^2 T^2 + 4kT(k+1)S}}{2C} > \underline{\theta}$. When the premium is larger (i.e., $T \geq T_2(S)$), the fund manager reimburses the entire medical fee of the enrolled residents; that is, $\theta^* = 1$. In fact, we can show that for any given subsidy $S$, the reimbursement rate increases in the premium $T$.

## 4.3 Government decision

In this section, we study the government's decision on contract design $(T, S, \theta)$. Ensuring a required reimbursement rate (i.e., $\theta \geq \theta_0$), the government aims to maximize the number of enrollees, within a finance budget $B$. According to Theorem 1, the government's problem can be formulated as

$$
\begin{aligned}
\max_{T,S,\underline{\theta}} \ & \lambda(T,S,\underline{\theta}) = \Lambda(1 - \frac{T}{C\theta})^k, \\
s.t. \ & \lambda(T,S,\underline{\theta})S \leq B, \\
& 0 \leq T \leq C, \\
& \theta_0 \leq \underline{\theta} \leq 1, \\
& S \geq 0.
\end{aligned}
$$

**Theorem 3.** *The government's optimal contract $(T^*, S^*, \underline{\theta}^*)$ is given in the following cases.*

1. *When $B < \dfrac{\Lambda C\theta_0}{k+1}$, then*

   □□□□□□□

$$
\begin{cases}
T^* = C\theta_0 - (k+1)(\dfrac{B}{\Lambda})^{\frac{1}{k+1}}(\dfrac{C\theta_0}{k+1})^{\frac{k}{k+1}}, \\
S^* = (\dfrac{B}{\Lambda})^{\frac{1}{k+1}}(\dfrac{C\theta_0}{k+1})^{\frac{k}{k+1}}, \\
\underline{\theta}^* = \theta_0.
\end{cases}
$$

*In this case, the associated number of enrollees is* $\lambda^* = \Lambda^{\frac{1}{k+1}}(\dfrac{(k+1)B}{C\theta_0})^{\frac{k}{k+1}}$.

2. *When $B \geq \dfrac{\Lambda C\theta_0}{k+1}$, then*

   □□□□□□□

$$\begin{cases} T^* = 0, \\ S^* = \dfrac{B}{\Lambda}, \\ \underline{\theta}^* = \min\{\dfrac{(k+1)B}{\Lambda C}, 1\}. \end{cases}$$

*In this case, all of the villagers enroll in NCMS, i.e., $\lambda^* = \Lambda$.*

*Proof.* See the supporting information in Appendix B.

**Remark 1.** *In Case 2 of Theorem 3, the finance budget is substantial (i.e., $B \geq \dfrac{\Lambda C \theta_0}{k+1}$) ,thus the government can guarantee that all of the villagers are involved in NCMS. Hence, the model (B) has infinite optimal solutions. In this case, we assume that the government shall choose the solution maximizing the reimbursement rate.*

Theorem 3 shows that whether NCMS is a multi-channel fund-raising system or a fully funded government system depends on finance budget $B$. When the budget is low( $B < \dfrac{\Lambda C \theta_0}{k+1}$ ), the optimal policy of the government is to run NCMS as a multi-channel fund-raising system. In this case, the government pays the subsidy, $(\dfrac{B}{\Lambda})^{\frac{1}{k+1}}(\dfrac{C\theta_0}{k+1})^{\frac{k}{k+1}}$, for each enrolled resident, and each enrolled resident pays the premium $C\theta_0 - (k+1)(\dfrac{B}{\Lambda})^{\frac{1}{k+1}}(\dfrac{C\theta_0}{k+1})^{\frac{k}{k+1}}$. In this case, only a portion of the residents will enroll in NCMS, which is the case in China. In 2011, the Chinese central government issued the NCMS contract as $T = 30$ yuan, $S = 200$ yuan, and $\underline{\theta} = 0.45$. According to China health statistics yearbook 2012, 832 million (around 97%) of the residents were enrolled in NCMS by the end of June 2011. When the budget is sufficiently large ( $B \geq \dfrac{\Lambda C \theta_0}{k+1}$ ), NCMS becomes a fully funded government system and achieves universal health coverage. In this case, the government sets a required reimbursement rate as large as possible; that is, $\underline{\theta} = \min\{\dfrac{(k+1)B}{\Lambda C}, 1\}$. Meanwhile, the government pays for the entire insurance premiums $\dfrac{B}{\Lambda}$ for the residents. In other words, the residents enroll in NCMS for free.

The threshold of the budget depends not only on the population of villagers, but also on the age structure of the village population. A population that is larger or has more aging villagers requires a substantial budget. In the subsequent discussion, we focus on the behavior of the fund manager under the optimal contract adopted by the government. It is easy to check that under the optimal contract, the condition specified in Case 2. in Theorem 2 holds (i.e., $\max\{C\underline{\theta}^* - (k+1)S^*, 0\} \leq T^* < T_1(S^*)$), and thus, the following result is immediate.

**Corollary 1.** *Under the optimal contract, the resulting fund manager's decision on the reimbursement rate is $\theta^* = \underline{\theta}$. In this case, the profit of the fund manager is zero.*

Taking the advantage of the first mover, the government would provide a contract that compels the fund manager to earn zero profit, though the fund manager chooses the most possibly low reimbursement rate $\theta$. Thus, effectively the government is setting contract terms for the fund manager which offset the fund manager's ability to exploit its monopoly position to earn excess profits, thus ensuring that the entire budget is channeled to the villagers.

**Proposition 1.**

1. $\partial T^* / \partial B \leq 0, \partial S^* / \partial B \geq 0, \partial \underline{\theta}^* / \partial B \geq 0, \partial \lambda(T^*, S^*, \underline{\theta}^*) / \partial B \geq 0$;

2. $\partial T^* / \partial \theta_0 \geq 0, \partial \underline{\theta}^* / \partial \theta_0 \geq 0, \partial S^* / \partial \theta_0 \geq 0, \partial \lambda(T^*, S^*, \underline{\theta}^*) / \partial \theta_0 \leq 0$;

3. When $B < \dfrac{\Lambda C \theta_0}{k+1}$, $\partial T^* / \partial k < 0, \partial \underline{\theta}^* / \partial k = 0, \partial^2 S^* / \partial k^2 < 0, \partial S^* / \partial k \, \pounds \, 0$,

   $\partial^2 \lambda(T^*, S^*, \underline{\theta}^*) / \partial^2 k > 0, \partial \lambda(T^*, S^*, \underline{\theta}^*) / \partial k \, \text{¤} \, 0$.

4. When $B = 0, \lambda = 0$.

*Proof.* See the supporting information in Appendix B3.

We draw attention first to part 4. of Proposition 1, which is important in understanding the scheme dynamics. Part 4. highlights that in the absence of government subsidy, the scheme is not viable: it exhibits an "adverse selection death spiral" (Cutler and Zeckhauser, 1998), whereby at equilibrium none enrolls in the scheme. The optimality of this solution reflects a dynamic whereby lower risk individuals opt out of the insurance program because they do not wish to cover the costs of their higher risk colleagues, resulting in a constantly shrinking enrollee base. In reality, health insurance is viable because of risk aversion, but in our model there is no risk aversion, and the only incentive for individuals to enroll is the subsidy.

The other parts of the proposition give other insights. Part 1. shows that, unsurprisingly, as the budget increases, the government can issue a more flexible contract, such as reducing the premium and increasing the subsidy. Part 2. shows that when the external required reimbursement rate increases, the premium, the subsidy, and the required reimbursement rate increases, but the number of enrollees decreases. Part 3. provides the answer to our main motivating questions about the impact of population aging. The population aging matters (i.e., the optimal contract and the number of enrollees vary with respect to $k$) only when the budget is insufficient to ensure full coverage. In this case, as the population ages, the premium becomes larger. Meanwhile, it is interesting to observe that the relationship between population aging and subsidy and the number of enrollees are not monotone. Specifically, $S$ is concave while $\lambda$ is convex in $k$. These results will be explored further in next section.

# 5  Empirical Example

The preceding analysis demonstrates that the degree of population aging $k$ plays an important role in the optimal contract and the number of enrollees. Thus, it is of great interest to estimate the value of $k$. In this section, we use the real data to do this. Similar to Kasteridis et al, (2014), we use the probability of getting a chronic disease to represent the probability of seeking medical

service. (In fact, it is very challenging to get the probability of seeking medical service for rural resident.) We first collect the population data in diff erent age ranges from China Population and Employment Statistics Yearbook 2014, and then collect the chronic diseases rate of rural residents in diff erent age ranges from China's health and family planning statistics yearbook 2015, which are listed in following table.

Table 1: Population profile

| Age | Chronic diseases(%) $x_i$ | Population $n_i$ | Percentage to Total Population(%) $p_i$ |
|---|---|---|---|
| years15-24 | 1.22 | 71107 | 17.1111 |
| years 25-34 | 3.82 | 68274 | 16.4295 |
| years 35-44 | 11.84 | 76718 | 18.4614 |
| years 45-54 | 23.00 | 77699 | 18.6974 |
| years 55-64 | 36.78 | 64670 | 15.5622 |
| over 65 years | 48.17 | 57091 | 13.7383 |

Based on $x_i$ and $p_i$ in the above table, we use $\chi^2$ test to check how well the $\mathscr{F}^k$ distribution fits the sampling distribution, where the parameter $k$ of $\mathscr{F}^k$ distribution is estimated by the maximum likelihood function. As convention (e.g., Turnbull 1976, Marwa et al. 2013), we divide the range of variation $(t_0, t_6)$ into 6 intervals. Let $t_1, t_2, t_3, t_4, t_5$ be the dividing points, such that $0 = t_0 < t_1 < t_2 < t_3 < t_4 < t_5 < t_6 = 1$ and $t_i \in (x_i, x_{i+1})(i = 1, 2, 3, 4, 5, 6)$ where $x_7 = 1$. The interval $(t_{i-1}, t_i)$ will be referred to as the $i$ th interval or the $i$ th group, and the points $t_i, (i = 0, 1, ..., 6)$ will be called group limits, the number of observation falling into one of the $i$ adjacent groups, say $i$, will be denoted by $n_i$. Then, the optimal $t_i$ and $k$ are derived from the following optimization problem.

$$\max_{t_i, k} L(t_i; k) = \sum_{i=1}^{6} n_i \ln(F(t_i) - F(t_{i-1}))$$

$$s.t. \qquad k > 0$$
$$t_0 = 0$$
$$t_6 = 1 \qquad\qquad (5.1)$$
$$x_i < t_i < x_{i+1}$$
$$x_7 = 1$$

Solving the above optimization problem yields solution as follows:
$$k = 4.4, \ t_0 = 0, t_1 = 0.0356, \ t_2 = 0.0863, \ t_3 = 0.1519, \ t_4 = 0.2438, \ t_5 = 0.3679, \ t_6 = 1 \qquad (5.2$$
Then, we can get the value of $F(t_i)$ associated with the solution. On this basis, we use $\chi^2$ test to check how well the $\chi^2$ distribution with $k = 4.4$ fits the sampling distribution. In this case, the null hypothesis is $k = 4.4$ and the $\chi^2$ statistic is
$$\sum_{i=1}^{6} \frac{(p_i - F(t_i) + F(t_{i-1}))^2}{p_i}$$

By computing, p−value for $\chi^2$ test is 1 (which is greater than 0.1), so we do not reject the null hypothesis, thereby indicating that the $\mathscr{F}^k$ distribution with $k = 4.4$ fits the real data very well.

The value of $k$ should be different across regions. However, because the specific data of villages or provinces is unavailable or inaccessible, we do not have a complete data set to estimate the value of $k$ for each region. Instead, we use the value estimated by the data aggregated into the nation level to represent the value of $k$ for different regions. Next, we use this estimated $k$ to demonstrate whether the contract currently implemented by the government is optimal.

In our model, the medical expense of each person within a population $C$ is homogeneous. The real data about per capita health expenditure for different regions in China can be found in China statistical yearbooks, and we take each region's per capita health expenditure in 2014 as the value of $C$ for the corresponding region. Also, the rural population of different regions can be found in some statistical yearbooks. For instance, one can find in Henan Statistical Yearbook 2014 that the rural population of Henan in 2013 was 82.56557 million. Moreover, each year's budget granted by the governments at all levels equals the difference between the amount of capital raised this year and the amount of personal financing. Table 2 summarizes these real data.

Table 2: Population, budget and per capita health expenditure

| Region | Total population ($\Lambda$, million person) | Budget this year ($B$, million yuan) | Per capita health expenditure (C, yuan) |
|--------|-----------------|-----------------|-----------------|
| Henan | 82.56557 | 23782.30 | 1791.68 |
| Sichuan | 62.92285 | 17939.70 | 2066.41 |
| Hainan | 4.94421 | 1413.75 | 2067.69 |
| Hunan | 47.78415 | 13746.19 | 1953.09 |
| Hubei | 39.56919 | 11976.84 | 2123.10 |
| Jiangxi | 34.07022 | 9481.22 | 1632.17 |
| Hebei | 56.59960 | 14731.16 | 2026.92 |
| Anhui | 51.18700 | 15858.26 | 2015.71 |
| Shanxi | 22.08497 | 6285.78 | 2018.84 |
| Chonging | 22.67730 | 8036.36 | 2482.61 |

From this table, we can see that per capita health expenditure is different across regions and the range is wide; for instance, the health expenditure per person was 2482.61 yuan in Chongqing , whereas is only 1632.17 yuan in Jiangxi. For this reason, the optimal contract for different regions should be different. The calculation result is presented in Table 3, of which we denote $T^*, S^*, \underline{\theta}^*$, and $UR(T^*, S^*, \underline{\theta}^*)$ as the premium, subsidy, required reimbursement rate and utilization of fund in our model. To distinguish the theoretical result and the real data, we use the superscript "$a$" to refer to the real data; that is, the actual premium, subsidy, required reimbursement rate and utilization fund are denoted by $T^a, S^a, \underline{\theta}^a$ and $UR(T^a, S^a, \underline{\theta}^a)$ respectively.

Table 3: Comparison between the actual data and the theoretical result

| Region | Premium (yuan) | Subsidy (yuan) | Required Reimbursement Rate(%) | Utilization Rate of Funds(%) |
|--------|----------------|----------------|--------------------------------|------------------------------|

| | $T^a$ | $T^*$ | $S^a$ | $S^*$ | $\underline{\theta}^a$ | $\underline{\theta}^*$ | $UR(T^a, S^a, \underline{\theta}^a)$ | $UR(T^*, S^*, \underline{\theta}^*)$ |
|---|---|---|---|---|---|---|---|---|
| Henan | 60 | 0 | 280 | 288 | 80 | 87 | 90.7 | 100 |
| Sichuan | 60 | 1.4 | 280 | 286.7 | 75 | 75 | 98.2 | 100 |
| Hainan | 60 | 1.2 | 280 | 286.9 | 75 | 75 | 97.3 | 100 |
| Hunan | 60 | 1.2 | 280 | 287.7 | 75 | 79.54 | 96.4 | 100 |
| Hubei | 60 | 0 | 280 | 302.7 | 75 | 76.99 | 94.1 | 100 |
| Jiangxi | 60 | 0 | 280 | 278.3 | 75 | 92.07 | 93.5 | 100 |
| Hebei | 60 | 2.5 | 280 | 262.3 | 70 | 70 | 98 | 100 |
| Anhui | 60 | 0 | 280 | 310.0 | 80 | 82.64 | 93.4 | 100 |
| Shanxi | 60 | 0 | 280 | 284.6 | 75 | 76.09 | 96.3 | 100 |
| Chongqing | 60 | 0 | 280 | 4 | 75 | 77.08 | 73.2 | 100 |

From Table 3, we find that in general, the government should lower the premium, improve the subsidy, and improve the reimbursement rate. Take Henan as an example, according to our model, the government should set premium at 0 (rather than 60) and subsidy at 288 (rather than 280); to improve the utilization of funds, the government should improve the reimbursement rate to 87% at which the fund will be used out. In addition, the actual premium and subsidy are almost the same across regions. Our model result, however, demonstrates that the government should offer different premium and subsidy for different regions. This is because the economic level, the health expenditure and the population are different across regions. We also observe that the actual subsidy is close to the theoretical one for each region, whereas the premium is not. Thus, reducing the premium can be the first step for the government to improve the efficiency of funds.

The government realized that the utilization extent of funds is low in many regions, and to address this, some regions have implemented so-called the secondary compensation policy for patients. To some extent, this post-adjustment measure improves the utilization of fund and benefit patients. However, it brings many issues meanwhile. The implement not only requires a vast of human and material resources, but also requires extra efforts from patients and medical institution, $e.g.$, the handling cost and travel cost. Moreover, the secondary reimbursement policy will be released at the end of a year, and thus, it cannot meet desperate need of low wealth patients. As Table 3 shows, the fund's utilization in Chongqing was 73.2% even after the secondary compensation policy was implemented. This highlights the important of designing the scheme optimally region by region.

# 6 Extensions

Consistent with the current goal of NCMS in China, the government's objective in the basic model is assumed to maximize the number of enrollees. Maximizing the populace welfare, in fact, can be another important key performance indicator for the government. In this extended model, we assume that the government aims to maximize the populace welfare, which consists of rewards and costs of receiving medical service. We assume that the reward is homogeneously given, and hence, maximizing the populace welfare is equivalent to minimizing the cost. Denote by the total payment of villagers TC, which contains the payment of villagers who enroll in NCMS $\Lambda(1-\frac{T}{C_\theta})^k((1-\theta)pC+T)$ and the payment of villagers who do not enroll in NCMS $\Lambda(1-(1-\frac{T}{C_\theta}))^k)pC$. Moreover, we assume that the fund for NCMS is collected from the medical service providers. Specifically, the government taxes the providers with a tax rate $r(0 < r \le r_0 < 1)$. Note that the tax policy does not affect the decision of the villagers and the fund manager directly. That is, in the presence of tax policy, when the contract is given, the villagers and the fund manager still act according to Theorems 1 and 2. Thus the amount of fund satisfies

$$\mathrm{B} = r\left[\Lambda\int_{\frac{T}{C_\theta}}^1 (1-\theta)CpdF(p) + \Lambda\int_0^{\frac{T}{C_\theta}} CpdF(p)\right],$$

and the government's optimization problem can be formulated as

$$\min_{r,T,S,\underline{\theta}} TC = \Lambda\int_{\frac{T}{C_\theta}}^1 ((1-\theta)Cp+T)dF(p) + \Lambda\int_0^{\frac{T}{C_\theta}} CpdF(p)$$

$$s.t. \quad \Lambda(1-\frac{T}{C_\theta})^k S \le \mathrm{B}$$

$$\mathrm{B} = r\left[\Lambda\int_{\frac{T}{C_\theta}}^1 (1-\theta)CpdF(p) + \Lambda\int_0^{\frac{T}{C_\theta}} CpdF(p)\right] \qquad\qquad (E)$$

$$0 \le T \le C$$

$$S \ge 0$$

$$\theta_0 \le \underline{\theta} \le 1$$

$$0 < r \le r_0$$

Solving the government's problem yields the optimal contract given as follows.

**Theorem 4.** When the objective of the government is to maximize the populace welfare, the optimal contract is as follows.
 (1.) If $r_0 < 1/k\theta_0$, then $\underline{\theta}^* = 1/kr_0, S^* = S_1, T^* = C/kr_0 - (k+1)S_1, r^* = r_0$.
 (2.) If $r_0 > 1/k\theta_0$, then $\underline{\theta}^* = \theta_0, S^* = S_1, T^* = C\theta_0 - (k+1)S_1, r^* = r_0$.

Theorem 4 demonstrates that the optimal tax rate equals the most possibly large value $r_0$ such that the government can increase the amount of funding as much as possible. Then, the government would use out the funding to subsidize the villagers such that their welfare would be increased. Interestingly, Theorem 4 demonstrates that when the tax rate is low, a high reimbursement rate is adopted. It is somehow counterintuitive, but reasonable because a high

premium is charged.

# 7 Conclusion

In this paper we have presented a novel model of China's New Cooperative Medical System (NCMS). We believe that our model is insightful and captures important features of the dynamics as the dynamics of the NCMS. Our model highlights the important role of the government in preventing the local insurance fund from exploiting its local monopoly through setting contract terms, and preventing an adverse selection death spiral through the provision of subsidy to enrollees. Our model highlights important implications for China's pathway to universal coverage as the population ages. Obviously, and not surprisingly, as the population grows older, the need for healthcare will increase, and so will associated expenditures. Unless central government budgets increase, this expenditure will fall back on the enrollees. Using actual data to calibrate the model, we find that the premium is higher while the reimbursement rate is lower in the contract currently implemented by the government than these in the theoretically optimal one. Meanwhile, the subsidy is in general lower than the optimum. Moreover, our model result demonstrates that the government should offer different premium and subsidy for different regions, rather than providing the uniform scheme of premium and subsidy. One of the perhaps surprising features of the model is that the relationship between population aging, i.e., older populations do not necessarily have lower enrollment. This highlights an important insight, namely that as the population ages, the costs of health insurance increase, but so do the benefits. However the age structure of the population interacts with ability of the subsidy to mitigate an adverse selection death spiral, giving rise to the nonmonotonic relationship which we see in Proposition 1. The scenario in which the government aims to maximize the populace welfare and the government finances the fund via taxation is discussed. The result shows that the government would charge a tax rate as high as possible to subsidize the enrollees, but a high tax rate does not lead to a high reimbursement rate.

Our study can be extended in several directions. First, considering the case that the villagers don't see a doctor when Second, for ease of exploration, we set the probability of seeking medical service obeying a distribution according to relationship between the aging of degree and the probability of requiring medical service, in further, we can extend this specific distribution function to a general distribution function. Lastly, although the details of the model are specific to China, the policy issue is of much more general concern. An interesting question for further research would be to compare the impact of population aging on the healthcare systems or other countries which are striving for universal coverage, in order to better understand the strengths and weaknesses of different national models. We also hope that this paper stimulates further theoretic studies of the Chinese healthcare system. Considerable care and attention has gone into the design of the Chinese system and it merits greater international study and attention than it has received hitherto.

# Acknowledgments

# References

Audibert, M., Mathonnat. J., Pelissier. A., Huang, X. X., & Ma, A. (2013). Health insurance reform and efficiency of township hospitals in rural China: An analysis from survey data. China Economic Review, 27: 326-338.

Bisceglia, M., Cellini, R., & Grilli, L. (2017). Regional regulators in healthcare service under quality competition: A game theoretical model.

China Population Employment Statistics Yearbook 2014. National Bureau of Statistics of the People's Republic of China, China Statistics Press; 2014.

China's health and family planning statistics yearbook 2015. National Health and Family Planning Commission of the People's Republic of China, China Union Medical University Press; 2015.

Cutler, D. M., & Zeckhauser, R. J.(1998). Adverse Selection in Health Insurance. In: Garber AM (Ed), Frontiers in Health Policy Research. MIT: Boston.

Debo, L. G., Toktay, L. B., & Van Wassenhove, L. N.(2005). Market segmentation and product technology selection for remanufacturable products. Manage Science, 51: 1193-1205.

Eggleston, K., Ling, L., Meng, Q. Y., Lindelow, M., & Wagstaff, A.(2008). Health service delivery in China: A literature review. Health Economic,17: 149-165.

Heilmann, S.(2009). Maximum Tinkering under Uncertainty Unorthodox Lessons from China. Mod China, 35(4): 450-462.

Heilmann, S., & Perry, E. J.(2011). Mao's Invisible Hand: The Political Foundations of Adaptive Governance in China. (Eds), Harvard Contemporary China Series. Harvard University Asia Center: Boston, MA.

Hu, H. H., Qi, Q., &Yang, C. H.(2012). Analysis of hospital technical efficiency in China:Effect of health insurance reform. China Economic Review, 23(4): 865-877.

Kateridis, P., Street, A., Dolman, M., Gallier, L., Hudson, K., Martin, J., & Wyer, I.(2014). The importance of multimorbidity in explaining utilisation and costs across health and social care

settings: evidence from South Somerset's Symphony Project. CHE Research Paper 96. Centre for Health Economics, University of York: York.

Jones, A. M.(2012). The Elgar Companion to Health Economics, Second Edition. (Ed). Elgar: Cheltenham.

Liu, H., Gao, S., & Rizzo, J. A.(2011). The expansion of public health insurance and the demand for private health insurance in rural China. China Economic Review, 22(1): 28-41.

Lu, J. F. R., & Hsiao, W. C. (2003). Does universal health insurance make health care unaffordable? Lessons from Taiwan. Health affairs, 22(3), 77-88.

Marwa, A. A., Zaher, H., & Elsherpieny, E. A. (2013). Optimum group limits for maxi- mum likelihood estimation of the exponentiated Frchet distribution based on grouped data. British Journal of Applied Science & Technology, 3(4), 1464.

Reddy, K. S., Patel, V., Jha, P., Paul, V. K., Kumar, A. K., & Dandona, L. (2011). Towards achievement of universal health care in India by 2020: a call to action. Lancet, 377, 760C768.

Rosenau, P. V., & Lako, C. J. (2008). An experiment with regulated competition and individual mandates for universal health care: the new Dutch health insurance system. Journal of Health Politics, Policy and Law, 33(6), 1031-1055.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. Journal of the Royal Statistical Society. Series B (Methodological), 290-295.

United Nations (2015). World population prospects, the 2015 revision.

Wagstaff, A., Lindelow, M., Wang, S., &Zhang, S.(2009). Reforming China?s Rural Health Systems. Wold Bank: Washington DC.

Wagstaff, A., Yip, W., Lindelow, M., &Hsiao, W. C.(2009). China's Health System and Its Reform: A Review of Recent Studies. Health Economics, 18: S7-S23.

WHO (2016). What is Universal Coverage?   http://www.who.int/health_financing/en/

Yip, W., Wagstaff, A., &Hsiao, W. C.(2009). Economic Analysis of China's Health Care System: Turning a New Page. Health Economics, 18(S2): S3-S6.

Yip, W. C. M., Hsiao, W. C., Chen, W., Hu, S. L., Ma, J.,& Maynard, A.(2012). Early appraisal of China's huge and complex health-care reforms. The Lancet, 379(9818): 833-842.

**Appendix A: Notations**

$$f = -4Ck^4(k+1)^2\beta((k+2)\beta - (k+1))S^3 + 2k^2C^2S^2(\beta^2(k^6(p-1) + k^5(3p-4) - k^4(p+4) - k^3(5p+12)$$
$$-k^2(2p+29) - 24k - 4) + \beta(-2k^6(p-1) + k^5(7-5p) + k^4(p+9) + 7k^3(p+3) + k^2(3p+35) + 22k + 4)$$
$$+(k+1)^3(k^3(p-1) - k^2p - 2k - 1)) + 2kC^3S(\beta^2(k^6(p-1)^2 + k^5(4p-2) + k^4(-p^2+10p+2) + 2k^3(2p+9)$$
$$+k^2(21-4p) - 4) + \beta(2k^6(p-1)^2 + k^5(8p-4) + k^4(-2p^2+19p+1) + k^3(4p+25) + k^2(23-7p) - 5k - 6)$$
$$+k^6(-(p-1)^2) + k^5(2-4p) + k^4(p^2-9p+1) - 9k^3 + 3k^2(p-2) + 3k + 2)$$
$$+4k^3(1-p)(\beta-1)C^4((k^2p+3k+2)\beta - (k^2p+2k+1))$$

$S_0$ is the root of $f = 0$ with respect to $S$, which minimize $TC$.

$$TC_1 = \beta\Lambda(\frac{k(1-p)(1-\beta)}{k+1-(2+k)\beta})^k \frac{k(1-\beta)(1-p)C\theta_0}{(k+1)(k+1-(2+k)\beta)} + (1-\beta)\Lambda(\frac{k(1-p)(1-\beta)}{k+1-(2+k)\beta})^k \frac{(1-p)(1-2\beta)C\theta_0}{k+1-(k+2)\beta}$$

$$TC_2 = \beta\Lambda(\frac{k((k+1)(1-p) + \beta(kp+p-k))}{(k+1)^2(1-\beta)})^k \frac{C\theta_0}{k+1}$$
$$+(1-\beta)\Lambda(\frac{k((k+1)(1-p) + \beta(kp+p-k))}{(k+1)^2(1-\beta)})^k \frac{(pk+p-2k-1)\beta + (k-1)(1-p))C\theta_0}{(k+1)^2(1-\beta)}$$

$$TC_3 = \beta\Lambda(\frac{2k+1-\sqrt{4k+1}}{2k})^k \frac{C\theta_0}{k+1} + (1-\beta)\Lambda(\frac{2k+1-\sqrt{4k+1}}{2k})^k \frac{\sqrt{4k+1} - (1+2kp)}{2k}$$

$$TC_4 = \beta(1 - \frac{T_2(S_0)}{C})^k S_0 + (1-\beta)(1 - \frac{T_2(S_0)}{C})^k (T_2(S_0) - Cp)$$

$$\beta_1 = \frac{(k+1)((2k^2p+3k+1) - (k+1)\sqrt{4k+1})}{5k^2 + 4k + 1 + 2k^2p + 2k^3p - (k+1)^2\sqrt{4k+1}}$$

$$\beta_2 = \frac{(k+1)(\sqrt{4k+1} - 1)((2k^2p+3k+1) - (k+1)\sqrt{4k+1})}{2((k^3p + k^2p + 3k^2 + 5k + 2)\sqrt{4k+1} - (k^3p + k^2p + 2k^3 + 7k^2 + 9k + 2))}$$

$$Q_1 = \{\beta \mid T_1 > 0, T_1 < T_4\} \bigcap ((\{\beta \mid T_1 < T_2\} \bigcap \{\beta \mid \beta < \beta_1\}) \bigcup (\{\beta \mid \beta_1 < \beta < \frac{1+pk}{2+pk}\} \bigcap \{\beta \mid T_1 < T_3\}))$$

$$Q_2 = \{\beta \mid T_2 > 0, T_2 < T_4\} \bigcap ((\{\beta \mid T_2 < T_1\} \bigcap \{\beta \mid \beta < \beta_1\}) \bigcup (\{\beta \mid \frac{1+pk}{2+pk} < \beta < \frac{(1+k)(1+kp)}{k^2p+kp+2k+1}\} \bigcap \{\beta \mid T_2 < T_3\}))$$

$$Q_3 = \{\beta \mid T_3 > 0, T_3 < T_4\} \bigcap ((\{\beta \mid T_3 < T_1\} \bigcap \{\beta \mid \beta < \beta_1 < \frac{1+pk}{2+pk}\})$$

$$\bigcup (\{\beta \mid \frac{1+pk}{2+pk} < \beta < \frac{(1+k)(1+kp)}{k^2p+kp+2k+1}\} \bigcap \{\beta \mid T_3 < T_2\}) \bigcup \{\beta \mid \beta > \frac{(1+k)(1+kp)}{k^2p+kp+2k+1}\})$$

$$Q_4 = \{\beta \mid f = 0 \text{ exist root}, \ T_4 > 0\} \bigcap \{\beta \mid T_4 < \min\{T_1, T_2, T_3\}\}$$

# Appendix B1:   the proof of Theorem 2

According to Theorem 1, $\lambda = 0$ when $\theta < T/C$. Hence, we only need to prove the case of $\theta \geq T/C$. In this case, the profit function of the fund manager is

$$FM(\theta) = \Lambda(1 - \frac{T}{\theta C})^k(T+S) - \Lambda[\frac{T}{\theta C}(1 - \frac{T}{\theta C})^k + \frac{(1 - \frac{T}{\theta C})^{k+1}}{k+1}]\theta C = \Lambda(1 - \frac{T}{\theta C})^k(S - \frac{C\theta - T}{k+1}).$$

Deriving the following equations is easy:

$$\frac{\partial FM(\theta)}{\partial \theta} = \Lambda(1-\frac{T}{C\theta})^{k-1}(-\frac{1}{k+1}C+\frac{(1-k)T}{(k+1)\theta}+\frac{kT(T+(k+1)S)}{(k+1)C\theta^2}),$$

$$\frac{\partial^2 FM(\theta)}{\partial \theta^2} = \frac{\Lambda(1-\frac{T}{C\theta})^{k-2}}{(k+1)C\theta^3}[-(k-1)^2T^2+(1-k)T^2-2kT(T+(k+1)S)+\frac{k(k+1)T^2(T+(k+1)S)}{C\theta}]<0.$$

According to the FOC, the maximizer of $FM(\theta)$ (denoted by $\theta_r$) is given by

$$\theta_r = \frac{(1-k)T+\sqrt{(1-k)^2T^2+4kT(T+(k+1)S)}}{2C}.$$

$\theta_r$ is increasing in $T$, and verifying that $T_1(S)$ is the unique root of $\theta_r=\theta$ with respect to $T$ and $T_2(S)$ is the unique root of $\theta_r=1$ with respect to $T$ is easy. Thus

$$T_1(S) > \frac{-k(k+1)S-C\underline{\theta}(1+k)+\sqrt{k^2(k-1)^2S^2+C^2\underline{\theta}^2(k+1)^2+2C\underline{\theta}Sk(1-k)(1+k)}}{2k}$$

$$= \frac{-k(k+1)S-C\underline{\theta}(1-k)+k(k-1)S+C\underline{\theta}(1+k)}{2k}$$

$$= C\underline{\theta}-S > C\underline{\theta}-(k+1)S,$$

$$T_1(S) < \frac{-k(k+1)S-C\underline{\theta}(1-k)+\sqrt{k^2(k+1)^2S^2+C^2\underline{\theta}^2(k+1)^2+2C\underline{\theta}Sk(1+k)(1+k)}}{2k}$$

$$= \frac{-k(k+1)S-C\underline{\theta}(1-k)+k(k+1)S+C\underline{\theta}(1+k)}{2k}$$

$$= C\underline{\theta},$$

$$T_2(S) > \frac{-k(k+1)S-C(1+k)+\sqrt{k^2(k-1)^2S^2+C^2(k+1)^2+2CSk(1+k)(1+k)}}{2k}$$

$$= \frac{-k(k+1)S-C(1-k)+k(k-1)S+C(1+k)}{2k}$$

$$> C-(k+1)S,$$

$$T_2(S) < \frac{-k(k+1)S-C(1-k)+\sqrt{k^2(k+1)^2S^2+C^2(k+1)^2+2CSk(1+k)(1+k)}}{2k}$$

$$= \frac{-k(k+1)S-C(1-k)+k(k+1)S+C(1+k)}{2k} = C.$$

Then, the decision model of the fund manager can be given by

$$\max_\theta FM(\theta) = \Lambda(1-\frac{T}{\theta C})^k(S-\frac{C\theta-T}{k+1}),$$

$$s.t. \quad \underline{\theta} \le \theta \le 1,$$

$$\theta \ge T/C. \tag{A.1}$$

Comparing $\underline{\theta}$ with $T=C$, Model (A.1) can be solved using the following two models:

$$\max_\theta FM(\theta) = \Lambda(1-\frac{T}{\theta C})^k(S-\frac{C\theta-T}{k+1}),$$

$$s.t. \quad \underline{\theta} \le \theta \le 1,$$

$$T < C\underline{\theta} \tag{A.2}$$

and

$$\max_\theta FM(\theta) = \Lambda(1-\frac{T}{\theta C})^k(S-\frac{C\theta-T}{k+1}),$$

$$s.t. \quad T/C \le \theta \le 1,$$

$$T \ge C\underline{\theta}. \tag{A.3}$$

We initially solve Model (A.2), and the solution can be classified into three subcases:

When $0 \le T < T_1(S)$, by the definition of $\theta_r$ and $\theta_r$ is increasing in $T$, $\theta_r < \underline{\theta}$. By the concavity of $FM(\theta)$, $FM(\underline{\theta}) \ge FM(\theta)$ for any $\theta \in [\underline{\theta}, 1]$. For $T \ge \max\{C\underline{\theta} - (k+1)S, 0\}$,

$$FM(\underline{\theta}) \ge \Lambda(1 - \frac{T}{\underline{\theta}C})^k (S - \frac{C\underline{\theta} - \max\{C\underline{\theta} - (k+1)S, 0\}}{k+1})$$

$$\ge \Lambda(1 - \frac{T}{\underline{\theta}C})^k (S - \frac{C\underline{\theta} - C\underline{\theta} + (k+1)S}{k+1})$$

$$= 0.$$

Recall that $C\underline{\theta} - (k+1)S < T_1(S) < C\underline{\theta}$, such that when $\max\{C\underline{\theta} - (k+1)S, 0\} < T_1(S) < C\underline{\theta}$, then

$\theta^* = \underline{\theta}$, For $0 < T < C\underline{\theta} - (k+1)S$ and $0 < S < \frac{C\underline{\theta}}{k+1}$, as $FM(\underline{\theta}) < 0$, the fund manager will quit

the market. The fund manager will quit the market when $0 < T < \max\{C\underline{\theta} - (k+1)S, 0\}$.

1. When $T_1(S) \le T \le \min\{T_2(S), C\underline{\theta}\}$, according to the definitions of $T_1(S)$ and $T_2(S)$, $\underline{\theta} \le \theta_r \le 1$, such that

$$FM(\theta_r) > FM(\underline{\theta}) \ge \Lambda(1 - \frac{T}{\underline{\theta}C})^k (S - \frac{C\underline{\theta} - C\underline{\theta} + (k+1)S}{k+1}) = 0.$$

Thus, $\theta^* = \theta_r$.

2. When $\min\{T_2(S), C\underline{\theta}\} < T < C\underline{\theta}$, $\theta_r > 1$ and $FM(\theta)$ is increasing in $[\underline{\theta}, 1]$. $T_2(S) > C-(k+1)S$, thus, for $\min\{T_2(S), C\underline{\theta}\} < T < C\underline{\theta}$,

$$FM(1) = \Lambda(1 - \frac{T}{C})^k (S - \frac{C-T}{k+1}) \ge \Lambda(1 - \frac{C\underline{\theta}}{C})^k (S - \frac{C-C+(k+1)S}{k+1}) = 0. \qquad (7.1)$$

We obtain $\theta^* = 1$.

Next, we solve Model (A.3). Note that

$$\theta_r = \frac{(1-k)T + \sqrt{(1+k)^2 T^2 + 4kT(k+1)S)}}{2C} > \frac{(1-k)T + (1+k)T}{2C} \ge \frac{T}{C} \ge \underline{\theta}$$

by the definition of $T_1(S)$, $T_1(S) < T$. Hence, Model (A.3) can be solved by the following two subcases:

1. When $\underline{\theta} < \theta_r \le 1$, that is, $T_1(S) < T \le T_2(S)$. $FM(\theta_r) > FM(\frac{T}{C}) = 0$ and $T_2 < C$; thus, if $T_2(S) \le C\underline{\theta}$, then $T_1(S) < T \le T_2(S)$ conflicts with the constraint that $T \ge C\underline{\theta}$, implying no feasible solutions; if $T_2(S) > C\underline{\theta}$, then for $C\underline{\theta} < T \le T_2(S)$, $\theta^* = \theta_r$. In summary, if $C\underline{\theta} < T \le \max\{T_2(S), C\underline{\theta}\}$, then $\theta^* = \theta_r$.

2. When $\theta_r > 1$, that is, $T > T_2(S)$. $FM(1) > 0$ ( see (7.1)) and the constraint $T \ge C\underline{\theta}$; thus, if $\max\{T_2(S), C\underline{\theta}\} < T < C$, then $\theta^* = 1$.

Combining the results of Models (A.2) and (A.3), we can obtain the result in Theorem 2.

# Appendix B2: the proof of Theorem 3

According to Theorem 2, the decision Model (B) can be divided into the following three models:

$$\max_{T,S,\theta} \lambda(T,S,\underline{\theta}) = \Lambda(1-\frac{T}{\underline{\theta}C})^k$$

$$s.t. \quad \Lambda(1-\frac{T}{\underline{\theta}C})^k S \le B \qquad\qquad (B.1)$$

$$\max\{C\underline{\theta}-(k+1)S,0\} \le T < T_1(S)$$

$$S \ge 0, \theta_0 \le \underline{\theta} \le 1$$

$$\max_{T,S,\underline{\theta}} \lambda(T,S,\underline{\theta}) = \Lambda(1-\frac{2}{(1-k)+\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}})^k S \le B$$

$$s.t. \quad \Lambda(1-\frac{2}{(1-k)+\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}})^k S \le B \qquad\qquad (B.2)$$

$$T_1(S) \le T \le T_2(S)$$

$$S \ge 0, \theta_0 \le \underline{\theta} \le 1$$

$$\max_{T,S,\underline{\theta}} \lambda(T,S,\underline{\theta}) = \Lambda(1-\frac{T}{C})^k$$

$$s.t. \quad \Lambda(1-\frac{T}{C})^k S \le B \qquad\qquad (B.3)$$

$$T_2(S) < T \le C$$

$$S \ge 0, \theta_0 \le \underline{\theta} \le 1$$



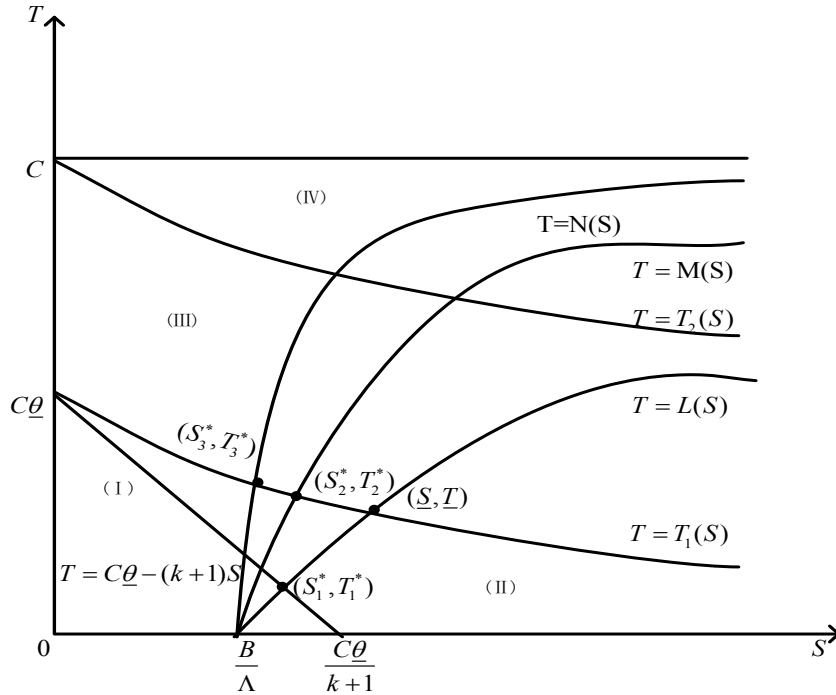Figure 3: The curves of $L(S)$, M(S), N(S)

For convenience, we denote $(T_i^*, S_i^*, \underline{\theta}_i^*)$ as the optimal solution of Model $(B.i)$ and define

$L(S) = C\underline{\theta} - C\underline{\theta}(\frac{B}{S\Lambda})^{\frac{1}{k}}$. We solve Models (B.1), (B.2), and (B.3) and begin with Model (B.1), which

can be divided into the following two models:

$$\max_{T,S,\underline{\theta}} \lambda(T,S,\underline{\theta}) = \Lambda(1-\frac{T}{\underline{\theta}C})^k$$

$$s.t. \quad T \geq L(S)$$

$$C\underline{\theta} - (k+1)S \leq T < T_1(S) \qquad \qquad \text{(B11)}$$

$$0 \leq S \leq \frac{C\underline{\theta}}{k+1}, \theta_0 \leq \underline{\theta} \leq 1$$

and

$$\max_{T,S,\underline{\theta}} \lambda(T,S,\underline{\theta}) = \Lambda(1-\frac{T}{\underline{\theta}C})^k$$

$$s.t. \quad T \geq L(S)$$

$$0 \leq T < T_1(S) \qquad \qquad \text{(B12)}$$

$$S \geq \frac{C\underline{\theta}}{k+1}, \quad \theta \leq \underline{\theta} \leq 1$$

We also denote $(T_{1i}^*, S_{1i}^*, \underline{\theta}_{1i}^*)$ as the optimal solution of Model$(B.i)$. To solve Model (B.11), we use the following two-step procedure:

1) Given $\theta$, we derive the optimum $(S_1^*(\underline{\theta}), T_1^*(\underline{\theta}))$ of Model (B.11);

2) Model (B.11) is recast into an optimal model with decision variable θ by substituting $(B,T)$ with $(S_1^*(\underline{\theta}), T_1^*(\underline{\theta}))$ into Model (B.11) .

We begin with Step 1 and solve the following model:

$$\max_{T,S} \lambda(T,S,\underline{\theta}) = \Lambda(1-\frac{T}{\underline{\theta}C})^k$$

$$s.t. \quad T \geq L(S) \qquad \qquad \text{(B.111)}$$

$$C\underline{\theta} - (k+1)S \leq T < T_1(S)$$

$$0 \leq S \leq \frac{C\underline{\theta}}{k+1}$$

$L(S) \in (-\infty, C\underline{\theta} - C\underline{\theta}(\frac{(k+1)B}{\Lambda C\underline{\theta}})^{\frac{1}{k}}]$ and is increasing with $S$; thus, we can solve Model (B.111) by

considering two different cases: $C\underline{\theta} - C\underline{\theta}(\frac{(k+1)B}{\Lambda C\underline{\theta}})^{\frac{1}{k}} \geq 0$ and $C\underline{\theta} - C\underline{\theta}(\frac{(k+1)B}{\Lambda C\underline{\theta}})^{\frac{1}{k}} < 0$,which can be

recast as $\underline{\theta} \geq \frac{(k+1)B}{\Lambda C}$ and $\underline{\theta} < \frac{(k+1)B}{\Lambda C}$. Let $(\hat{T}_{11a}^*, \hat{S}_{11a}^*, \hat{\underline{\theta}}_{11a}^*)$ $((\hat{T}_{11b}^*, \hat{S}_{11b}^*, \hat{\underline{\theta}}_{11b}^*))$ be the optimal

solution of Model (B.11) in the case $\underline{\theta} \geq \frac{(k+1)B}{\Lambda C}$ $(\underline{\theta} < \frac{(k+1)B}{\Lambda C})$.

a) Suppose $\underline{\theta} \geq \frac{(k+1)B}{\Lambda C}$. Recall that $T \geq \max\{C\underline{\theta} - (k+1)S, L(S)\}$ and $\lambda(S,T)$ is decreasing in $T$,

implying that $T_1^*(\underline{\theta}) = \min(\max\{C\underline{\theta} - (k+1)S, L(S)\})$.

Recall that $L(S) \in (-\infty, C\underline{\theta} - C\underline{\theta}(\frac{(k+1)B}{\Lambda C\underline{\theta}})^{\frac{1}{k}}]$ and is increasing with $S$ note that

$C\underline{\theta} - C\underline{\theta}(\frac{(k+1)B}{\Lambda C\underline{\theta}})^{\frac{1}{k}} \geq 0$. Moreover, $C\underline{\theta} - (k+1)S \in [0, C\underline{\theta}]$ and is decreasing in $S$. Thus, $T_1^*(\underline{\theta})$ is obtained at the intersecting point between $T = C\underline{\theta} - (k+1)S$ and $T = L(S)$ with respect to $S$, i.e.,

$S = (\frac{B}{\Lambda})^{\frac{1}{k+1}}(\frac{c\underline{\theta}}{k+1})^{\frac{k}{k+1}}$ (see Figure 3).

Hence, $S_1^*(\underline{\theta}) = (\frac{B}{\Lambda})^{\frac{1}{k+1}}(\frac{c\underline{\theta}}{k+1})^{\frac{k}{k+1}}$ and $T_1^*(\underline{\theta}) = L(S_1^*(\underline{\theta})) = C\underline{\theta} - (k+1)(\frac{B}{\Lambda})^{\frac{1}{k+1}}(\frac{c\underline{\theta}}{k+1})^{\frac{k}{k+1}}$. Regarding Step 2, incorporating $T_1^*(\underline{\theta})$ and $S_1^*(\underline{\theta})$ into Model (B.11), the model becomes

$$\max_{\underline{\theta}} \lambda(\underline{\theta}) = \Lambda(\frac{(k+1)B}{\Lambda C\underline{\theta}})^{\frac{1}{k+1}}$$
$$s.t. \quad \underline{\theta} \geq \frac{(k+1)B}{\Lambda C}, \tag{B.11a}$$
$$\theta_0 \leq \underline{\theta} \leq 1$$

$\lambda(\underline{\theta})$ is decreasing in $\underline{\theta}$ and thus the solution of Model (B.11a) (denoted by $\underline{\theta}_1^*$) is

$\underline{\theta}_1^* = \max\left\{\frac{(k+1)B}{\Lambda C}, \theta_0\right\}$.

According to this two-step procedure, $(\hat{T}_{11b}^*, \hat{S}_{11b}^*, \hat{\underline{\theta}}_{11b}^*) = \Lambda$, which can be specified as follows:

a-i) When $B < \frac{\Lambda C\theta_0}{k+1}$, i.e., $\frac{(k+1)B}{\Lambda C} < \theta_0$, then

$$(\hat{T}_{11a}^*, \hat{S}_{11a}^*, \hat{\underline{\theta}}_{11a}^*) = (C\theta_0 - (k+1)(\frac{B}{\Lambda})^{\frac{1}{k+1}}(\frac{C\theta_0}{k+1})^{\frac{k}{k+1}}, (\frac{B}{\Lambda})^{\frac{1}{k+1}}(\frac{C\theta_0}{k+1})^{\frac{k}{k+1}}, \theta_0),$$

and $\lambda(\hat{T}_{11a}^*, \hat{S}_{11a}^*, \hat{\underline{\theta}}_{11a}^*) = \Lambda^{\frac{1}{k+1}}(\frac{(k+1)B}{C\theta_0})^{\frac{k}{k+1}}$.

a-ii) When $\frac{\Lambda C\theta_0}{k+1} \leq B < \frac{\Lambda C}{k+1}$, $\theta_0 \leq \frac{(k+1)B}{\Lambda C} < 1$, then $\hat{S}_{11a}^* = \frac{C\hat{\underline{\theta}}_{11a}^*}{k+1}$, $\hat{S}_{11a}^* = \frac{B}{\Lambda} \in [\frac{C\theta_0}{k+1}, \frac{C}{k+1})$, $\hat{T}_{11a}^* = 0$, and $\lambda(\hat{T}_{11a}^*, \hat{S}_{11a}^*, \hat{\underline{\theta}}_{11a}^*) = \Lambda$.

a-iii) When $B \geq \frac{\Lambda C}{k+1}$, i.e., $\frac{(k+1)B}{\Lambda C} \geq 1$, this case has no solution.

b) Suppose $\underline{\theta} \leq \frac{(k+1)B}{\Lambda C}$, we can obtain the following:

b-i) When $B < \frac{\Lambda C\theta_0}{(k+1)}$, i.e., $\frac{(k+1)B}{\Lambda C} < \theta_0$, this case also has no solution.

b-ii) When $\frac{\Lambda C\theta_0}{(k+1)} \leq B < \frac{\Lambda C}{(k+1)}$, i.e., $\theta_0 \leq \frac{(k+1)B}{\Lambda C} < 1$, then $\hat{\underline{\theta}}_{11b}^* \in [\theta_0, \frac{(k+1)B}{\Lambda C})$, $\hat{S}_{11b}^* \in \frac{C\underline{\theta}^*}{k+1}$, $\hat{T}_{11b}^* \in 0$,

and $(\hat{T}_{11b}^*, \hat{S}_{11b}^*, \hat{\underline{\theta}}_{11b}^*) = \Lambda$.

b-iii) When $B \geq \frac{\Lambda C}{k+1}$, i.e., $\frac{(k+1)B}{\Lambda C} \geq 1$, then $\hat{\underline{\theta}}_{11b}^* \in [\theta_0, 1]$, $\hat{S}_{11b}^* = \frac{C\hat{\underline{\theta}}}{k+1}$, $\hat{T}_{11b}^* = 0$, and $\lambda(\hat{T}_{11b}^*, \hat{S}_{11b}^*, \hat{\underline{\theta}}_{11b}^*) = \Lambda$.

By combining cases a) and b) in Model (B.11), we can obtain the following:

1) When $B < \dfrac{\Lambda C\theta_0}{k+1}$, the optimal solution is

$$\lambda(T_1^*, S_1^*, \underline{\theta}_1^*) = (0, [\dfrac{C\underline{\theta}^*}{k+1}, \dfrac{B}{\Lambda}], [\theta_0, \dfrac{(k+1)B}{\Lambda C}])$$

then $\lambda(\hat{T}_{11a}^*, \hat{S}_{11a}^*, \hat{\underline{\theta}}_{11a}^*) = \Lambda^{\frac{1}{k+1}} (\dfrac{(k+1)B}{C\theta_0})^{\frac{k}{k+1}}$

2) When $\dfrac{\Lambda C\theta_0}{k+1} \le B \le \dfrac{\Lambda C}{k+1}$, the optimal solution is

$$(\hat{T}_{11}^*, \hat{S}_{11}^*, \hat{\underline{\theta}}_{11}^*) = (0, \dfrac{C\underline{\theta}_{11}^*}{k+1}, [\theta_0, \dfrac{(k+1)B}{\Lambda C}])$$

then $\lambda(T_{11}^*, S_{11}^*, \underline{\theta}_{11}^*) = \Lambda$

3) When $B \ge \dfrac{\Lambda C}{k+1}$, the optimal solution is

$$(\hat{T}_{11}^*, \hat{S}_{11}^*, \hat{\underline{\theta}}_{11}^*) = (0, \dfrac{C\underline{\theta}_{11}^*}{k+1}, [\theta_0, 1]) \cdot$$

then $\lambda(\hat{T}_{11}^*, \hat{S}_{11}^*, \hat{\underline{\theta}}_{11}^*) = \Lambda$

Now, we focus on Model (B.12). Similar to the proof of Model (B.11), we can also solve Model (B.12) in two cases: $\underline{\theta} \ge \dfrac{(k+1)B}{\Lambda C}$ and $\underline{\theta} < \dfrac{(k+1)B}{\Lambda C}$. We similarly let $(\hat{T}_{12a}^*, \hat{S}_{12a}^*, \hat{\underline{\theta}}_{12a}^*)$

$((\hat{T}_{12b}^*, \hat{S}_{12b}^*, \hat{\underline{\theta}}_{12b}^*))$ be the optimal solution of Model (B.12) when $\underline{\theta} \ge \dfrac{(k+1)B}{\Lambda C}$ $(\underline{\theta} < \dfrac{(k+1)B}{\Lambda C})$, thus:

a) Suppose $\underline{\theta} \ge \dfrac{(k+1)B}{\Lambda C}$, we can obtain the optimal solution as follows:

a-i) When $B < \dfrac{\Lambda C\theta_0}{k+1}$, i.e., $\dfrac{(k+1)B}{\Lambda C} < \theta_0$, then

$$(\hat{T}_{11a}^*, \hat{S}_{11a}^*, \hat{\underline{\theta}}_{11a}^*) = (C\theta_0 - C\theta_0(\dfrac{(k+1)B}{C\theta_0\Lambda})^{\frac{1}{k}}, \dfrac{C\theta_0}{k+1}, \theta_0)$$

and $\lambda(\hat{T}_{11a}^*, \hat{S}_{11a}^*, \hat{\underline{\theta}}_{11a}^*) = \dfrac{(k+1)B}{C\theta_0}$

a-ii) When $\dfrac{\Lambda C\theta_0}{k+1} \le B < \dfrac{\Lambda C}{k+1}$, i.e., $\theta_0 \le \dfrac{(k+1)B}{\Lambda C}$. Then $\hat{\underline{\theta}}_{11a}^* \in [\theta_0, 1], \hat{S}_{11a}^* = \dfrac{C\hat{\underline{\theta}}_{11a}^*}{k+1}, \hat{T}_{11a}^* = 0$, and

$\lambda(\hat{T}_{11a}^*, \hat{S}_{11a}^*, \hat{\underline{\theta}}_{11a}^*) = \Lambda \cdot$

a-iii) When $B \ge \dfrac{\Lambda C}{(k+1)}$, i.e., $\dfrac{(k+1)B}{\Lambda C} \ge 1$, this case has no solution.

b) Suppose $\theta_0 \le \dfrac{(k+1)B}{\Lambda C}$, we can obtain the following:

b-i) When $B < \dfrac{\Lambda C\theta_0}{(k+1)}$, i.e., $\dfrac{(k+1)B}{\Lambda C} < \theta_0$, this case also has no solution.

b-ii) When $\dfrac{\Lambda C\theta_0}{k+1} \le B < \dfrac{\Lambda C}{k+1}$, i.e., $\theta_0 \le \dfrac{(k+1)B}{\Lambda C} < 1$, then $\hat{\underline{\theta}}_{1b}^* \in [\theta_0, \dfrac{(k+1)B}{\Lambda C}), \hat{S}_{11b}^* = \dfrac{C\theta^*}{k+1}, \hat{T}_{11b}^* = 0$ and

$$\lambda(\hat{T}_{11b}^*, \hat{S}_{11b}^*, \hat{\underline{\theta}}_{1b}^*) = \Lambda.$$

b-iii)   When $B \geq \dfrac{\Lambda C}{k+1}$ ,   i.e.,   $\dfrac{(k+1)B}{\Lambda C} \geq 1$ ,   then $\hat{\underline{\theta}}_{1b}^* \in [\theta_0, 1], \hat{S}_{11b}^* = \dfrac{C\theta^*}{k+1}, \hat{T}_{11b}^* = 0$ ,   and

$$\lambda(\hat{T}_{11b}^*, \hat{S}_{11b}^*, \hat{\underline{\theta}}_{1b}^*) = \Lambda.$$

By combining cases a) and b) in Model (B.11), we can obtain the following:

1)   When $B < \dfrac{\Lambda C\theta_0}{k+1}$ , the optimal solution is $(T_{12}^*, S_{12}^*, \underline{\theta}_{12}^*) = (C\theta_0 - C\theta_0(\dfrac{(k+1)B}{C\theta_0\Lambda})^{\frac{1}{k}}), \dfrac{C\theta_0}{k+1}, \theta_0)$ , and

$$\lambda(T_{12}^*, S_{12}^*, \underline{\theta}_{12}^*) = \dfrac{(k+1)B}{C\theta_0}.$$

2)   When $B > \dfrac{\Lambda C\theta_0}{k+1}$ , the optimal solution is $(T_{12}^*, S_{12}^*, \underline{\theta}_{12}^*) = (0, [\dfrac{C\theta^*}{k+1}, \dfrac{B}{\Lambda}], [\theta_0, \dfrac{(k+1)B}{\Lambda C}])$ , and

$$\lambda(T_{12}^*, S_{12}^*, \underline{\theta}_{12}^*) = \Lambda$$

3)   When $B \geq \dfrac{\Lambda C}{k+1}$ , the optimal solution is $(T_{12}^*, S_{12}^*, \underline{\theta}_{12}^*) = (0, [\dfrac{C\theta^*}{k+1}, \dfrac{B}{\Lambda}], [\theta_0, 1])$ , and

$$\lambda(T_{12}^*, S_{12}^*, \underline{\theta}_{12}^*) = \Lambda$$

Comparing the solution of Model (B.11) to the solution of Model (B.12), we obtain the solution of Model(B.1) as follows:

(1)   When $B < \dfrac{\Lambda C\theta_0}{k+1}$ , as

$$\lambda(T_{11}^*, S_{11}^*, \underline{\theta}_{11}^*) = \Lambda^{\frac{1}{k+1}}(\dfrac{(k+1)B}{C\theta_0})^{\frac{k}{k+1}} > (\dfrac{(k+1)B}{C\theta_0})^{\frac{1}{k+1}}(\dfrac{(k+1)B}{C\theta_0})^{\frac{k}{k+1}} = \lambda(T_{12}^*, S_{12}^*, \underline{\theta}_{12}^*),$$

it follows that the solution of Model (B.1) is exactly the same as the solution of Model (B.11), that

is,  $(T_1^*, S_1^*, \underline{\theta}_1^*) = (C\theta_0 - (k+1)(\dfrac{B}{\Lambda})^{\frac{1}{k+1}}(\dfrac{C\theta_0}{k+1})^{\frac{k}{k+1}}, (\dfrac{B}{\Lambda})^{\frac{1}{k+1}}(\dfrac{C\theta_0}{k+1})^{\frac{k}{k+1}}, \theta_0)$

then  $\lambda(T_1^*, S_1^*, \underline{\theta}_1^*) = \Lambda^{\frac{1}{k+1}}(\dfrac{(k+1)B}{C\theta_0})^{\frac{k}{k+1}}.$

2) When $\dfrac{\Lambda C\theta_0}{k+1} \leq B < \dfrac{\Lambda C}{k+1}$ , as

$$\lambda(T_{11}^*, S_{11}^*, \underline{\theta}_{11}^*) = \Lambda = \lambda(T_{12}^*, S_{12}^*, \underline{\theta}_{12}^*)$$

then the solution of Model (B.1) is the same as the solution of Model (B.11) or (B.12), that is,

$$\lambda(T_1^*, S_1^*, \underline{\theta}_1^*) = (0, [\dfrac{C\theta^*}{k+1}, \dfrac{B}{\Lambda}], [\theta_0, \dfrac{(k+1)B}{\Lambda C}])$$

then $\lambda(T_1^*, S_1^*, \underline{\theta}_1^*) = \Lambda$

For convenience, we define

$$M(S) = \dfrac{k(k+1)S[1 - (\dfrac{B}{S\Lambda})^{\frac{1}{k}}]^2}{(\dfrac{B}{S\Lambda})^{\frac{1}{k}}(1 + k - k(\dfrac{B}{S\Lambda})^{\frac{1}{k}})}$$

We can easily obtain that $M(S)$ is increasing with $S$. Therefore, Model (B.2) can be divided into the following two models:

$$\max_{T,S,\underline{\theta}} \lambda(T,S,\underline{\theta}) = \Lambda(1 - \frac{2}{1-k+\sqrt{(1+k)^2 + 4k(k+1)\frac{S}{T}})})^k,$$

$$s.t. \quad T \geq M(S),$$
$$T_1(S) \leq T \leq T_2(S),$$
$$\theta_0 \leq \underline{\theta} \leq 1$$
$$S > \frac{B}{\Lambda}. \tag{B.11}$$

$$\max_{T,S,\underline{\theta}} \lambda(T,S,\underline{\theta}) = \Lambda(1 - \frac{2}{1-k+\sqrt{(1+k)^2 + 4k(k+1)\frac{S}{T}})})^k$$

$$s.t. \quad T \geq 0,$$
$$T_1(S) \leq T \leq T_2(S),$$
$$\theta_0 \leq \underline{\theta} \leq 1,$$
$$S \leq \frac{B}{\Lambda}. \tag{B.22}$$

Denote by $(T_{2i}^*, S_{2i}^*, \underline{\theta}_{2i}^*)$ the optimal solution of Model $(B.2i)$.

We initially solve Model $(B.21)$, which can be solved in the same manner as Model $(B.11)$ (i.e., the two-step procedure) in the sense that $L(S)$ is replaced by $M(S)$ and that $C\underline{\theta} - (k+1)$ is replaced by $T_1(S)$.

The optimal solution is given by

$$(T_{21}^*, S_{21}^*, \underline{\theta}_{21}^*) = (\frac{-k(k+1)S_{21}^* - C\theta_0(1-k) + \sqrt{k^2(k+1)^2 S_{21}^{*2} + C^2\theta_0^2(k+1)^2 + 2C\theta_0 S_{21}^* k(1-k)(1+k)}}{2k}, S_2^*(\theta_0), \theta_0) \text{ and}$$

$$\lambda(T_{21}^*, S_{21}^*, \underline{\theta}_{21}^*) = \Lambda(1 - \frac{T_{21}^*}{C\theta_0})^k.$$

The solution of Model $(B.22)$ is likewise given by

$$(T_{22}^*, S_{22}^*, \underline{\theta}_{22}^*) = \left( \frac{-k(k+1)\frac{B}{\Lambda} - C\theta_0(1-k) + \sqrt{k^2(k+1)^2(\frac{B}{\Lambda})^2 + C^2\theta_0^2(k+1)^2 + 2C\theta_0\frac{B}{\Lambda}k(1-k)(1+k)}}{2k}, \frac{B}{\Lambda}, \theta_0 \right),$$

and $\lambda(T_{22}^*, S_{22}^*, \underline{\theta}_{22}^*) = \Lambda(1 - \frac{T_{22}^*}{C\theta_0})^k.$

We subsequently compare the solution of Models $(B.21)$ and $(B.22)$. We define $f_1(S) = M(S) - T_1(S)$, then $f_1(S)$ is increasing with $S$ as $M(S)$ is increasing with $S$ and $T_1(S)$ is increasing with $S$. We can obtain $f_1(S_{22}^*) = f_1(\frac{B}{\Lambda}) < 0 = f_1(S_{21}^*)$, thus $S_{21}^* > S_{22}^*$. Furthermore, as $T_1(S)$ is increasing with $S$, then $T_{21}^* = T_1(S_{21}^*) < T_1(S_{22}^*) = T_{22}^*$ and

$$\lambda(T_{21}^*, S_{22}^*, \underline{\theta}_{22}^*) = \Lambda(1 - \frac{T_{22}^*}{C\theta_0})^k < \Lambda(1 - \frac{T_{21}^*}{C\theta_0})^k = \lambda(T_{21}^*, S_{21}^*, \underline{\theta}_{21}^*).$$

Thus, the solution of Model (B.2) is the same as the solution of Model (B.21), that is,

$$(T_2^*, S_2^*, \underline{\theta}_2^*) = (\frac{-k(k+1)S_2^* - C\theta_0(1-k) + \sqrt{k^2(k+1)^2 S_2^{*2} + C^2\theta_0^2(k+1)^2 + 2C\theta_0 S_2^* k(1-k)(1+k)}}{2k}, S_2^*(\theta_0), \theta_0) \text{ and}$$

$$\lambda(T_2^*, S_2^*, \underline{\theta}_2^*) = \Lambda(1 - \frac{T_2^*}{C\theta_0})^k.$$

Model (B.3), which can be also solved in a similar manner, and the solution is given by

$(T_3^*, S_3^*, \underline{\theta}_3^*) = (N(S_3^*), S_3^*, \theta_0)$, and $\lambda(T_3^*, S_3^*, \underline{\theta}_3^*) = \dfrac{B}{S_3^*}$, where $S_3^*$ is the unique root of the equation

$N(S) = T_2(S)$ with respect to $S$ and $N(S) = C - C(\dfrac{B}{S\Lambda})^{\frac{1}{k}}$.

Finally, we compare among the solution of Models (B.1), (B.2) and (B.3).

1. When $B < \dfrac{\Lambda C\theta_0}{k+1}$, from the solutions of Model (B.1), (B.2) and (B.3), it is clear that

$$\lambda(T_1^*, S_1^*, \underline{\theta}_1^*) = \Lambda(1 - \frac{T_1^*}{C\theta_0})^k = \frac{B}{S_1^*},$$

$$\lambda(T_2^*, S_2^*, \underline{\theta}_2^*) = \Lambda(1 - \frac{T_2^*}{C\theta_0})^k,$$

$$\lambda(T_3^*, S_3^*, \underline{\theta}_3^*) = \frac{B}{S_3^*},$$

and $\underline{\theta}_1^* = \underline{\theta}_2^* = \underline{\theta}_3^* = \theta_0$.

We initially compare between $(T_2^*, S_2^*, \underline{\theta}_2^*)$ and $(T_1^*, S_1^*, \underline{\theta}_1^*)$. Let point $(\underline{S}, \underline{T})$ satisfy $T_1(\underline{S})|_{\underline{\theta}=\theta_0} = L(\underline{S})|_{\underline{\theta}=\theta_0} = \underline{T}$.

Note that $T_1(S) > C\underline{\theta} - (k+1)S$ holds for any $S$ and $\theta$, $L(S)$ is increasing with $S$ and $L(S_1^*) = C\underline{\theta} - (k+1)S_1^*$ holds for any $\underline{\theta}$, it follows that $L(S_1^*) = C\underline{\theta} - (k+1)S_1^* < T_1(S_1^*)$, i.e., $T_1(S_1^*) - L(S_1^*) > 0$. Moreover, as $T_1(S)$ is decreasing in $S$, it follows that $T_1(S) - L(S)$ is decreasing in $S$. By the definition of $\underline{S}$, it follows that given $\underline{\theta} = \theta_0$, we obtain $T_1(S_1^*) - L(S_1^*) > T_1(\underline{S}) - L(\underline{S})$ holds, thus $\underline{S} > S_1^*$ and $\underline{T} = L(\underline{S}) > L(S_1^*) = T_1^*$ (See Figure 3).

We then define

$$f_2(S) = M(S) - L(S) = \frac{k(k+1)S[1 - (\frac{B}{S\Lambda})^{\frac{1}{k}}]^2}{(\frac{B}{S\Lambda})^{\frac{1}{k}}(1 + k - k(\frac{B}{S\Lambda})^{\frac{1}{k}})} - C\underline{\theta} + C\underline{\theta}(\frac{B}{S\Lambda})^{\frac{1}{k}},$$

which is increasing with $S$. Obviously, $f_2(\dfrac{B}{\Lambda}) = 0$. Thus, when $S > \dfrac{B}{\Lambda}$, it follows that $f_2(S) > 0$, that is, $M(S) > L(S)$. Recall that $T_1(S)$ is decreasing in $S$ and $T_1(S_2^*) = M(S_2^*)$ holds for any $\underline{\theta}$. Hence, $T_1(S_2^*) = M(S_2^*) > L(S_2^*)$, i.e., $T_1(S_2^*) - L(S_2^*) > 0$.

Since $T_1(S) - L(S)$ is decreasing in $S$. Given $\underline{\theta} = \theta_0$, by the definition of $\underline{S}$, we can obtain that $T_1(S_2^*) - L(S_2^*) > T_1(\underline{S}) - L(\underline{S}) = 0$, thus, we can obtain $\underline{S} > S_2^*$ and $T_2^* = T_1(S_2^*) > T_1(\underline{S}) = \underline{T}$.

Since $\underline{T} > T_1^*$, so we can get $T_2^* > T_1^*$, then

$$\lambda(T_1^*, S_1^*, \underline{\theta}_1^*) = \Lambda(1 - \frac{T_1^*}{C\theta_0})^k > \Lambda(1 - \frac{T_2^*}{C\theta_0})^k = \lambda(T_2^*, S_2^*, \underline{\theta}_2^*),$$

Which implies that $(T_2^*, S_2^*, \underline{\theta}_2^*)$ is dominated by $(T_1^*, S_1^*, \underline{\theta}_1^*)$.

We subsequently compare $(T_1^*, S_1^*, \underline{\theta}_1^*)$ with $(T_3^*, S_3^*, \underline{\theta}_3^*)$. We define $f_3(S) = N(S) - T_2(S)$. Recall that $N(S)$ is increasing with $S$ and $T_2(S)$ is decreasing in $S$. It is obvious that $f_3(S)$ is increasing with $S$. As

$$f_3(S_1^*) = -\frac{-k(k+1)S_1^* - C(1-k) + \sqrt{k^2(k+1)^2 S_1^{*2} + C^2(k+1)^2 + 2CS_1^* k(1-k)(1+k)}}{2k} + C - C(\frac{B}{S_1^* \Lambda})^{\frac{1}{k}}$$

$$< C - C(\frac{B}{S_1^* \Lambda})^{\frac{1}{k}} - \frac{2k(C - S_1^*)}{2k}$$

$$= C - C + S_1^* - C(\frac{B}{S_1^* \Lambda})^{\frac{1}{k}}$$

$$= (\frac{B}{\Lambda})^{\frac{1}{k+1}} (\frac{C\theta_0}{k+1})^{\frac{k}{k+1}} - C^{\frac{k}{k+1}} (\frac{(k+1)B}{\Lambda \theta_0})^{\frac{1}{k+1}}$$

$$= C^{\frac{k}{k+1}} (\frac{B}{\Lambda})^{\frac{1}{k+1}} ((\frac{\theta_0}{k+1})^{\frac{k}{k+1}} - (\frac{k+1}{\theta_0})^{\frac{1}{k+1}}) < 0$$

and $f_3(S_3^*) = 0$, then $S_1^* < S_3^*$. Thus, $\lambda(T_1^*, S_1^*, \underline{\theta}_1^*) = \frac{B}{S_1^*} > \frac{B}{S_3^*} = \lambda(T_3^*, S_3^*, \underline{\theta}_3^*)$, which implies that $(T_3^*, S_3^*, \underline{\theta}_3^*)$ is also dominated by $(T_1^*, S_1^*, \underline{\theta}_1^*)$.

In summary, the optimal solution of Model (B) is $(T_1^*, S_1^*, \underline{\theta}_1^*)$, that is,

$$(T^*, S^*, \underline{\theta}^*) = (C\theta_0 - (k+1)(\frac{B}{\Lambda})^{\frac{1}{k+1}} (\frac{C\theta_0}{k+1})^{\frac{k}{k+1}}, (\frac{B}{\Lambda})^{\frac{1}{k+1}} (\frac{C\theta_0}{k+1})^{\frac{k}{k+1}}, \ \theta_0).$$

In this case, $\lambda(T^*, S^*, \underline{\theta}^*) = \Lambda^{\frac{1}{k+1}} (\frac{(k+1)B}{C\theta_0})^{\frac{k}{k+1}}$.

2. When $B \geq \frac{\Lambda C\theta_0}{k+1}$, recall that $T_2^* > 0$, we can easily obtain that

$$\lambda(T_2^*, S_2^*, \underline{\theta}_2^*) \neq \Lambda \ (1 + \frac{T_2^*}{C\theta_0}^k) < \Lambda \neq T_1^* \ S_1^* \theta$$

Moreover, we can determine that $f_3(\frac{B}{\Lambda}) < 0 = f_3(S_3^*)$, thus $S_3^* > \frac{B}{\Lambda}$ and then

$$\lambda(T_3^*, S_3^*, \underline{\theta}_3^*) = \frac{B}{S_3^*} < \Lambda = \lambda(T_1^*, S_1^*, \underline{\theta}_1^*).$$

Therefore, the optimal solution of Model (B) is $(T_1^*, S_1^*, \underline{\theta}_1^*)$, i.e.,

$(T_1^*, S_1^*, \underline{\theta}_1^*) = (0, [\frac{C\underline{\theta}^*}{k+1}, \frac{B}{\Lambda}], [\theta_0, \min\{\frac{(k+1)B}{\Lambda C}, 1\}])$. In this case, $\lambda(T^*, S^*, \underline{\theta}^*) = \Lambda$. In this paper, we choose the solution which make reimbursement rate large as the optimal solution, i.e.,

$(T_1^*, S_1^*, \underline{\theta}_1^*) = (0, \frac{B}{\Lambda}, \min\{\frac{(k+1)B}{\Lambda C}, 1\})$. In this case, $\lambda(T^*, S^*, \underline{\theta}^*) = \Lambda$.

The proof of Theorem 3 is complete.

## Appendix B3: the proof of Proposition 1

For part 1-3, According to Theorem 3, we prove each part of this proposition in two cases, namely $B < \frac{\Lambda C\theta_0}{k+1}$, $B \geq \frac{\Lambda C\theta_0}{k+1}$.

1. When $B < \frac{\Lambda C\theta_0}{k+1}$,

$$\frac{\partial T^*}{\partial B} = -(\frac{B}{\Lambda})^{\frac{-k}{k+1}} \frac{1}{\Lambda} (\frac{C\theta_0}{k+1})^{\frac{k}{k+1}} < 0,$$

$$\frac{\partial \underline{\theta}^*}{\partial B} = 0,$$

$$\frac{\partial S^*}{\partial B} = \frac{1}{k+1} (\frac{B}{\Lambda})^{\frac{-k}{k+1}} \frac{1}{\Lambda} (\frac{C\theta_0}{k+1})^{\frac{k}{k+1}} > 0,$$

$$\frac{\partial \lambda(T^*, S^*, \underline{\theta}^*)}{\partial B} = \Lambda^{\frac{1}{k+1}} \frac{k}{C\theta_0} (\frac{(k+1)B}{C\theta_0})^{\frac{-1}{k+1}} > 0.$$

When $B \geq \frac{\Lambda C\theta_0}{k+1}$, $T^*, \underline{\theta}^*, S^*$ and $\lambda(T^*, S^*, \underline{\theta}^*)$ are all constant with respect to $B$, that is

$\frac{\partial T^*}{\partial B} = 0, \frac{\partial \underline{\theta}^*}{\partial B} = 0, \frac{\partial S^*}{\partial B} = 0, \frac{\partial \lambda(T^*, S^*, \underline{\theta}^*)}{\partial B} = 0$. Therefore, $\frac{\partial T^*}{\partial B} \leq 0, \frac{\partial \underline{\theta}^*}{\partial B} = 0, \frac{\partial S^*}{\partial B} \geq 0, \frac{\partial \lambda(T^*, S^*, \underline{\theta}^*)}{\partial B} \geq 0$.

2. When $B < \frac{\Lambda C\theta_0}{k+1}$,

$$\frac{\partial T^*}{\partial \theta_0} = C - C\frac{k}{k+1}(\frac{(k+1)B}{\Lambda C\theta_0})^{\frac{1}{k+1}} > C - C\frac{k}{k+1} > 0,$$

$$\frac{\partial \underline{\theta}^*}{\theta_0} = 1 > 0,$$

$$\frac{\partial S^*}{\partial \theta_0} = \frac{Ck}{(k+1)^2}(\frac{B}{\Lambda})^{\frac{1}{k+1}}(\frac{C\theta_0}{k+1})^{\frac{-1}{k+1}} > 0,$$

$$\frac{\partial \lambda(T^*, S^*, \underline{\theta}^*)}{\partial \theta_0} = -\frac{k}{C\theta_0^2}\Lambda^{\frac{1}{k+1}}(\frac{(k+1)B}{C\theta_0})^{\frac{-1}{k+1}} < 0.$$

When $B \geq \frac{\Lambda C\theta_0}{k+1}$, $\frac{\partial T^*}{\partial \theta_0} = 0, \frac{\partial \underline{\theta}^*}{\partial \theta_0} = 1 > 0, \frac{\partial S^*}{\partial \theta_0} = \frac{C}{k+1} > 0, \frac{\partial \lambda(T^*, S^*, \underline{\theta}^*)}{\partial \theta_0} = 0.$

To conclude, $\frac{\partial T^*}{\partial \theta_0} \geq 0, \frac{\partial \underline{\theta}^*}{\partial \theta_0} > 0, \frac{\partial S^*}{\partial \theta_0} > 0, \frac{\partial \lambda(T^*, S^*, \underline{\theta}^*)}{\partial \theta_0} \leq 0.$

3. When $B < \frac{\Lambda C\theta_0}{k+1}$, it implies that $\frac{\partial \underline{\theta}^*}{k} = 0$. To prove the monotonicity of $T^*$ with respect to $k$, we first define

$$H(k) = (k+1)(\frac{B}{\Lambda})^{\frac{1}{k+1}}(\frac{C\theta_0}{k+1})^{\frac{k}{k+1}}.$$

We can obtain that $\ln H(k) = \ln(k+1) + \frac{1}{k+1}\ln\frac{B}{\Lambda} + \frac{k}{k+1}\ln\frac{C\theta_0}{k+1}.$

Differentiating the above equation with respect to $k$, we obtain

$$H'(k) = (k+1)(\frac{B}{\Lambda})^{\frac{1}{k+1}}(\frac{C\theta_0}{k+1})^{\frac{k}{k+1}}\frac{1+\ln\frac{\Lambda C\theta_0}{(k+1)B}}{(k+1)^2} > 0.$$

Note that $T^* = C\theta_0 - H(k)$, hence, $\frac{\partial T^*}{\partial k} = -H'(k) < 0.$

Below we investigate the relationship between $S^*$ and $k$.

We first define $G(k) = \ln\frac{\Lambda C\theta_0}{(k+1)B} - k.$

Obviously, $G'(k) < 0$.

When $B < \frac{\Lambda C \theta_0}{2e}$, $G(1) > 0$ for any $k \in (0,1]$. Thus, $\frac{\partial S^*}{\partial k} = S^* \frac{G(k)}{(k+1)^2} > 0$.

When $\frac{\Lambda C \theta_0}{2e} \leq B < \frac{\Lambda C \theta_0}{k+1}$, $G(0) > 0$ and $G(1) \leq 0$. There exists a unique root for $G(k)$ in $(0,1)$. This unique root is denoted by $k_0$. Note that $\frac{\partial S^*}{\partial k} = S^* \frac{G(k)}{(k+1)^2}$ and $G'(k) < 0$, it implies that $\frac{\partial S^*}{\partial k} > 0$ for $k < k_0$ and $\frac{\partial S^*}{\partial k} < 0$ for $k > k_0$, that is, $\frac{\partial S^*}{\partial k} \, \pounds \, 0$.

$$\frac{\partial^2 S^*}{\partial k^2} = S^* ((\frac{\ln \frac{\Lambda C \theta_0}{(k+1)B} - k}{(k+1)^2})^2 + (\frac{(k-2-2\ln \frac{\Lambda C \theta_0}{(k+1)B})}{(k+1)^3}))$$

$$= S^* ((\frac{k + \ln \frac{(k+1)B}{\Lambda C \theta_0}}{(k+1)^2})^2 - \frac{2 - k - 2\ln \frac{(k+1)B}{\Lambda C \theta_0}}{(k+1)^3})$$

$$= S^* \frac{(k + \ln \frac{(k+1)B}{\Lambda C \theta_0})(k + \ln \frac{(k+1)B}{\Lambda C \theta_0}) - (k+1)(2 - k - 2\ln \frac{(k+1)B}{\Lambda C \theta_0})}{(k+1)^4}$$

$$< S^* \frac{(k+1)(k + \ln \frac{(k+1)B}{\Lambda C \theta_0}) - (k+1)(2 - k - 2\ln \frac{(k+1)B}{\Lambda C \theta_0})}{(k+1)^4}$$

$$= S^* \frac{2k - 2 + 3\ln \frac{(k+1)B}{\Lambda C \theta_0}}{(k+1)^3} < 0$$

In a similar manner, we can prove the relationship between $\lambda(T^*, S^*, \underline{\theta}^*)$ and $k$. Hence the proof is omitted.

We prove part 4. as follows. Observe that given the first constraint of Problem (B), if $B$ is zero then $S$ is also zero. In conjunction with Proposition 1, this implies that $\lambda T - \gamma c \theta = 0$ and hence $\frac{\gamma}{\lambda} = \frac{T}{C\theta}$. Substituting into the equations for $\lambda$ and $\gamma$ of Theorem 1. Part 1. gives (a)

$\lambda = \Lambda(1 - \frac{\gamma}{\lambda})^k$, and (b) $\gamma = \Lambda[\frac{\gamma}{\lambda}(1 - \frac{\gamma}{\lambda})^k + \frac{(1 - \frac{\gamma}{\lambda})^{k+1}}{k+1}]$. Substituting (a) into (b) yields

$\gamma = \frac{\gamma}{\lambda}\lambda + \frac{(1 - \frac{\gamma}{\lambda})^{k+1}}{k+1}$, hence $\frac{(1 - \frac{\gamma}{\lambda})^{k+1}}{k+1} = 0$ which implies $\frac{\gamma}{\lambda} = 1$ and substituting this back into (a) gives us $\lambda = 0$.

The proof of Proposition 2 is complete.

**Appendix B4:   the proof of Theorem 4**

By calculating, we can get $\Lambda \int_{\frac{T}{C\theta}}^{1} ((1-\theta)pC + T)dF(p) + \Lambda \int_0^{\frac{T}{C\theta}} CpdF(p) = \Lambda \frac{C}{k+1}(1 - \theta(1 - \frac{T}{C\theta})^{k+1})$,

$r(\Lambda \int_{\frac{T}{C\theta}}^{1} (1-\theta)pCdF(p) + \Lambda \int_0^{\frac{T}{C\theta}} CpdF(p)) = r\Lambda(\frac{C}{k+1}(1 - \theta(1 - \frac{T}{C\theta})^{k+1}) - T(1 - \frac{T}{C\theta})^k)$. According to

Theorem 2, we can divided model (E) into the following three models:

$$\min_{r,T,S,\underline{\theta}} TC = \Lambda \frac{C}{k+1}(1-\underline{\theta}(1-\frac{T}{C\underline{\theta}})^{k+1})$$

$$s.t. \quad \Lambda(1-\underline{\theta}(1-\frac{T}{C\underline{\theta}})^k)S \leq r\Lambda(\frac{C}{k+1}(1-\underline{\theta}(1-\frac{T}{C\underline{\theta}})^{k+1})-T(1-\frac{T}{C\underline{\theta}})^k),$$

$$\max\{C\underline{\theta}-(k+1)S,0\} \leq T < T_1(S), \quad (E.1)$$

$$S \geq 0,$$

$$\theta_0 \leq \underline{\theta} \leq 1,$$

$$0 < r \leq r_0.$$

$$\min_{r,T,S,\underline{\theta}} TC = \Lambda \frac{C}{k+1}(1-\frac{(1-k)T+\sqrt{(1+k)^2T^2+4kT(k+1)S}}{2C}(1-\frac{2}{(1-k)+\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}})^{k+1})$$

$$s.t. \quad \Lambda(1-\frac{T}{C\underline{\theta}})^k S \leq r\Lambda(\frac{C}{k+1}(1-\frac{(1-k)T+\sqrt{(1+k)^2T^2+4kT(k+1)S}}{2C}(1-\frac{2}{(1-k)+\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}})^k),$$

$$\quad (E.2)$$

$$T_1(S) \leq T < T_2(S),$$

$$S \geq 0,$$

$$\theta_0 \leq \underline{\theta} \leq 1,$$

$$0 < r \leq r_0.$$

$$\min_{r,T,S,\underline{\theta}} TC = \Lambda \frac{C}{k+1}(1-(1-\frac{T}{C})^{k+1})$$

$$s.t. \quad \Lambda(1-\frac{T}{C})^k S \leq r\Lambda(\frac{C}{k+1}(1-(1-\frac{T}{C})^{k+1})-T(1-\frac{T}{C})^k),$$

$$T_2(S) \leq T \leq C, \quad (E.3)$$

$$S \geq 0,$$

$$\theta_0 \leq \underline{\theta} \leq 1,$$

$$0 < r \leq r_0.$$

For convenience, we denote $(T_{im}^*, S_{im}^*, \underline{\theta}_{im}^*, r_{im}^*)$ as the optimal solution of Model $(E.i)$. We solve Models $(E.1), (E.2), (E.3)$ and begin with Model $(E.1)$. Obviously, when $r$ is higher, the taxes collected from hospital is higher, which is more beneficial to villagers. So, the optimal tax rate is always equal to $r_0$. And, from $\Lambda(1-\frac{T}{C\theta})^k S = r\Lambda(\frac{C}{k+1}(1-(1-\frac{T}{C\theta})^{k+1})-T(1-\frac{T}{C\theta})^k)$, we

can get $S = \dfrac{r\frac{C}{k+1}-\Lambda(1-\frac{T}{C\theta})^k\frac{(kT+C)r}{k+1}}{\Lambda(1-\frac{T}{C\theta})^k}$, which is increasing in $T$. Next, we begin to solve model

$(E.1)$.

Obviously, $TC$ is increasing in $T$, the optimal premium to be on line $T = \max\{C\underline{\theta}-(k+1)S,0\}$.

From $\Lambda(1-\frac{T}{C\underline{\theta}})^k S \leq r\Lambda(\frac{C}{k+1}(1-(1-\frac{T}{C\underline{\theta}})^{k+1})-T(1-\frac{T}{C\underline{\theta}})^k)$, we can get $\Lambda(1-\frac{T}{C\underline{\theta}})^k(S+\frac{kT+C}{(k+1)\underline{\theta}}r_0) \leq r_0\frac{C}{k+1}$.

Then, we must discuss two cases about $S$.

Case 1: $S > \frac{C\underline{\theta}}{k+1}$, then $T^* = 0$, but $\Lambda S \leq 0$ which is conductive with $S > \frac{C\underline{\theta}}{k+1}$, this case is impossible.

Case 2: $S \leq \frac{C\underline{\theta}}{k+1}$, then $\Lambda(1-\frac{T}{C\underline{\theta}})^k(S+\frac{kT+C}{(k+1)\underline{\theta}}r_0) \leq r_0\frac{C}{k+1}$ translate into

$$\Lambda(\frac{(k+1)S}{C\underline{\theta}})^k((1-\frac{kr_0}{\underline{\theta}})S+Cr_0) \leq r_0\Lambda\frac{C}{k+1}. \quad \text{If } \underline{\theta}<\frac{1}{kr_0}, \quad \Lambda(\frac{(k+1)S}{C})^k((1-\frac{kr_0}{\underline{\theta}})S+Cr_0) \quad \text{in}$$

decreasing in $S$, Denoted by $S_1$ the root of $\Lambda(\frac{(k+1)S}{C})^k((1-\frac{kr_0}{\underline{\theta}})S+Cr_0)-r_0\Lambda\frac{C}{k+1}=0$, and

$S_1 > \frac{C\underline{\theta}}{k+1}$. Then, there has no solution. If $\underline{\theta}>\frac{1}{kr_0}$, then $\Lambda(\frac{(k+1)S}{C})^k((1-\frac{kr_0}{\underline{\theta}})S+Cr_0)$ is

increasing in $S$. And $S_1 < \frac{C\underline{\theta}}{k+1}$. So, the optimal $T$ be at the point $T=C\underline{\theta}-(k+1)S_1$. Thus, we can

get the optimal contract is

If $\frac{1}{k}<r_0<\frac{1}{k\theta_0}$, then $\underline{\theta}_{1m}^*=\frac{1}{kr_0}$, $S_{1m}^*=S_1$, $T_{1m}^*=C\frac{1}{kr_0}-(k+1)S_1$, $r_{1m}^*=r_0$.

If $r_0>\frac{1}{k\theta_0}$, then $\underline{\theta}_{1m}^*=\theta_0$, $S_{1m}^*=S_1$, $T_{1m}^*=C\theta_0-(k+1)S_1$, $r_{1m}^*=r_0$.

Next, using the same method to solve model $(E.2)$.

$$\frac{\partial TC}{\partial T}=-(1-\frac{2}{(1-k)+\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}})^k(\frac{(1-k)\sqrt{(1+k)^2T^2+4kT(k+1)S}+(1+k)^2T+2k(k+1)S}{2C\sqrt{(1+k)^2T^2+4kT(k+1)S}}\frac{-1-k+\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}}{(1-k)+\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}})$$

$$-\frac{4k(k+1)^2S}{2C((1-k)T+\sqrt{(1+k)^2T^2+4kT(k+1)S})\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}}$$

$$=(1-\frac{2}{(1-k)+\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}})^k\frac{1}{2C((1-k)T+\sqrt{(1+k)^2T^2+4kT(k+1)S})\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}}$$

$$=(1-\frac{2}{(1-k)+\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}})^k\frac{(4k(k+1)^2S-((1-k)\sqrt{(1+k)^2T^2+4kT(k+1)S}+(1+k)^2T+2k(k+1)S)(-1-k+\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}))}{2C((1-k)T+\sqrt{(1+k)^2T^2+4kT(k+1)S})\sqrt{(1+k)^2+4k(k+1)\frac{S}{T}}}$$

$$\frac{2k(k+1)}{}((5k+1)S+(1+k)T-\sqrt{(1+k)^2T^2+4k(k+1)TS})>0,$$

which means $TC$ is increasing in $T$, the optimal $T$ must be on line $T=T_1(S)$, and we know when

$T=T_1(S)$, $\frac{(1-k)T+\sqrt{(1+k)^2T^2+4kT(k+1)S}}{2C}=\underline{\theta}$. From

$\Lambda(1-\frac{T_1(S)}{C\underline{\theta}})^kS \leq r\Lambda(\frac{C}{k+1}(1-(1-\frac{T_1(S)}{C\underline{\theta}})^{k+1})-T_1(S)(1-\frac{T_1(S)}{C\underline{\theta}})^k)$, we can get

$\Lambda(1-\frac{T_1(S)}{C\underline{\theta}})^k(S+\frac{Cr_0}{k+1}+\frac{((k+1)\underline{\theta}-1)T_1(S)}{(k+1)\underline{\theta}}r_0) \leq r_0\frac{C}{k+1}$. $\Lambda(1-\frac{T_1(S)}{C\underline{\theta}})^k(S+\frac{Cr_0}{k+1}+\frac{((k+1)\underline{\theta}-1)T_1(S)}{(k+1)\underline{\theta}}r_0)$ is

increasing in $S$, denoted by $S_2$ the root of $\Lambda(1-\frac{T_1(S)}{C\underline{\theta}})^k(S+\frac{Cr_0}{k+1}+\frac{((k+1)\underline{\theta}-1)T_1(S)}{(k+1)\underline{\theta}}r_0)-r_0\frac{C}{k+1}=0$.

Using the same method as showed in Model $(E.1)$, we can get the optimal contract is
$T_{2m}^*=T_1(S_2), S_{2m}^*=S_2, \underline{\theta}_{2m}^*=\theta_0, r_{2m}^*=r_0$.

The solving process of Model $(E.3)$ is similar to model $(E.1)$, here, we omit the detailed

procedure and present the optimal contract is $T_{3m}^*=T_1(S_3), S_{3m}^*=S_3, \underline{\theta}_{3m}^*\in[\theta_0,1], r_{3m}^*=r_0$, where

$S_3$ is the root of $\Lambda(1-\frac{T_2(S)}{C})^k(S+\frac{k(T_2(S)+C)}{k+1}r_0)-r_0\frac{C}{k+1}=0$.

Comparing the optimal contract of model $(E.1)$, model $(E.2)$, model $(E.3)$, it is obvious that

$$\Lambda \frac{C}{k+1}(1-\underline{\theta}_{1m}^{*}(1-\frac{T_{1m}^{*}}{C\underline{\theta}_{1m}^{*}})^{k+1}) < \Lambda \frac{C}{k+1}(1-\underline{\theta}_{1m}^{*}(1-\frac{T_{2m}^{*}}{C\underline{\theta}_{1m}^{*}})^{k+1}) = \Lambda \frac{C}{k+1}(1-\underline{\theta}_{2m}^{*}(1-\frac{T_{2m}^{*}}{C\underline{\theta}_{2m}^{*}})^{k+1})$$

$$< \Lambda \frac{C}{k+1}(1-\underline{\theta}_{2m}^{*}(1-\frac{T_{3m}^{*}}{C\underline{\theta}_{2m}^{*}})^{k+1}) < \Lambda \frac{C}{k+1}(1-(1-\frac{T_{3m}^{*}}{C})^{k+1})$$

.

So, the optimal contract is as follows:

If $r_0 < \dfrac{1}{k\theta_0}$, then $\underline{\theta}^{*} = \dfrac{1}{kr_0}, S^{*} = S_1, T^{*} = C\dfrac{1}{kr_0} - (k+1)S_1, r^{*} = r_0$.

If $r_0 > \dfrac{1}{k\theta_0}$, then $\underline{\theta}^{*} = \theta_0, S^{*} = S_1, T^{*} = C\theta_0 - (k+1)S_1, r^{*} = r_0$.