TEAM MEMBERS: Boxian WANG Zhiming JING

SCHOOL： Nanjing Foreign Language School

PROVINCE： Jiangsu

COUNTRY： China

Advisor： Xuebin Lv

SUBJECT： Mathematical Models for Pricing and Reimbursement in DRGs of Chinese Health Care Reform

Table of Contents

Abstract………………………………………………………………………..……3
1. Background and problem restatement…………………………………….…….4
2. Assumptions……………………………………………………………………5
3. Medical costs of cerebral infraction patients and pricing methods……………….5
3.1 Estimation of the distribution of CI medical costs based on Gaussian Mixture Model…..5
3.2 Pricing methods in DRGs system for CI patients……………………………….8
3.2.1 Pricing method I………………………………………………………….9
3.2.2 Pricing method II………………………………………………...…….9
3.2.3 Comparison of the two pricing methods…………………………..…….11
3.3 Factor analysis of costs of CI patients based on Hierarchical Multinomial Logistic Regression Model……………………………….………………..…….11
3.4 Summary……………………………………………………………...…….14
4. The model for government subsidies based on DRGs payment system………....15
4.1 A brief introduction to Lundberg-Cramer Ruin theory…………………………....15
4.2 The model for government subsidies and individual patients……………………..…16
4.3 The model for government subsidies and hospital income………………………..…22
4.4 Combined consideration of hospital and patient benefits……………………...….24
4.5 Summary……………………………………………………………………….26
5. Conclusions and suggestions……………………………………..……………27
6. Model evaluation…………………………………………………………….28
7. References…………………………………………………..……………….29
8. Appendix……………………………………………………………………29

Abstract

The new health care reform in China aims to replace the old payment system that is based on service items with a new Diagnosis Related Groups (DRGs) payment system. The DRGs payment system categorizes different diseases and set a single price for each group of disease. This payment system is expected to overcome flaws in the old one such as hospital overcharging patients with unnecessary items, and increase the allocation efficiency of medical resources. In this paper we aim to answer two fundamental questions while establishing DRGs payment system: how to set prices, and how to coordinate prices and government subsidies to optimize the welfare of both patients and hospitals.

To answer the first question, we first take the disease Cerebral Infraction (CI) as an example, and estimate the distribution of 306 CI patients' medical costs with Gaussian Mixture Model. Then, based on the estimated distribution, we propose two pricing methods and compare them in terms of advantages and disadvantages, offering suggestions to hospitals. Last, we utilize Hierarchical Multinomial Logistic Regression to analyze factors that potentially influence the medical cost.

To answer the second question, we construct a model for individual patients based on Lundberg-Cramer Ruin Theory, and discuss the ruin probability caused by high medical fees for patients with different income levels. We also construct a second model to illustrate the relation among medical prices, government subsidies and hospital income. Finally, we combine the welfare of individual patients and hospital to obtain a utility function, using which we could optimize medical prices and government subsidies. In the end, based on the analysis and optimization, we offer our suggestions.

Keywords:
Diagnosis Related Groups；Gaussian Mixture Model；Hierarchical Multinomial Logistic Regression Model；Lundberg-Cramer Ruin Model；Monte Carlo Simulation

1. Background and problem restatement

China has established a basic health care system and has been extending the coverage of basic health insurance. Until 2015, over 1.3 billion people had been covered by health insurance, raising the coverage rate to 96.5 percent. Along with the constant growth of coverage and rising level of insurance, the health insurance system also faces emerging problems such as unreasonable growth of medical expenses and increased pressure on maintaining the insurance funds. In 2015, national health care expenses had gone up to 4.05 trillion Chinese yuan, and medical expenses per capita had reached 2952 Chinese yuan [1]. The unreasonable growth of healthcare expenses offset part of government spending in healthcare, and increased the burdens of patients. Hence, determining the adequate health care payment system is essential to guarantee the effective usage of healthcare resources, controlling the growth of medical costs, and lessening the economical burdens of patients.

The existing payment system in China is based on separate items during the medical treatment. Each item is separately priced and patients are charged when they are provided with the service. Although this payment mode is easy to operate, and the relation between patients, healthcare providers and health insurance funds is relatively simple, the price of each item could not be accurately determined. Furthermore, this payment mode stimulates the hospital to include many unnecessary items during the treatment to increase income, which ultimately leads to uncontrolled growth of healthcare costs.

To amend the flaws of the old payment mode based on separate items, the new healthcare reform in China aims to introduce Diagnosis Related Groups (DRGs). This new payment mode categorizes patients into different groups (sometimes more than five hundreds of them) according to their respective diagnosis. Then different patients would be charged by different standards, mostly preset payments, according to their related DRGs. In the meanwhile, hospitals would be subsidized by the authority with respect to their usage of medical resources for patients in different DRGs. In this way, payment mode based on DRGs could maximize the allocation efficiency of medical resources while significantly control the unreasonable growth of healthcare costs. China has begun to experiment this new payment mode in several provinces.

However, during the implementation of DRGs, several questions need immediate attention. For instance, we need a quantitative method to set the price for each group of patients, considering factors such as reasonable healthcare expenses, incomes of hospitals, and means of government subsidies. In this paper, we aim to study hospital pricings and government subsidies of DRGs. Specifically, we address the following two questions:

1. How to set prices in the DRGs payment mode?
2. How to coordinate prices and governmental subsidies to maximize the welfare

of both hospitals and patients?

2. Assumptions

1. Due to limited data, we select Cerebral Infractions (CI) patients as an example of a certain DRG, and utilize the data of total medical costs of 306 CI patients discharged from a hospital in 2017 as the training set.
2. For simplicity, assume the incidence rate of disease is constant.
3. Only the direct expenses and costs during the treatment are accounted for. We do not consider other indirect income or expenses of both hospitals and patients.
4. The government subsidies go either to the hospitals or to the patients.

3. Medical costs of cerebral infraction patients and pricing methods

In this section we take CI patients as an example, and utilize the data of total medical costs of 306 CI patients in 2017 as the training set. We first estimate the distribution of medical costs of CI. Then, based on the distribution model, we propose two pricing methods for hospitals, and present a comparison between the two methods. Last, we apply Hierarchical Multinomial Logistic Regression to analyze factors that potentially lead to high medical costs.

3.1 Estimation of the distribution of CI medical costs based on Gaussian Mixture Model

The medical costs in CI treatment could be influenced or be related to many factors, such as surgeries, medicine, and examinations during the treatment. In this part we estimate the distribution of the medical costs in the training set.

The training set includes the total medical costs of 306 CI patients in a public hospital in 2017. The costs vary from 1227 Chinese yuan to 100271 Chinese yuan, with an average cost of 19119.5 yuan. A histogram of the data is shown below:

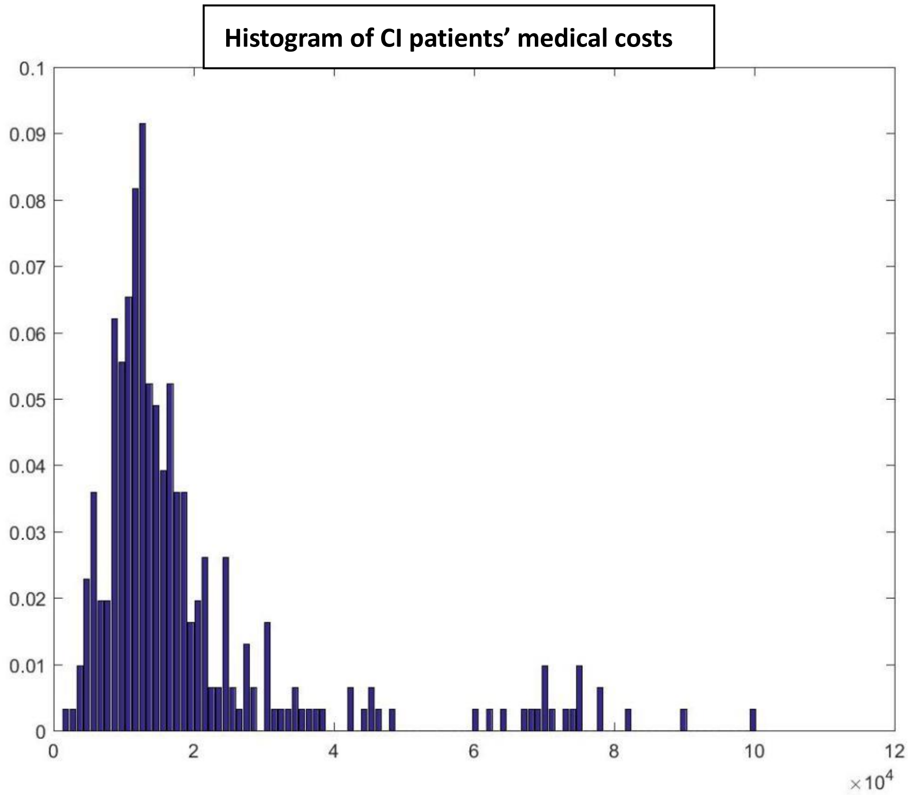

Figure 3.1: Histogram of the total costs of 306 CI patients

Let the cost of a CI patient be a random variable X. Observing that the histogram includes multiple peak values, we choose to fit its distribution with Gaussian Mixture Model (GMM). Thus the probability density function of X is:

$$p(x) = \sum_{j=1}^{g} \pi_j \varphi(x; \mu_j, \sigma_j) \tag{3.1}$$

where g is the order of the model, $\pi_j \geq 0$ are mixing coefficients, $\sum_{j=1}^{g} \pi_j = 1$,

$\varphi(x; \mu_j, \sigma_j)$ is a Gaussian distribution with mean μ_j and variance $\sigma_j^{\ 2}$:

$$\varphi(x; \mu_j, \sigma_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^{\ 2}}}, -\infty < x < +\infty \tag{3.2}$$

The expectation and variance of a Gaussian Mixture Model are:

$$E(X)=\int_{-\infty}^{+\infty} x\sum_{j=1}^{g}\pi_j\varphi(x;\mu_j,\sigma_j)dx=\sum_{j=1}^{g}\pi_j\mu_j,$$

$$E(X^2)=\int_{-\infty}^{+\infty} x^2\sum_{j=1}^{g}\pi_j\varphi(x;\mu_j,\sigma_j)dx=\sum_{j=1}^{g}\pi_j\int_{-\infty}^{+\infty} x^2\varphi(x;\mu_j,\sigma_j)dx=\sum_{j=1}^{g}\pi_j(\mu_j^{\ 2}+\sigma_j^{\ 2}),$$

$$Var(X)=E(X^2)-(E(X))^2=\sum_{j=1}^{g}\pi_j(\mu_j^{\ 2}+\sigma_j^{\ 2})-(\sum_{j=1}^{g}\pi_j\mu_j)^2. \tag{3.3}$$

For a GMM with given order g, we need to estimate the following parameters: mixing coefficients $\pi=\left\{\pi_1,\cdots,\pi_g\right\}$ and the parameter (μ_j,σ_j) for each Gaussian distribution. We use the Expectation Maximization Algorithm to estimate the parameters, which maximizes the likelihood function [2, 3]:

$$L(\pi_1,\cdots,\pi_g,\mu_1,\cdots,\mu_g,\sigma_1,\cdots,\sigma_g)=\prod_{i=1}^{n}\left\{\sum_{j=1}^{g}\pi_j\varphi(x_i;\mu_j,\sigma_j)\right\} \tag{3.4}$$

We could also apply Akaike Information Criterion (AIC) to determine the order g:

$$AIC=-2\log(L(\Psi))+2k_d \tag{3.5}$$

where $L(\Psi)$ is the likelihood function of the model, k_d is the number of unknown parameters. Let $g=1,2,3,4$, and apply EM algorithm to estimate parameters for each g. Then compare the AIC of each model, and select the model with the least AIC.

Table 3.1

Number of Distributions	1	2	3
AIC	6820	6471	6446

We select the model with g=3. When g=3, we estimate the parameters to be:

$$\pi_1=0.7185,\pi_2=0.2151,\pi_3=0.0664,$$
$$\mu_1=12264.8,\mu_2=25354.94,\mu_3=73117.98,$$
$$\sigma_1=10046,\sigma_2=9658,\sigma_3=4256.$$

The estimated expectation and variance of X are:

$$EX = 19121$$
$$Var(X) = 3.2951 \times 10^{8}$$

Using GMM with order 3 to fit the distribution of medical costs of CI patients, we obtain the graph of the probability density function:

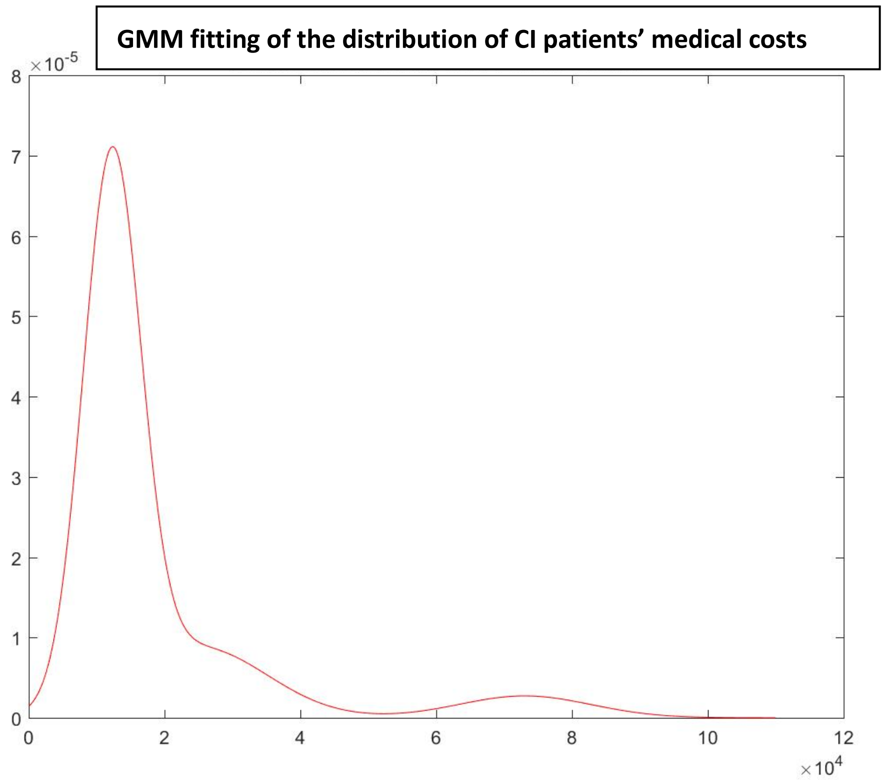

Figure 3.2: probability density function of cost of CI patients

3.2 Pricing methods in DRGs system for CI patients

Without government subsidies, the hospital must set the price for each group of patients such that the risk of treatment cost exceeding the set payment (the amount patients pay) is lowered. Based on the distribution model established in section 3.1, we propose two pricing methods and briefly discuss their advantages and disadvantages.

3.2.1 Pricing method I

Let the actual medical cost of a single patient be random variable X. We set the price at K, so that the probability of X exceeding K is less than p:

$$P(X \geq K) \leq p$$

To determine the lower limit of K, we utilize Monte-Carlo method. First we generate 10000 random numbers based on the PDF of GMM we established in section 3.1. Next, we arrange the generated numbers in decreasing order. For given p, we select the $10000p$-th to be the estimated value of K. Repeating the aforementioned procedures and take the average to obtain a more accurate lower limit of K. We let p=0.2, 0.15, 0.1, 0.05 and 0.01. The result is shown as

Table 3.2

p	0.2	0.15	0.1	0.05	0.01
K	23082	27978	35611	67592	82962

3.2.2 Pricing method II

Consider N patients in a period of time, with actual costs denoted as random variables $X_1,\ X_2,\cdots,\ X_N$, which are independent and identically distributed. Let the price be set at K, so that the probability of the total amount of medical costs exceeding the total payment received is less than p:

$$P\left\{\sum_{i=1}^{N} X_i \geq NK\right\} \leq p \qquad (3.6)$$

As X_1, $X_2, \cdots$, X_N are independent and identically distributed, their expectation and variance both exist. If the number of patients, N, is sufficiently large, we could apply central limit theorem. It follows that $\frac{\sum_{i=1}^{N} X_i - NE(X)}{\sqrt{N \operatorname{var}(X)}}$ approximately obeys standard normal distribution. Thus

$$P\left\{\sum_{i=1}^{N} X_i \geq NK\right\} \approx P\left\{\frac{\sum_{i=1}^{N} X_i - NE(X)}{\sqrt{N \operatorname{var}(X)}} \geq \frac{\sqrt{N}(K - E(X))}{\sqrt{\operatorname{var}(X)}}\right\} \leq p \qquad (3.7)$$

Solving for K yields

$$K \geq E(X) + \sqrt{\frac{\operatorname{var}(X)}{N}}\Phi^{-1}(1-p) \qquad (3.8)$$

Where Φ^{-1} stands for the inverse function of $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$, the cumulative distribution function of standard normal distribution.

With the help of MATLAB, we calculate the lower limit of K for different N and p as the following:

Table 3.3

N \ P	0.2	0.15	0.1	0.05	0.01
50	21282	21782	22411	23344	25093
500	19804	19963	20162	204546	21010
5000	19337	19387	19450	19543	19718
10000	19274	19309	19354	19420	19543

It could be seen that as N increases, K decreases. As p decreases, K increases.

3.2.3 Comparison of the two pricing methods

The two pricing methods are both based on the GMM established in section 3.1 that describes the distribution of actual medical costs of patients.

For the first method, we only consider the payment and cost of a single patient, and control the risk. It is thus safer for the hospital, as this method take patients with extremely high costs into account. However, as the distribution of medical cost is skewed to the left, meaning that most patients have a lower cost than average cost, this pricing method harms the welfare of low-cost patients, who take up the majority of CI patients.

For the second method, however, we consider the case of N patients, and their total amount of costs and payments. This method significantly lowers the set price as the number of patient increases, and seems to be friendlier to the welfare of majority of patients. However, the hospital needs to bear greater risks due to some high-cost patients. Additionally, when N is small, central limit theorem is no longer accurate, and so is our second method.

3.3 Factor analysis of costs of CI patients based on Hierarchical Multinomial Logistic Regression Model

In this part we apply Hierarchical Multinomial Logistic Regression Model to analyze factors affecting the costs of CI patients, based on the data of 306 CI patients. We hope to estimate the cost range based on the information of the patient, and thus provide suggestions to public hospitals.

We first divide patients into three groups according to their costs: low-cost (0~15000 Chinese yuan), mid-cost (15001~45000 yuan), and high-cost (more than 45000 yuan). Then, we consider five factors: gender, age, surgery, length of stay, and infection. We aim to study the influence of these five factors on the medical costs of CI patients.

Thus let the response variable be the cost group, and the five independent variables are $Gender$, Age, $Surgery$, $Length$ and $Infection$. $Gender=0$ for female, $Gender=1$ for male. $Surgery=1$ for surgery performed, 0 for no surgery. $Infection=1$ for infection having occurred, 0 for no infection. We construct the following regression model:

$$\ln\left(\frac{P(X\le 15000)}{P(X>15000)}\right)=\beta_{10}+\beta_{11}Age+\beta_{12}Gender+\beta_{13}Surgery+\beta_{14}Length+\beta_{15}Infection$$

$$\ln\left(\frac{P(15001\le X\le 45000)}{P(X>45000)}\right)=\beta_{20}+\beta_{21}Age+\beta_{22}Gender+\beta_{23}Surgery+\beta_{24}Length+\beta_{25}Infection \tag{3.11}$$

where $\beta_{ij}, i=1,2, j=0,1,2,\cdots,5,$ stand for the intercepts and regression coefficients in the model. Based on our patient data, we utilize MATLAB to conduct the regression. We obtained the estimation of intercepts and regression coefficients as:

$$B=\begin{pmatrix} \beta_{10} & \beta_{20} \\ \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \beta_{13} & \beta_{23} \\ \beta_{14} & \beta_{24} \\ \beta_{15} & \beta_{25} \end{pmatrix}=\begin{pmatrix} 2.9452 & 1.7298 \\ -0.0028 & 0.0279 \\ -0.5437 & 0.0109 \\ -102.27 & -3.8236 \\ -0.1440 & -0.0786 \\ -2.2708 & -1.5902 \end{pmatrix},$$

After running the significance test, we further obtain the P-value

$$p=\begin{pmatrix} 0.0014 & 0.3499 \\ 0.8095 & 0.2151 \\ 0.0780 & 0.9862 \\ 1.0000 & 0.0028 \\ 0.0000 & 0.0010 \\ 0.0383 & 0.0638 \end{pmatrix}$$

For P-values 0.0000, 0.0010 and 0.0383, 0.0638, we conclude that length of stay and infection significantly affect the medical costs of CI patients. P-values 0.8095, 0.2151 tell that age of patients does not significantly affect medical costs. In fact, most CI patients are above fifty years old. P-values 0.0780, 0.9862 show that gender does affect the ratio between low-cost group and mid-to-high-cost group. The coefficient -0.5437 further tells that a patient is $e^{-0.5437}$ times more likely to fall into a mid-cost or high-cost group if the gender changes from female to male. However, gender does not significantly affect the ratio between mid-cost and high-cost groups. P-value 0.0028 shows that surgeries significantly affect the ratio between mid-cost and high-cost groups.

Taking a step further, we eliminate the insignificant variable Age, and establish a new Logistic regression:

$$\ln\left(\frac{P(X \le 15000)}{P(X > 15000)}\right) = \beta_{10} + \beta_{12}Gender + \beta_{13}\text{surgery} + \beta_{14}length + \beta_{15}\text{Infection}$$

$$\ln\left(\frac{P(15001 \le X \le 45000)}{P(X > 45000)}\right) = \beta_{20} + \beta_{22}Gender + \beta_{23}\text{surgery} + \beta_{24}length + \beta_{25}\text{Infection} \tag{3.12}$$

Conduct the regression by calculating the estimation of coefficients and the P-value:

$$B = \begin{pmatrix} \beta_{10} & \beta_{20} \\ \beta_{12} & \beta_{22} \\ \beta_{13} & \beta_{23} \\ \beta_{14} & \beta_{24} \\ \beta_{15} & \beta_{25} \end{pmatrix} = \begin{pmatrix} 2.7447 & 3.8314 \\ -0.5385 & -0.1907 \\ -102.27 & -3.8441 \\ -0.1445 & -0.0758 \\ -2.2706 & -1.4775 \end{pmatrix}, \qquad p = \begin{pmatrix} 0.0000 & 0.0000 \\ 0.0800 & 0.7527 \\ 1.0000 & 0.0023 \\ 0.0000 & 0.0011 \\ 0.0383 & 0.0856 \end{pmatrix}$$

The regression shows that length of stay and infections significantly affect the cost groups, and gender has a significant influence on the ratio of low-cost groups. Thus male patients have a higher probability of falling into the mid-cost or high-cost category. Surgeries have a more significant effect when distinguishing mid-cost and high-cost patients. Specifically, the ratio between mid-cost and high-cost is changed by $e^{-3.8441}$ if surgeries are performed.

3.4 Summary

We first determine the distribution of medical costs of CI. Then, based on the distribution model, we propose two pricing methods for hospitals, and present a comparison between the two methods. Last, we apply Hierarchical Multinomial Logistic Regression to analyze five factors' respective influence on the medical cost, providing a theoretical approach for hospitals to predict the patient's cost. Again, due to limited data, we could only examine the instance of cerebral infarction patients. However, the same methods could be utilized when analyzing other diseases.

4. The model for government subsidies based on DRGs payment system

In this section we take government subsidies into consideration. As the role of government spending is crucial to the existing healthcare system, we must study the proper way of subsidizing hospitals and patients under the DRGs payment system. We would first establish a model for individual patients in the light of Lundberg-Cramer Ruin Theory. Then we would discuss the influence of government subsidies on hospital income. Finally, we combine the welfare for patients and the hospital, and define a utility function that would describe the combined welfare. We would then maximize the utility function to determine the optimal government subsidies and hospital pricing.

4.1 A brief introduction to Lundberg-Cramer Ruin theory

The Lundberg-Cramer Ruin theory [2, 3] is a common model in stochastic process. It describes the following situation:

An insurance company starts with some amount of capital and its income in unit time is constant. However, the insurance company has to pay for the insurance claims from time to time. We are interested in the probability that the insurance company ultimately goes bankrupt. We express the situation above in mathematical language:

u : The initial capital

c : The income of insurance company in unit time

X_k : The payment to the n-th insurance claim

$N(t)$:The total amount of insurance claims at time t, a Poisson Process with parameter λ

Thus the balance of the insurance company at time t $U(t)$:

$$U(t) = u + ct - \sum_{k=1}^{N(t)} X_k \qquad (4.1)$$

For this model, we define a function $\Psi(u)$ that describes the probability that the insurance company ultimately goes bankrupt.

Let $T = \inf\{t : U(t) < 0\}$ be the first time the insurance company goes bankrupt, then

$$\Psi(u) = P\{T < 0 | U(0) = u\}, u \geq 0$$

which is the ruin probability.

For the model above, we have the following theorems:

1. $\Psi(u) \leq e^{-Ru}, \forall u \geq 0$ (4.2)
2. There exists a positive constant C, such that $\Psi(u) \sim Ce^{-Ru}, u \to \infty$ (4.3)

where R is the adjustment coefficients, and satisfies

$$\frac{\lambda}{A}\int_0^{+\infty} e^{Rx}(1 - F(x))dx = 1 \qquad (4.4)$$

For proofs of the two theorems above, see reference [4].

4.2 The model for government subsidies and individual patients

Based on the model of insurance company presented in section 4.1, we construct a similar model for patients encountering serious diseases. For a single kind of disease, we propose the following:

a : The initial capital in the healthcare fund

I : The patient's income in unit time, which we assume is constant

$N(t)$: The number of times in hospital till time t, a Poisson Process with parameter λ

K : The set price for each treatment, which we assume is constant for a single disease under DRGs payment system

ρ : The government subsidiary coefficient, i.e. the fraction of medical fees paid by the government

Thus, the balance of individual healthcare fund $V(t)$ at time t is [5]

$$V(t) = a + It - (1-\rho)\sum_{i=1}^{N(t)} K \tag{4.5}$$

Similarly, let $T = \inf\{t : V(t) < 0\}$, we consider the function $f(K,\rho)$ that describes the probability of failing to pay the medical fees in unit time:

$$f(K,\rho) = P\{T \le 1\}, (K \ge 0, 0 \le \rho \le 1) \tag{4.6}$$

In other words, we use $f(K,\rho)$ to represent the welfare of individual patients as it stands for the probability that the patient goes 'bankrupt' because of serious illness in unit time. We assume that unit time is one year.

We utilize MATLAB to simulate this process [6], and in this way we are able to calculate $f(K,\rho)$ with given (K,ρ). We apply the following algorithm:

1. Initialize the value of a, I, $N(t)$, K, ρ

2. Divide the unit time equally into many small sections $\Delta t_1, \ldots, \Delta t_n$, thus the increment of Poisson Process $N(t)$ is independent and obeys Poisson distribution with parameter $\lambda \Delta t$.

3. We simulate a sample path for $N(t)$ according to the division.

4. Using the simulated sample path, and the initial values of other parameters, we could now calculate $V(t)$ for $\Delta t_1, \ldots, \Delta t_n$. And then we could examine whether $V(t) < 0$ occurs in unit time.

5. repeat step 3 and 4, and use the frequency of ruin as the approximation of ruin probability in unit time.

The average resident hospitalization rate in the province where the CI patient data is taken from is 0.167, so we let λ=0.167. Additionally, we consider the income level of residents in the province. We consider the quintiles of disposable income in the province.

Table 4.1

Quintiles	Low	Medium-low	Medium	Medium-high	High
Income/yuan	5221.2	11894.0	19320.1	29437.6	54543.5

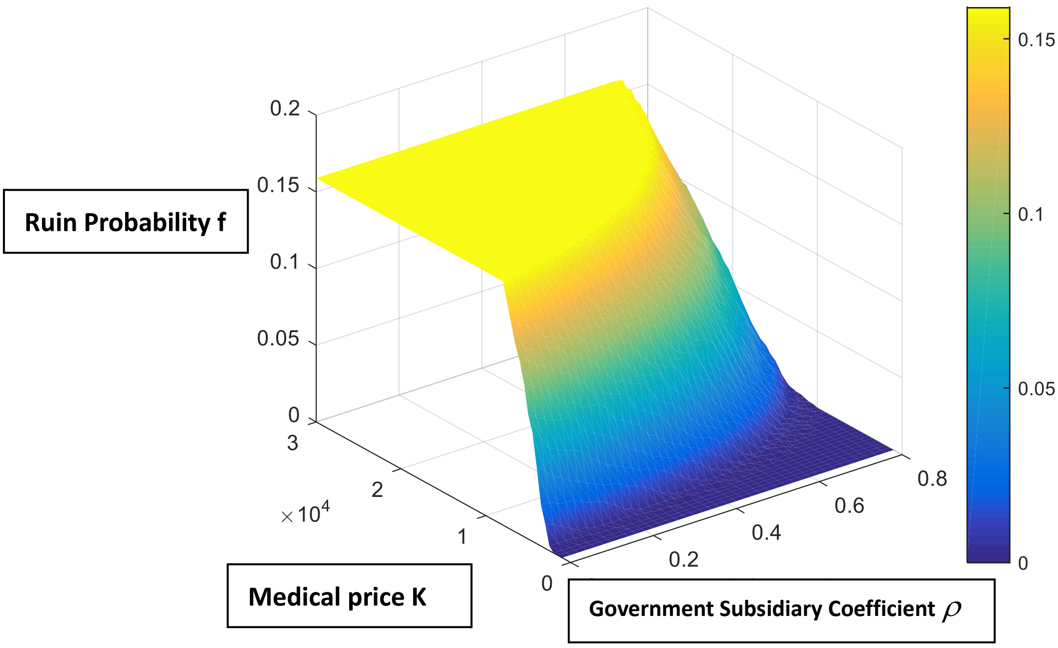

Figure 4.1: Low income, $I = 5221.2, a = 0.5I$

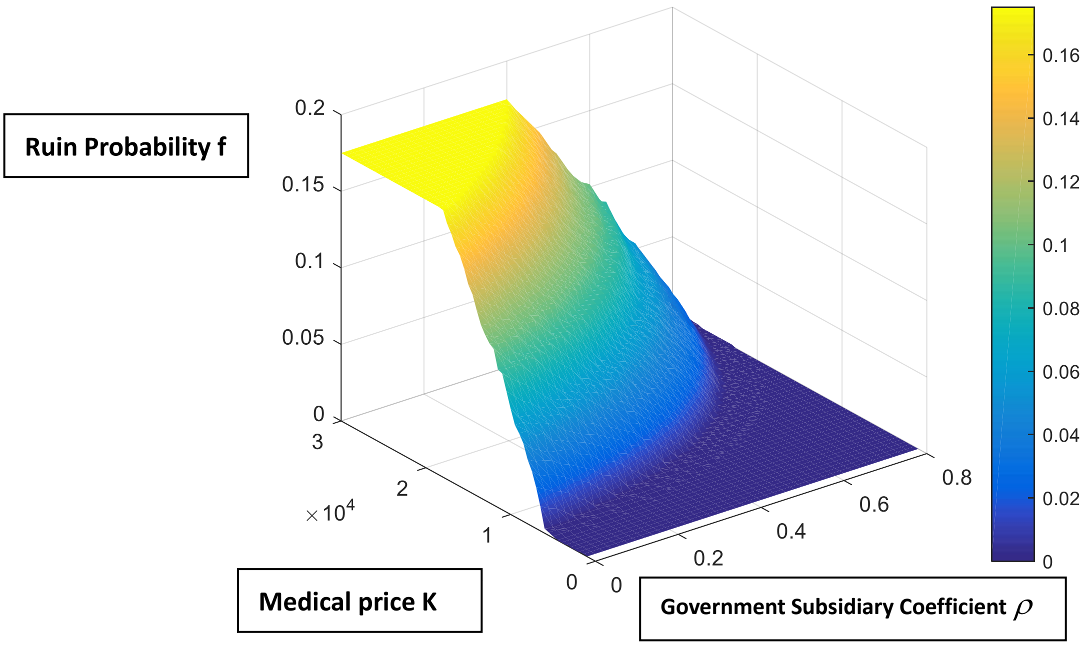

Figure 4.2: Medium-low income, $I = 11894.0, a = 0.5I$

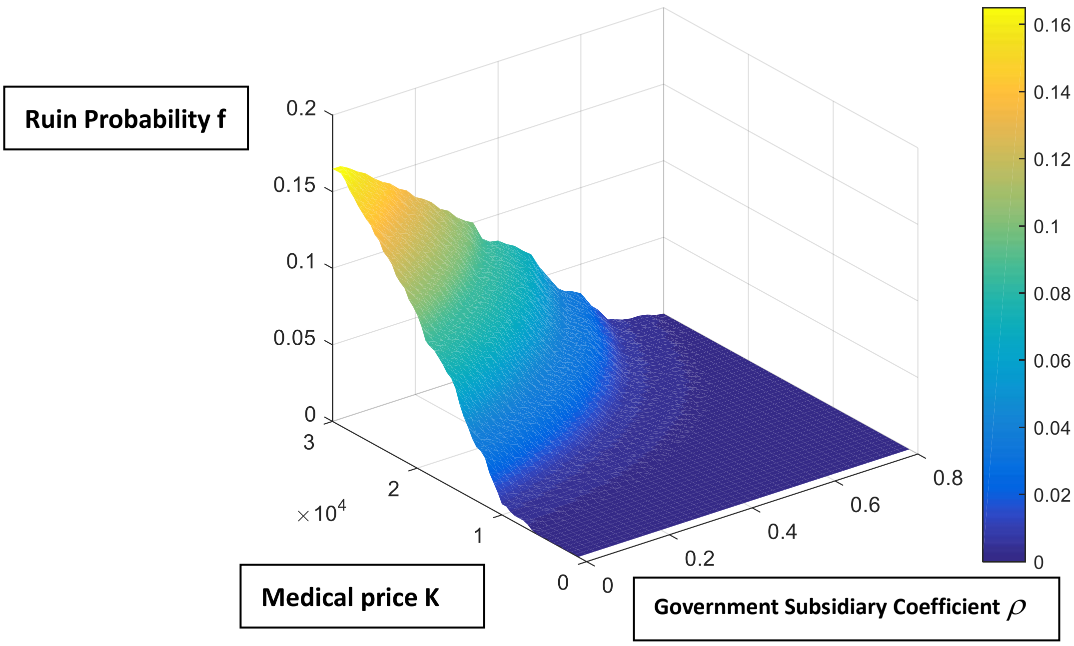

Figure 4.3: Medium income, $I = 19320, a = 0.5I$

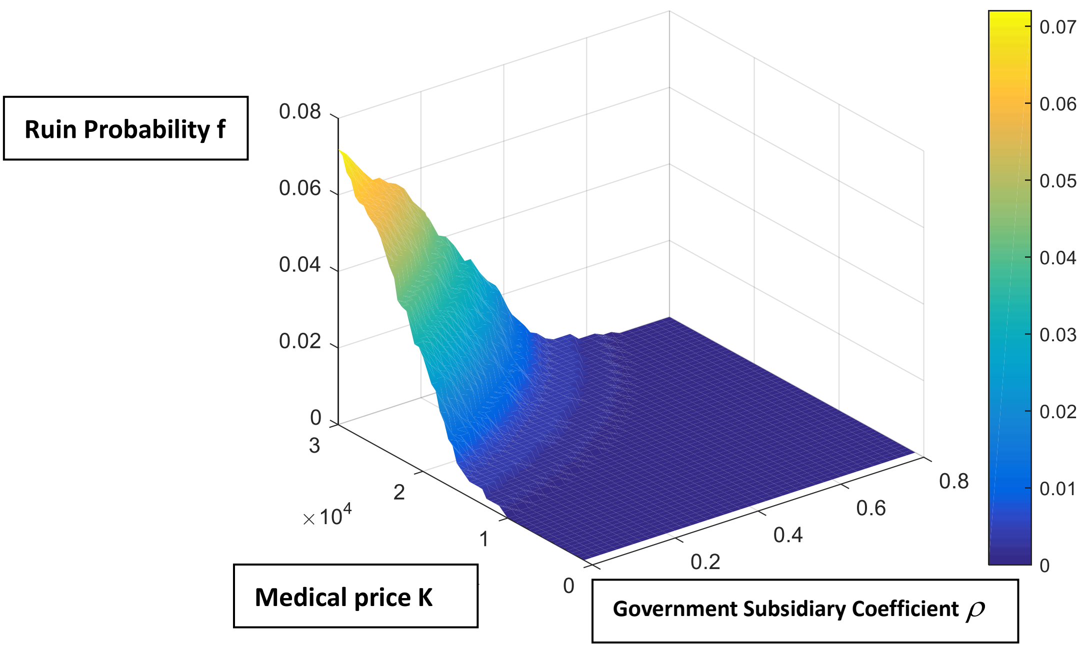

Figure 4.4: Medium-high income, $I = 29437, a = 0.5I$

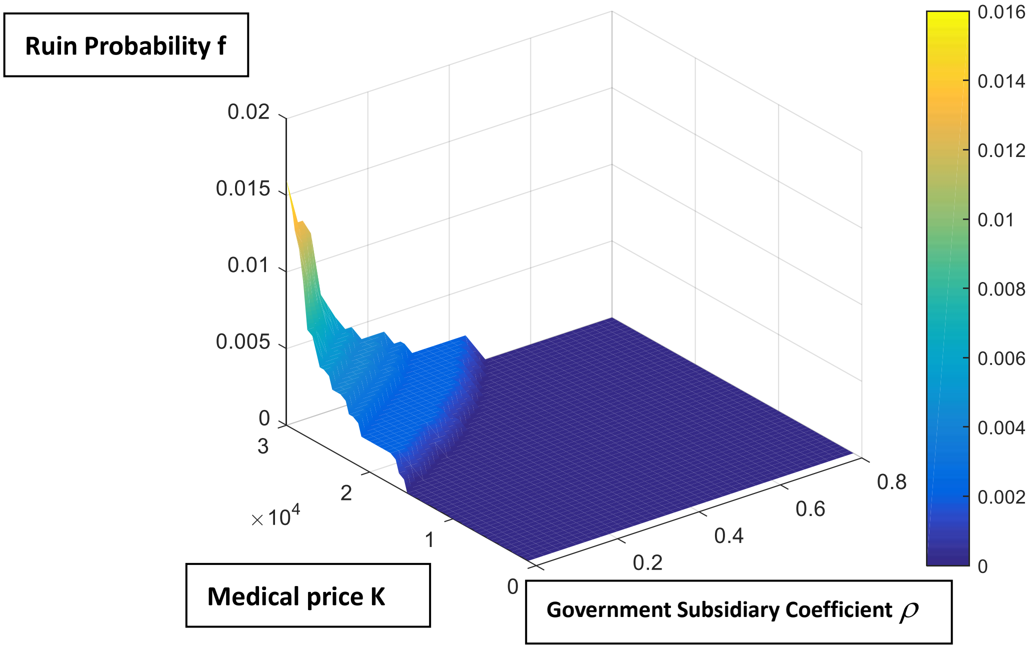

Figure 4.5: High income, $I = 54543, a = 0.5I$

From the graphs of $f(K,\rho)$ of different income groups, we could make a few conclusions.

First, as K increases, $f(K,\rho)$ increases. As ρ increases, $f(K,\rho)$ decreases, i.e. $\frac{\partial f}{\partial K} > 0, \frac{\partial f}{\partial \rho} < 0$. Thus, to increase patient's welfare, the set price K must be lowered while the subsidiary coefficient ρ must be increased. Second, the income level affects the ruin probability $f(K,\rho)$ significantly. Low income patients would be ten to twenty times more fragile when encountering serious illness. Compared to medium income patients, they are still more susceptible to 'bankruptcy' for large range of (K,ρ). Thus to improve the situation, besides raising the income level of residents, the government must also seek to further expand the coverage of health insurance to increase healthcare fund for low income patients.

4.3 The model for government subsidies and hospital income

In this part we discuss the relation between government subsidies and hospital income. Public hospitals are non-for-profit, thus they would need government funding, aside from medical fees, to support its operation and development. We would need to determine the proper amount of government subsidies that would maintain hospital income in reasonable range. We first construct a theoretical model.

In unit time, assume that N patients received medical treatment, each with actual medical cost of $X_1, X_2, \cdots X_N$. Based on our conclusion in section 3.1, $X_1, X_2, ..., X_N$ are independent and identically distributed, and they obey the GMM model

$$p(x) = \sum_{j=1}^{3} \pi_j \varphi(x; \mu_j, \sigma_j),$$

with parameters

$\pi_1 = 0.7185, \pi_2 = 0.2151, \pi_3 = 0.0664,$
$\mu_1 = 12264.8, \mu_2 = 25354.94, \mu_3 = 73117.98,$.
$\sigma_1 = 10046, \sigma_2 = 9658, \sigma_3 = 4256.$

According to DRGs payment system, the price each patient pays is constant. Assume the price is set at K. We further assume that the total government subsidies for healthcare are directly proportional to the number of patients in the region. For each patient, the government provides an additional funding of k. We also consider the government subsidiary coefficient ρ that we introduced in section 4.2. Because the government subsides go either to the patient or to the hospital. With ρK subsidized to each patient, the hospital receives a subsidy of $(k-\rho K)$ per patient. We also include a coefficient m to describe the ratio between the price that patients pay (in our data set, without DRGs) and the real medical cost of hospital.

With the assumptions and variables above, we obtain the balance of hospital in unit time

$$NK + N(k-\rho K) - m\sum_{i=1}^{N} X_i \tag{4.7}$$

Take the expectation of the expression, we obtain:

$$\begin{aligned}
&E[N(K+(k-\rho K))] - mE[\sum_{i=1}^{N} X_i] \\
&= [(1-\rho)K + k]EN - m\sum_{n\in R^+} E[\sum_{i=1}^{N} X_i \mid N=n]P\{N=n\} \\
&= [(1-\rho)K + k]EN - m\sum_{n\in R^+} nEX\,P\{N=n\} \\
&= [(1-\rho)K + k]EN - mEXEN \\
&= [(1-\rho)K + k - mEX]EN
\end{aligned} \tag{4.8}$$

We let a function $g(K,\rho)$ describe the expected surplus rate of a hospital

$$\begin{aligned}
g(K,\rho) &= \frac{EN[(1-\rho)K + k - mEX]}{mENEX} \\
&= \frac{(1-\rho)K + k - mEX}{mEX}
\end{aligned} \tag{4.9}$$

Then if $g(K,\rho) \le 0$, the hospital is suffering economical losses and requires more proportion of total government subsidies. $g(K,\rho)$ must maintain in a certain range to ensure regular operation of the hospital.

We estimate $k \approx 5464$ by the ratio between the total government spending in healthcare province-wide and the total number of hospitalized patients in the province in 2016. Then we estimate that $m \approx 0.928$ by the ratio between the total medical income and the expenditure of public hospitals in the province.

We graph $g(K,\rho)$ as the following:

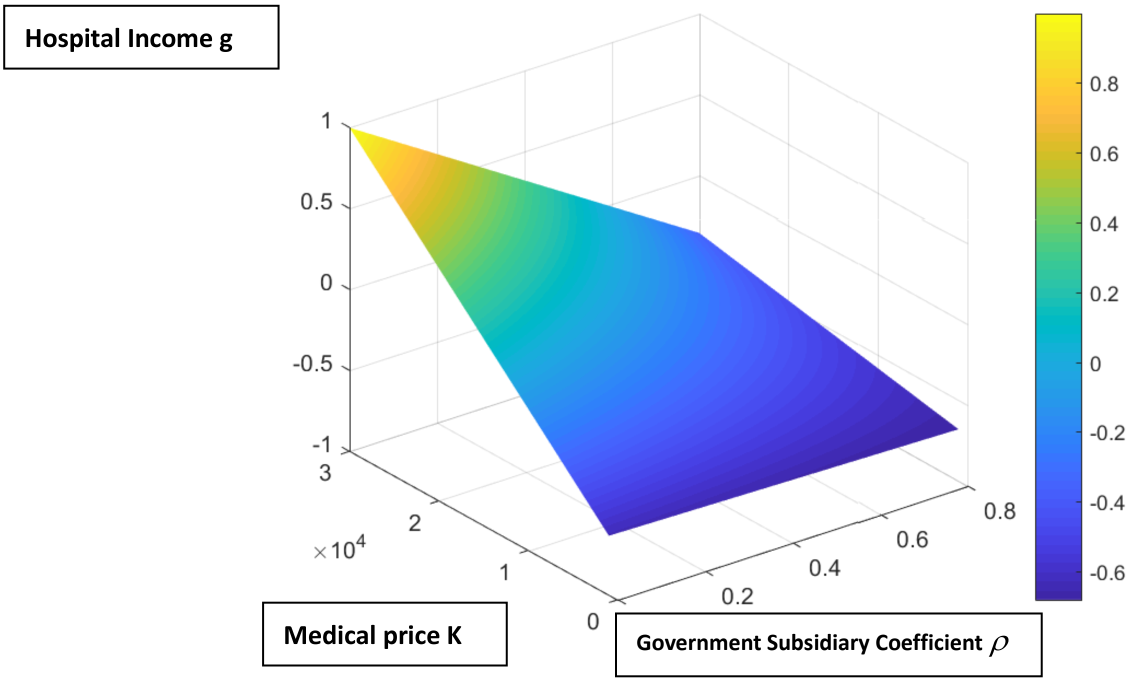

Figure 4.6

We could also conclude from the above graph (or the definition of function $g(K,\rho)$) that $\frac{\partial g}{\partial K} > 0, \frac{\partial g}{\partial \rho} < 0$. Thus the authorities could adjust K,ρ to keep the profit rate in reasonable range, ensuring the operation and development of the hospital.

4.4 Combined consideration of hospital and patient benefits

To consider the welfare for both hospital and individual patient and maximize the total benefit, we consider the combination of function $f(K,\rho)$ and $g(K,\rho)$.

We first introduce utility function $U(K,\rho)$ to express the combined utility of hospital and individual patient.

$$U(K,\rho)=q(1-f(K,\rho))+g(K,\rho) \qquad (4.10)$$

$(1-f(K,\rho))$ is the probability that the patient could afford the medical fees, thus describes individual patient's utility. $g(K,\rho)$ is the expected profit rate for hospital, thus describes hospital's utility.

We adjust q, so that the difference between maximum and minimum value for $q(1-f(K,\rho))$ and $g(K,\rho)$ is the same for the range $K\in(0,30000], \rho\in[0,0.8]$. We let $q=23.99\approx 24$ to compare the functions on the same quantitative level. We use MATLAB to graph $U(K,\rho)$:

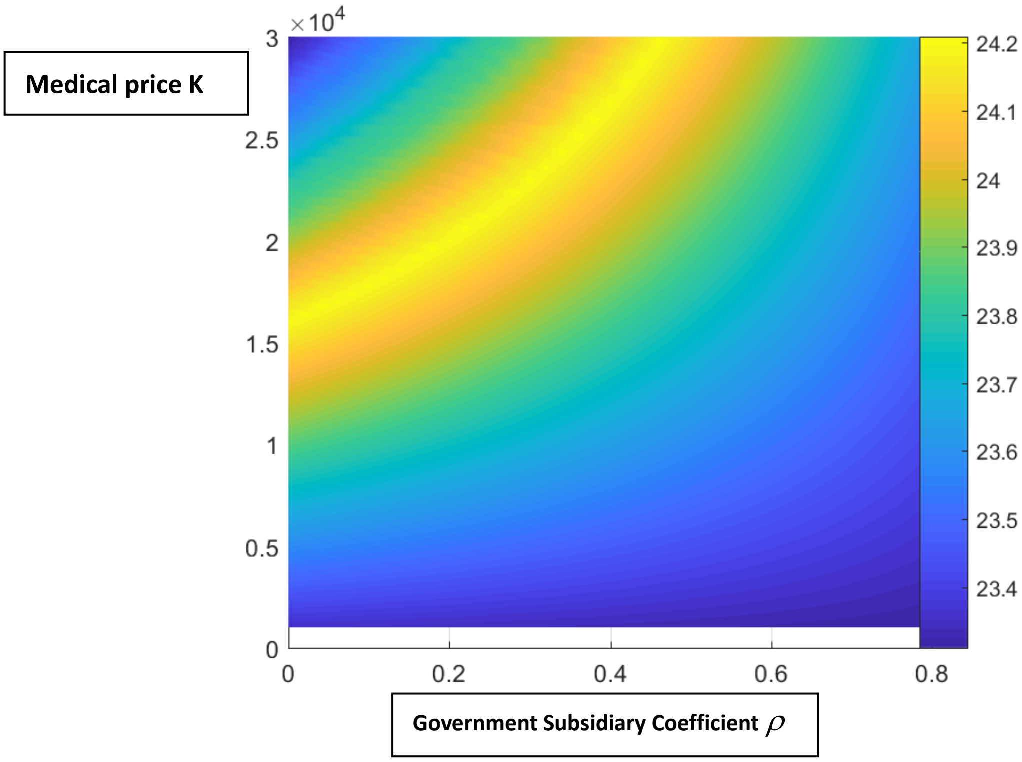

Figure 4.7

Then we select (K,ρ) of large $U(K,\rho)$ ($U(K,\rho)>24.2$), and apply polynomial curve fitting (degree 2) to study the relation between K,ρ.

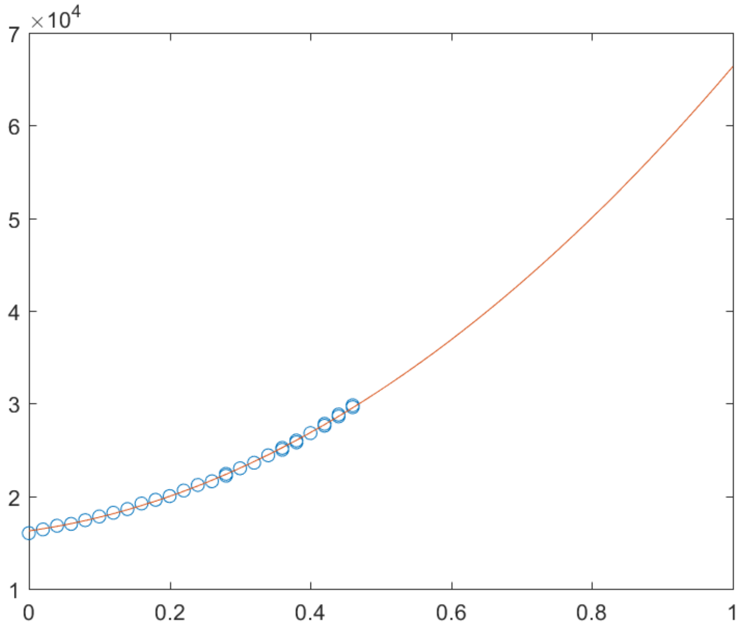

Figure 4.8

We get

$$K = 39311\rho^2 + 10819\rho + 16242 \tag{4.11}$$

Furthermore, we also aim to minimize the medical fees paid by each patient, $K(1-\rho)$. To minimize $K(1-\rho)$, we obatian that when $\rho = 0.13, K = 18313$, $K(1-\rho)$ reaches a minimum of $K(1-\rho) = 15932$. This cost is about 3000 yuan less than the original average which is greater than 19000 yuan. Meanwhile, we calculate $g(K,\rho) = 0.2058 > 0$ in this case, which means that hospitals are also benefitted.

Hence, optimizing hospital pricing and government subsidies would maximize benefit for both hospital and patients, lessening the economical burdens of patients while increasing the income and development of public hospital.

4.5 Summary

In this section we first introduce Lundberg-Cramer Ruin Theory, and construct a model for patients in a similar way. The two key variables are hospital pricing K and government subsidiary coefficient ρ. We first define a function $f(K,\rho)$, which is the probability for patients to go bankrupt in unit time. Utilizing Monte-Carlo method we are able to determine the value of and plot the graph of $f(K,\rho)$ for patients with different income. We then construct a model for the hospital. We build another function $g(K,\rho)$ that describes the expected profit rate for hospitals. Lastly we combine the two functions, and introduce a utility function $U(K,\rho)$, which is about $1-f$ and g, combing the welfare of hospital and individual patients. To maximize total utility and minimize the cost burden for patients, we finally obtain optimized value for K,ρ.

5. Conclusions and suggestions

1. DRGs payment system must be supported by scientific pricing and subsidiary method in order to function regularly. Meticulously planned hospital pricing would help lessen the burdens of patients and boost efficiency of medical resources.

2. Pricing methods could be determined after the distribution of medical cost for individual patient is estimated. The policy makers must consider the economic risks of hospital and the cost of patients. They must consider many factors such as the largest risk the hospital could bear and the estimated number of patients. Additionally, influencing factors of medical costs could be analyzed to predict

medical costs more accurately.

3. When determining the allocation of government subsidies, policy makers must consider the welfare of patients and hospitals, especially those with relatively low income who are more negatively affected by exorbitant medical costs.

6. Model evaluation

Strengths:

1. We utilize a variety of mathematical models in an effort to answer the two questions we proposed in the beginning, and give viable suggestions for policies and healthcare reforms.

2. We utilize MATLAB to plot many graphs presented in the paper. Graphs make the conclusions and results more lucid and easier to understand and interpret.

Weakness:

1. Due to limited data available, we only discuss one kind of disease. More diseases should be analyzed in the future, in order to fully implement DRGs payment system.

2. We assume that the incidence rate of disease is constant. In reality, we should also analyze the change in incidence rate with respect to time, region, and other factors.

7. References

[1] China Health Statistical Yearbook, http://www.nhfpc.gov.cn/zwgkzt/tjnj/list.shtml.

[2] Ross S.M., Introduction to Probability Models, Posts &Telecom Press, 2007.

[3] Webb A. R., Copsey K.D., Statistical Pattern, John Wiley &Sons, 2011.

[4] Zhang B., Shang H., Applied Stochastic Process (Chiness) , China Renmin University Press, 2009.

[5] Giordano F.R., Fox W.P., A first course in Mathematical Modeling, China Machine Press,2014.

[6] Fusai G., Marazzina D., PDEs Tools for Financial Engineering, a Matlab Introduction, 2013.

8. Appendix

Data

1. Average disposable income for residents in the province: 32,020 yuan

   Yearly hospitalization rate: 0.167

   Provincial public hospital income: 187,710,750,000 yuan, expenditure: 174,286,070,000 yuan, ratio m=0.928 (2015)

   Provincial government spending in healthcare: 71,531,560,000 yuan, number of hospitalized patients: 13,110,000 , government spending per capita k=5,456 yuan

   Data sources: Official website of provincial department of finance, and provincial commission of health and family planning, 2016 Chinese health statistical yearbook

2. 306 Cerebral Infraction Patients’ medical costs in a public hospital (2017.6)

No.	Age	Gender	Surgery	Infection	Length of stay /d	Cost/ yuan
1	57	M	No	No	10	12285
2	83	M	No	No	14	15069

3	77	M	No	No	15	27473
4	76	M	No	No	9	12480
5	71	M	No	No	12	8888
…	…	…	…	…	…	…
306	64	M	No	No	34	17010

MATLAB Code

Code 1

```
Data1mean=mean(Data1);
Data1var=var(Data1);
Data1sort=sort(Data1)
L = length(Data1);
[counts,centers] = hist(Data1,100);
figure(1)
bar(centers,counts/L);
C = 3;
GMM = fitgmdist(Data1,C);
Wei = GMM.ComponentProportion;
Mu = GMM.mu;
Sig1=sqrt(GMM.Sigma);
Sig =GMM.Sigma;
x = 0:10:110000;
y = 0;
for i = 1:C
   tmp = exp(-0.5*(x-Mu(i)).^2/Sig(:,:,i));
   tmpy = tmp/sqrt(2*pi*Sig(:,:,i));
   y = y + Wei(i)*tmpy;
end

figure(2)
plot(x,y,'r');


SamNum = 10000;
Samp = zeros(SamNum,1);
kk = rand(SamNum,1);
Ind = cell(3,1);
TmpVal = 0;
for i = 1:C
    TmpVal = TmpVal + Wei(i);
    Ind{i} = find(kk >= TmpVal - Wei(i) & kk < TmpVal);
```

```
    N = length(Ind{i});
    Samp(Ind{i}) = sqrt(Sig(:,:,i))*randn(N,1) + Mu(i);
end

[counts,centers] = hist(Samp,100);
figure(3)
bar(centers,counts/SamNum);

SortSamp = sort(Samp);
level = 0.99;
PValue = SortSamp(fix(SamNum*level));
```

Code 2

```
function [ K ] = prob( p )



pi=[0.7185,0.2151,0.0664];
mu=[12264.8,25355,73118];
sigma=[10046,9658,4256];
N=10000;
E=sum(pi.*mu);
Var=sum(pi.*(mu.*mu+sigma.*sigma))-E^2;
K=E+sqrt(Var/N)*norminv(1-p,0,1)

end
```

Code 3

```
X=ordinal(Cost,{'1' '2' '3'},[],[0 15000 45000 150000]);
Factors1=[Age Gender Surgery Stay Infection];
[B1h,devh,stats1h]=mnrfit(Factors1, X,'model','hierarchical')
Factors2=[Gender Surgery Stay Infection];
[B2h,dev,stats2h]=mnrfit(Factors2, X,'Model','hierarchical')
```

Code 4

```
function [ P ] = Part2_1c( Rho, K )
limit   = 1;
steps   = 250; %number of sub intervals
times   = 1000; %simulation times
lambda  = 0.1641; %frequency in unit time

salary     = 32020;
dt         = limit/steps;
dSALARY    = salary/steps;
```

```
initial    = 0.5*salary;

% initinlization
Len     = length(K);
P      =  zeros(Len,1);
count   = zeros(Len,1);
rep_K   = repmat(K,1,steps);
rep_Rho = repmat(Rho,1,steps);
Const_Incre = 0:dSALARY:steps*dSALARY;
rep_Const_Incre = repmat(Const_Incre,Len,1);

for i= 1:times %simulation cycle
   prnd = poissrnd(lambda*dt,[1,steps]);
   rep_prnd = repmat(prnd,Len,1);
   dN = rep_prnd .* rep_K .*
(1-rep_Rho);   %poissrnd(lambda*dt,[1,steps]) * K * (1-Rho);
   cdN = [zeros(Len,1), cumsum(dN,2)];

   RESULT = initial + rep_Const_Incre - cdN;

   for k = 1:Len
      if find(RESULT(k,:) < 0)
         count(k) = count(k)+1;
      end
   end
end

P = count/times;

end

x=0:0.02:0.8; Lx = length(x);
y=1000:500:30000; Ly = length(y);
[X,Y]=meshgrid(x,y);

F = Part2_1c(X(:),Y(:));
F = reshape(F,Ly,Lx);

%
figure(1)
surf(X,Y,F)
shading interp
colorbar;
```

```
figure(2)
imagesc(F(end:-1:1,:));
hold on
% axis([0 0.6 1000 30000])
```

Code 5

```
function [ profit_percentage ] = Hospital(rho, K )
k=5464;
m=0.928;
EX=19121;

profit_percentage=((1-rho)*K+k-m*EX)/(m*EX); %
end

m=0:0.02:0.8;
n=1000:200:30000;
[M,N]=meshgrid(m,n);
F=zeros(length(n),length(m));
for i= 1:length(n)
for j= 1:length(m)
F(i,j)=Hospital(M(i,j),N(i,j));
end
end
surf(M,N,F)

shading interp
colorbar;
```

Code 6（section 4.4）

```
P=0;
limit=1;
steps=250; %number of sub intervals
times=10000; %simulation times
lambda=0.1637; %frequency in unit time

count=0;
salary=32000;
dt=limit/steps;
dSALARY=salary/steps;
initial= 16000;
Simu = poissrnd(lambda*dt,[times,steps]);
Const_Incre = repmat(dSALARY:dSALARY:steps*dSALARY,times,1);
```

```
x=0:0.02:0.8;
y=1000:200:30000;
[X,Y]=meshgrid(x,y);
F=zeros(length(y),length(x));
c1=zeros(1,1000);
c2=zeros(1,1000);
pointer=0;
m=(Hospital(0,30000)-Hospital(0.8,1000))/0.07;
for i =1 : length(y)
    for j= 1:length(x)
        count=0;
        Rho= X(i,j);
        K= Y(i,j);
        Result=Simu.*(-1).*K.*(1-Rho)+Const_Incre+initial;
        for k= 1: times
            if find(Result(k,:)<0)
                count= count+1;
            end
        end
        P=count/times;
        F(i,j)=m*(1-P)+Hospital(X(i,j),Y(i,j));
        if F(i,j)>24.2
            pointer=pointer+1;
            c1(pointer)=Rho;
            c2(pointer)=K;
        end
    end
end
figure(1)
surf(X,Y,F)
shading interp
colorbar;

c1( pointer+1:1000 )=[];
c2( pointer+1:1000 )=[];
p=polyfit(c1,c2,2)

x1 = linspace(0,1,101);
f1 = polyval(p,x1);
figure(2)
plot(c1,c2,'o')
hold on
plot(x1,f1)
[pks,locs]=findpeaks(-f1.*(1-x1));
```

```
kmin=polyval(p,(locs-1)*0.01)
Hospital((locs-1)*0.01,kmin)
(locs-1)*0.01
```