Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 4265801, 13 pages
https://doi.org/10.1155/2018/4265801

Research Article

Incidence, Dependence Structure of Disease, and Rate Making for Health Insurance

Yuan-tao Xie,[1] **Juan Yang**,[2] **Chong-guang Jiang**,[1,3] **Zi-yu Cai**,[1] **and Joshua Adagblenya**[1]

[1] *School of Insurance and Economics, University of International Business and Economics, Beijing 100029, China*
[2] *Institute of Comprehensive Development, Chinese Academy of Science and Technology for Development, Beijing 100038, China*
[3] *Insurance Society of China, Beijing 100037, China*

Correspondence should be addressed to Juan Yang; yangj@casted.org.cn

Received 5 February 2018; Revised 25 May 2018; Accepted 2 July 2018; Published 12 August 2018

Academic Editor: Nazrul Islam

In order to analyze the two goals under the national strategy of "Healthy China", this paper attempts to solve the problem of coverage rate and guarantee level of health insurance, as well as the rational allocation of full life cycle health insurance resources. This paper uses pair copula to model the dependence of different disease incidence and proposes an actuarial model for rate making in health insurance based on the dependence captured by pair copula. These are far more accurate than any other model and more proper for covering a basket of several different diseases. The data for the paper was drawn from the experience incidence table of major diseases (malignant tumor, acute myocardial infarction, and stroke sequelae) from the ages 0-65 years in the Chinese life insurance industry. Extending the hypothesis of independence in actuarial modeling, the authors comprehensively use a hierarchical copula theory to extract the dependence structure of risk variable in insurance. The classification rate making technology and survival analysis method in traditional actuarial pricing were also considered. This paper applied the generalized linear model, which is commonly used in nonlife insurance pricing for empirical study of health insurance rate making. The authors discovered that the incidence of major diseases and the single premium rate calculated by the generalized linear model under HAC dependence structure were both significantly different from that calculated by the Manchester United method without dependency. The authors also stated that the rate based on the generalized linear model under HAC dependence structure was a bit different from that without dependency but both were generally the same as that of Care Expert in PICC Health. The underestimation or overestimation of systematic risks and the distortion of the rate system can be eliminated if we combine risk dependence into modeling.

1. Background

It is needed to consider a variety of diseases in health insurance pricing and their accompanied dependencies that are mutual excluded. For example, a person suffering from cerebrovascular diseases is highly likely to suffer from sudden cardiac death (coronary heart disease), heart failure, and stroke. Overlooking these relationships and moving further to carry out pricing will create bias and distortion of the rating system. However, once the dependence structure between the incidence rates is introduced, the analysis becomes complicated. In consideration, this paper introduces Hierarchical Archimedean Copula (HAC) for analysis. This is given by the dependent direction and degree of different diseases which is dissimilar holding on the characteristics of the edge distribution (for example, the edge distribution of the incidence should be the beta distribution, then the measurement of simple correlation coefficient is out of work).

The goal of "Healthy China" is to fully maintain and enhance people's health level and to achieve the coordinated development of social economy. Inferring from the planning outline of "Healthy China 2030", the primary goals are to ensure that all the citizens get full health protection and to optimize the allocation of health elements which cover full life cycle and the supply of health services as well as achieving a health protection of a full life cycle.

Fundamentally, the first issue is about the coverage and guarantee level of health insurance. Currently, the health insurance system is overall suffering from a loss and to reduce the risk, some exemption clauses such as indemnity limit, deductible, and self-paid ratio were adopted. Meanwhile, many poverty-stricken families faced with major diseases might not be able to pay for the portion of the self-paid ratio (if the ratio is higher than 50%) given by these clauses. Looking at all the complexity of the design of the insurance product, the authors have identified the key problem of the guarantee level and payment rate as being inaccurate. This paper proses a more precise and accurate pricing model.

Another issue under consideration by this paper is the rational allocation of life cycle health insurance resources. Social health insurance emphasizes the basic fairness of premium payments. It usually provides the basic part of the multilevel health insurance and the limited critical illness insurance. Yet, these insurance policies cannot reasonably balance the life cycle health risks and health insurance needs. Consistent with the design of insurance products, we need to solve the problem of rating and cost-sharing of different ages which is associated with the product design. The health risks differ in different age group, and "rating and cost-sharing" needs to be considered throughout the whole life cycle.

This study has theoretical significance and practical value since it attempts to solve the two problems presented.

The core factor in health insurance pricing is morbidity. Qiu and Chen (2014) [1] introduced the health insurance rate making model; however, they did not conduct a detailed and in-depth study on the method of health risk rates making. Olivieri and Pitacco (2011) [2] described the relationship between transition intensity and transition probability and constructed a three-dimensional Markov chain, which in turn changed the loss of income in insurance. Some scholars have discussed optimal health insurance from an optimized point of view; for example, Powell D. (2016) [3] provided theoretical guidance for the relationship of family preferences, cost-sharing, and premiums, and studied optimal health insurance and the distortionary effects of the tax subsidy.

The incidence rate is one of the core factors in this study. However, there may be some difficulties in directly modeling and pricing without considering some of its necessary features. One of them is that it lacks flexibility (it is difficult to extrapolate or interpolate the incidence table). Additionally, there has not been consideration for thick tail risk. Therefore, it is necessary to build a model using the random mortality model which is comparable to Carter et al. (1992) [4] to come with a better form. The Lee-Carter model has a simple form and a high precision which is very significant to be considered. However, Renshaw et al. (2006) [5] improved the Lee-Carter model and established the RH model, which could simultaneously model and analyze the cohort effect and period effect of a particular age. In addition to these models, a famous age-period-cohort by (Reither et al., 2015) [6] is considered useful in arriving at a general model suitable for solving the problem. Furthermore, some scholars have extended the idea of a multilevel APC model (Bell, 2014 [7]) and the CBD model (Cairns, A.J.G., Blake, D. and Dowd, 2007 [8]), which is a good predictor of mortality in the elderly population. Cairns, A.J.G., Blake, D., and Dowd (2007) [9] subsequently made a quantitative comparison of several currently useful models based on mortality data in United Kingdom, Wales, and United States. After that, the three as well as Guy D. Coughlan et al. (2011) [10] proposed an evaluation method model whereby eight random mortality models were compared. Donatien Hainaut (2012) [11] used the variable of death process to change the time-related variables and constructed a multielement Lee-Carter model that was employed to analyze the gender of France from 1946 to 2007. Giacometti et al. (2012) [12] discussed the problem of model comparison in his paper and highlighted some key important feature which worth considering. In China, some relevant modeling achievements are Wang and Huang (2011) [13] who compared several random mortality models based on the Bayesian information criterion and likelihood ratio test and He and Liu (2014) [14] presented a model characterized by introducing the state space model. They achieve this by combining the fitting stage and the prediction stage of the CBD model and estimated parameters based on Kalman filter method. They assumed that the two-factor state space model is superior to the traditional CBD model. On the other side, Debón et al. (2010) [15] used historical data from Sweden, the Czech Republic, and Spain to make some predictions. But their forecast was limited to only predicted future changes in several important parameters estimates and did not analyze the predictive results of APC model or the residual situation. Jöreskog et al. (2016) [16] in their newly published book introduced in detail an updated theoretical development of the generalized linear models. Ray et al. (2017) [17] used kernel conditional density estimation for modeling the dependence for copula and forecast the incidence. Cornelia et al. (2010) [18] discussed the HAC construction and Górecki et al. (2016) [19] discussed the determination and estimation with an Bayesian approach. When applied to health, Zimmer et al. (2017) [20], Li et al. (2017) [21], and Stöber et al. (2015) [22] give some good application examples. Nikoloulopoulos (2015) [23] and Sukcharoen (2017) [24] discussed another type of copula, i.e., vine copula, while Yang et al. (2018) [25] discussed the PCC constructing for different dependence structure.

2. Methods

2.1. Hierarchical Dependence Structure. According to Sklar (1959) theorem, the copula function can be constructed for any joint distribution function. Any copula function can be combined with any set of edge distributions to derive a multivariate distribution function. Here we use the symbols of Savu and Trede (2010) [19], denote the total level of the hierarchy with L, and let D be the dimension of HAC. Then the l level has n_l different copula functions and will be marked as (l, j), $j = 1, \cdots, n_l$. There are d_l variables in each level, $\sum_l d_l = D$, and starting from the bottom, it could be grouped into n_l subsets, and each subset has a multiple Archimedes copula.

Starting from the bottom of Figure 1, there is

$$C_{1,j}\left(u_{1,j}\right) = \phi_{1,j}^{-1}\left(\sum_{u_{1,j}} \phi_{1,j}\left(u_{1,j}\right)\right), \quad j = 1, \cdots, n_1 \tag{1}$$

where $C_{1,j}$ is the copula function between the variable 1 and variable j. $\phi_{1,j}$ is the generator of the copula $C_{1,j}$ (the generator can be different because each copula is different). $u_{1,j}$ represents the distribution function of the variable corresponding to that level. $\sum_{u_{1,j}} \phi_{1,j}(u_{1,j})$ represents the sum of the function values of all variables in the subset.

For the second level, there is

$$C_{2,j}\left(C_{1,j}, u_{2,j}\right) = \phi_{2,j}^{-1}\left(\sum_{C_{1,j}} \phi_{2,j}\left(C_{1,j}\right) + \sum_{u_{2,j}} \phi_{2,j}\left(u_{2,j}\right)\right), \quad j = 1, \cdots, n_2 \tag{2}$$

where $C_{2,j}$ is the copula function between the variable 2 and variable j. $\phi_{2,j}$ is the generator of the copula $C_{2,j}$ (as already mentioned, the generator can be different because each copula is different), $u_{2,j}$ represents the distribution function of the variable corresponding to that level. $\sum_{C_{1,j}} \phi_{2,j}(C_{1,j}) + \sum_{u_{2,j}} \phi_{2,j}(u_{2,j})$ represents the sum of the functional values of all variables in the subset.

In this paper, three diseases were selected to establish a double HAC model for analysis. These are

$$\begin{aligned} C_{1,1} &= \phi_{1,1}^{-1}\left(\phi_{1,1}\left(u_1\right) + \phi_{1,1}\left(u_2\right)\right) \\ C_{2,1}\left(u_1, u_2, u_3\right) &= \phi_{2,1}^{-1}\left(\phi_{2,1}\left(C_{1,1}\right) + \phi_{2,1}\left(u_3\right)\right) \\ &= \phi_{2,1}^{-1}\left(\phi_{2,1}\left(\phi_{1,1}^{-1}\left(\phi_{1,1}\left(u_1\right) + \phi_{1,1}\left(u_2\right)\right)\right) + \phi_{2,1}\left(u_3\right)\right) \end{aligned} \tag{3}$$

According to McNeil Theorem 4.4 (2008), the general condition of which a general hierarchical structure is an appropriate Archimedean Copula function is that all nodes occurred are completely monotonous. Only selective copulas are available in this situation. We could choose Hierarchical Archimedean Copula functions, including Clayton, Frank, or Gumbel copulas form.

We often operate on the model reversely in random simulation. That is, starting from the top of the first generation to generate random variables distributed evenly between 0 and 1 and then based on Laplace transform to obtain random variables. This is based on the conditional distribution to generate random variables that meet a given distribution and a generator.

Using the method of integration on different diseases and considering the actuarial models comprehensively, it is possible to arrive at a joint actuarial model of k diseases. For example, people who have heart disease may suffer from sudden cardiac death (coronary heart disease), heart failure, and stroke in a higher probability, while the relationship with an infectious disease can be seen as similar to independent.

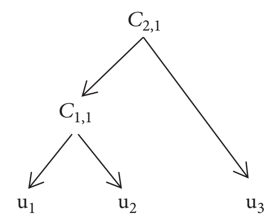

Figure 1: Hierarchical Archimedean Copula Structure.

Using copula technique to extract the dependence structure between diseased species, we have

$$\begin{aligned} {}_t p_0 = TP(t) &= \oiint_D P_{D_1, D_2 \cdots D_k} \\ &\cdot f_{1,2,\cdots k}\left(D_1, D_2 \cdots D_k\right) dD_1 \cdot dD_2 \cdots dD_k \\ &= \oiint_D P_{D_1, D_2 \cdots D_k}(t) \\ &\cdot copula\left(D_1, D_2 \cdots D_k\right) dF_1\left(D_1\right) \\ &\cdot dF_2\left(D_2\right) \cdots dF_k\left(D_k\right) \end{aligned} \tag{4}$$

where ${}_t p_0$ means the probability of incidence for a newborn child (age equals 0) to suffer disease at age t. $D_1, D_2 \cdots D_k$ are diseases (rate) factors and P represents the incidence rate that matches the rate factor. The $F(\cdot)$ means the integration (cumulative distribution function, CDF) of $f(\cdot)$ (PDF, probability density function). f_k is the joint probability density. F_k means the integration (cumulative distribution function) of f_k.

2.2. Premium Determination

2.2.1. Pricing Formula. The premium formula of long-term major disease insurance under unit insurance coverage (for example, ¥10,000) is

$$\mathrm{P}(x, n) = \int_0^n {}_t h_x \mathrm{v}^t dt = \int_0^n h(x+t) \mathrm{v}^t dt \tag{5}$$

where x represents the age at which the insured is insured. $\mathrm{P}(x, n)$ represents the premium rate to cover n years for a person aged. t represents t years after the insurance contract take into effect. v represents the discount factor. $h(\cdot)$ represents the disease intensity (hazard ratio function) of the insured during the policy term.

The key problem is how to get $h(x+t)$. This paper will try to apply the generalized linear model, which is commonly used in the nonlife actuarial to the health insurance pricing. We obtain ${}_t p_0$ equation by constructing the model and get the single premium through integration. This can avoid making the difference between the pure premium and the expected loss too large through several approximate calculations.

The model we build by CBD model about t on ${}_t p_0$ is as follows:

$$\ln\left(\frac{{}_t p_0}{1 - {}_t p_0}\right) = \beta_0 + Gender \cdot \beta_1 + t\beta_2 + \varepsilon \tag{6}$$

Variable *Gender* means the gender factor for the insured. According to (5), we also need to deduce hazard rate function $h(\cdot)$ according to ${}_tp_0$:

$$\begin{aligned} h(t) &= \frac{f(t)}{1-F(t)} = \frac{dF(t)/dt}{1-F(t)} = \frac{d\,{}_tp_0/dt}{1-{}_tp_0} \\ &= \frac{\beta_2 \cdot \exp\{\beta_0 + Gender \cdot \beta_1 + t\beta_2\}}{1+\exp\{\beta_0 + Gender \cdot \beta_1 + t\beta_2\}} = \beta_2 \cdot {}_tp_0 \end{aligned} \quad (7)$$

So

$$h(x+t) \approx \beta_2 \cdot {}_{t+x}p_0 \quad (8)$$

Substituting formula (8) and (4) into formula (5), we can derive the premium.

2.2.2. The Traditional Manchester United Method. The Manchester United method is simple to calculate, so far, many health insurance companies use the method to price an insurance product. Based on the Manchester United method, we can derive the calculation of the single premium formula of long-term major disease insurance [2]:

$$\widetilde{\mathrm{P}}(x,n) = \int_0^n \frac{l_{x+t}}{l_x} k_{x+t} v^t dt \cong \sum_{h=1}^{n} \frac{l_{x+h-1/2}}{l_x} v^{h-1/2} \alpha_{x+h-1} \quad (9)$$

where x represents the insured age. $\widetilde{\mathrm{P}}(x,n)$ represents the premium rate of long-term major disease insurance to cover n years for a person aged x. l_x indicates the population at age x in the life table. k_{x+t} indicates the morbidity at age $x+t$. v is the discount factor. α represents the incidence of disease center.

Although the Manchester United method is convenient, it used the approximating method to substitute the integration with summation, so there must be some errors when we estimate the premium by way of Manchester United method. And the statistics constructed according to the Manchester United method are not sufficient statistics.

3. Results

3.1. Modeling of Morbidity Dependence Structure. The data in this paper is from the experience incidence table of 25 major diseases in Chinese life insurance industry at the age of 0-65 for analysis. Recently, many insurance policies cover 36 kinds of major diseases that meet the national regulations. Some of the products even cover more than 60 kinds of major diseases. But in fact, the 36 diseases prescribed by the government are major diseases with relatively high incidence rates, and the incidence rates of other major diseases are extremely low, which can almost be negligible for premium pricing. And among the 36 major diseases, the malignant tumor, acute myocardial infarction, and stroke sequelae own the incidence of absolute dominance, and malignant tumor is the umbrella name for a wide range of diseases. Therefore, the selection of the incidence of these three major diseases is the finest representative for analysis.

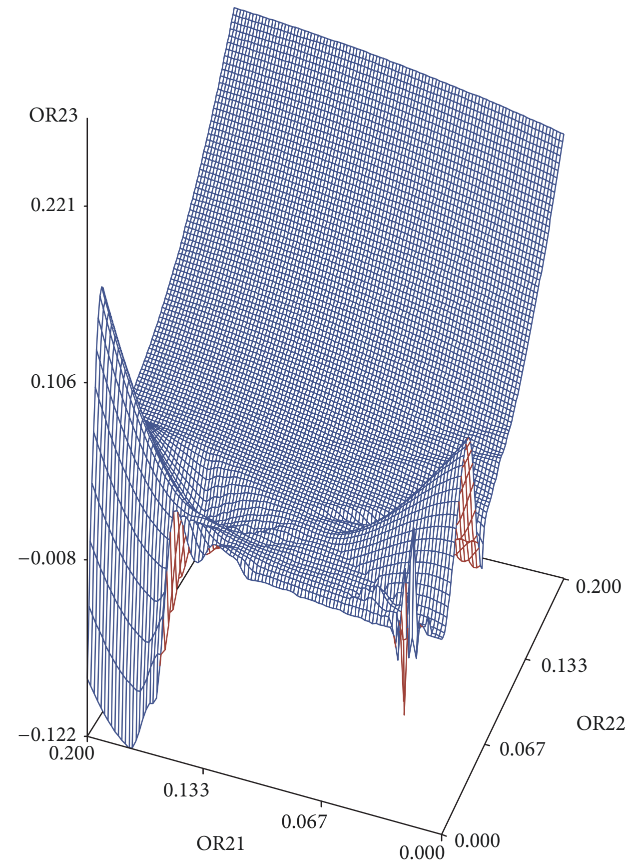

Figure 2: The 3D plot for incidences.

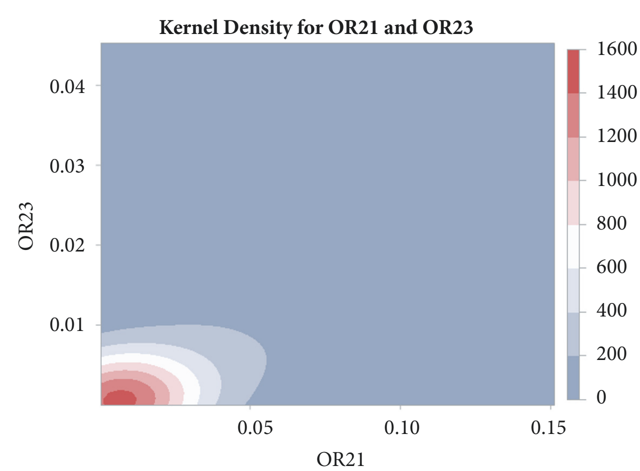

Figure 3: Density for OR21 and OR23.

The descriptive statistics for fitted incidence after the Hierarchical Archimedean Copula modeling and estimation are shown in Table 1.

From Figure 2 we can see the 3D plot for incidences. The kernel density for OR21 and OR23 is shown in Figure 3 while the kernel density for OR22 and OR23 is shown in Figure 4.

The probability density for OR21 and OR23 is shown in Figure 5 while the probability density for OR22 and OR23 is

TABLE 1: Descriptive statistics for male incidence of different major diseases.

Variable	Label	Mean	Std	Min	Max
OR21	Fitted incidence for malignant tumor	0.0279	0.0376	0.0003	0.1511
OR22	Fitted incidence for acute myocardial infarction	0.0048	0.0098	0.0000	0.0486
OR23	Fitted incidence for stroke sequelae	0.0053	0.0094	0.0000	0.0453

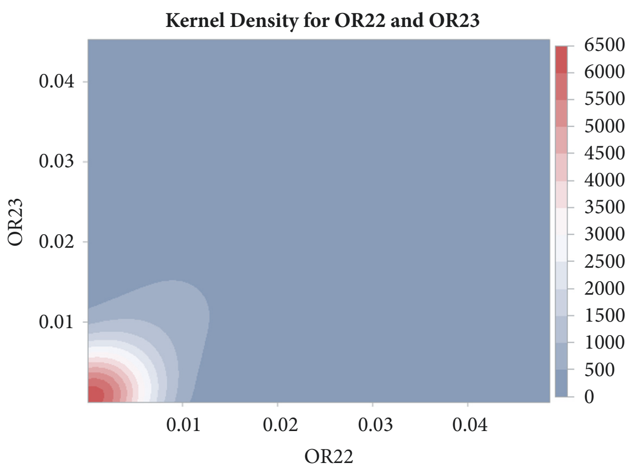

FIGURE 4: Density for OR22 and OR23.

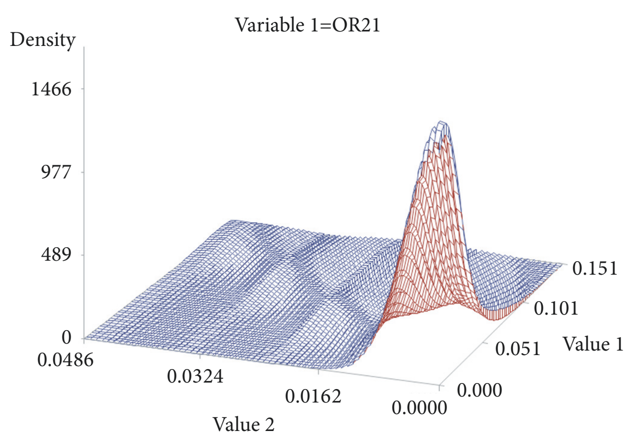

FIGURE 5: Density for OR21 and OR23.

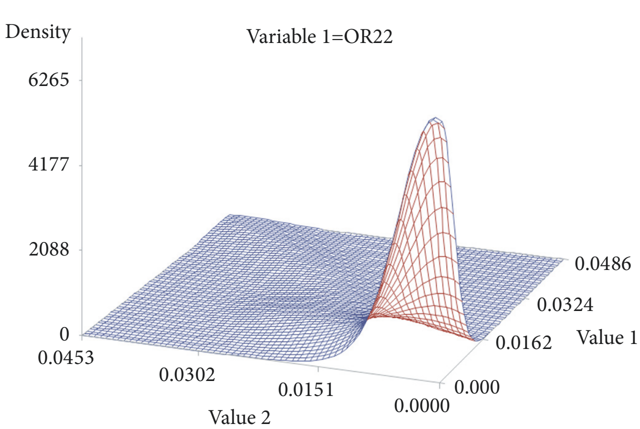

FIGURE 6: Density for OR22 and OR23.

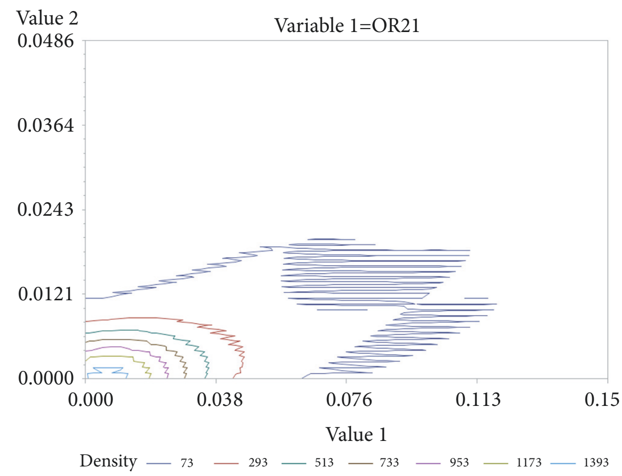

FIGURE 7: Density for OR21 and OR23.

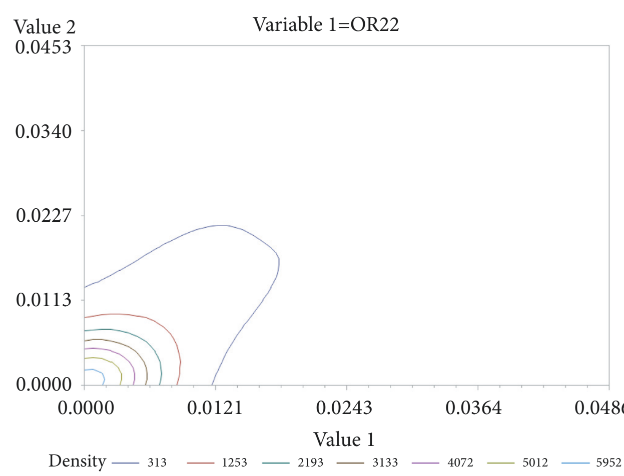

FIGURE 8: Density for OR22 and OR23.

shown in Figure 6. According to Figure 5, we can see that the probability density for the OR21 and OR23 has a very high uplift at a lower value, which means that the two variables show a great dependence on the lower value. Moreover, the entire diagonal part shows an obvious ridge, which showed dependence somehow. The basic situation in Figure 6 is similar to Figure 5, except that the ridge is not so obvious for OR22 and OR23, and the dependence is much smaller except the lower tail.

The contour plot for OR21 and OR23 is shown in Figure 7 while the contour plot for OR22 and OR23 is shown in Figure 8. It can be seen that, in a considerable range of values, OR21 and OR23 have equal levels. The contour map shows

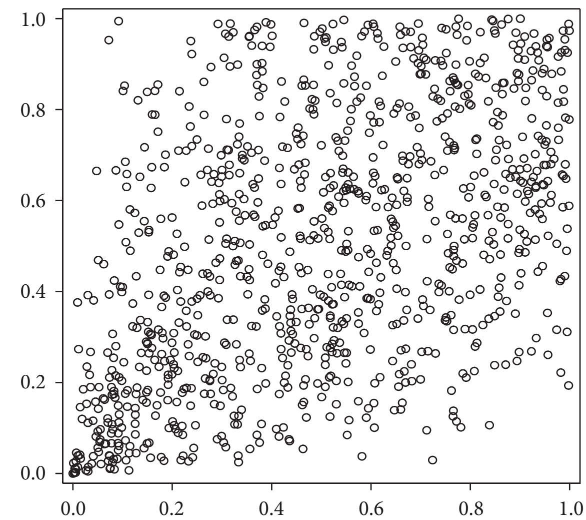

Figure 9: Upper copula fitting.

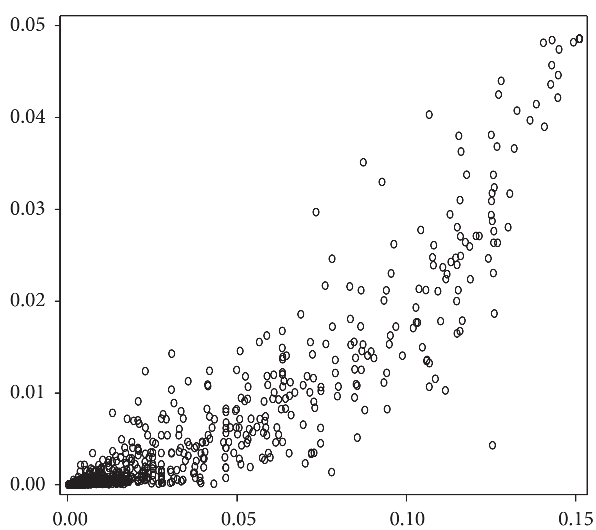

Figure 10: Lower copula fitting.

a lot of knots and overlaps. In Figure 8, the dependence structures for OR22 and OR23 are similar, but the contours are much smoother and the knot phenomenon is much better.

We can get a comprehensive incidence for different gender at different age.

The Chinese life insurance industry in the year 2010 to 2013 sampled 1 million people surviving up to the age of 105 to produce the experience incidence table for single disease at the ages from 0 to 65. Figures 9 and 10 show the upper copula and lower copula fitting.

3.2. Modeling of Generalized Linear Model. Now, we compare the results of the Manchester United method with the single premiums of major disease insurance products named Care Specialist with an insurance period of 20 years in PICC Health, where the single premiums are from male policy holders at different insuring age.

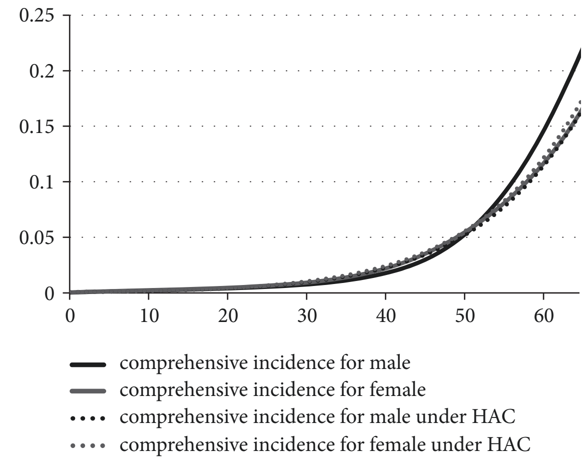

Figure 11: Incidence comparison for both gender.

The results of the analysis of the maximum likelihood estimate are shown in Table 2. We can see that the model works well through the analysis and we get the following generalized linear model:

$$\ln\left(\frac{{}_t\widehat{p}_0}{1-{}_t\widehat{p}_0}\right) = -7.1244 - 0.0685 \cdot Gender + 0.0858 \cdot t \quad (10)$$

The deviation of the model is 4.6300 and the adjusted deviation is 132.7671; the Pearson chi is 4.3255 and the P value is 0.0335; the logarithmic likelihood is 612.6400, the AICC is -1216.9650, the BIC reaches -1205.7487, and the whole model is very significant. According to the results of type III variance analysis of LR statistics, the LR statistics of age variables is 595.61 and the P value <.0001; the LR statistics of gender variables is 4.35 and the P value is 0.0369. Comparing with the results of type I variance analysis and the goodness test of the model, it is assumed that the whole model is very significant.

We obtain this form by solving

$$ {}_t\widehat{p}_0 = \frac{\exp\{-7.1244 - 0.0685 \cdot Gender + 0.0858 \cdot t\}}{1+\exp\{-7.1244 - 0.0685 \cdot Gender + 0.0858 \cdot t\}} \quad (11)$$

Figure 11 shows a comparison of the incidence of both gender and the nondependent incidence of both genders predicted by the generalized linear model under the HAC dependence structure. It can be seen that, for males, the incidence of major diseases predicted by the generalized linear model under the HAC dependence structure of the newborn (0-2 years old) is higher than that without dependency; but from the age of 3, the incidence of major diseases predicted by the generalized linear model under the HAC dependence structure of males is lower than that without dependency. For females, the characteristics are that it peaks at both sides and low in the middle. In order to show that the difference between the two is statistically significant; this paper compares the incidences of major diseases at ages that are integral by gender and that the P values are all less than 0.01.

According to formula (5), it is easy to calculate the single pure premium paid at 0-50 years old of major disease

Table 2: Maximum likelihood estimate.

Parameter		The degree of freedom	Estimation	Standard error	Wald 95% confidence interval	Wald chi-square	Pr > chi-square
Intercept		1	-7.1244	0.0346	-7.1922	-7.0565	42347
Age		1	0.0858	0.0008	0.0842	0.0874	10786
Gender	Male	1	-0.0685	0.0326	-0.1323	-0.0047	4.43
Gender	Female	0	0.0000	0.0000	0.0000	0.0000	
Scale		1	28.6756	3.5094	22.5601	36.4490	

insurance with an insurance period of 20 years; the results are shown in Table 2. Regardless of the dependencies between diseases, this paper uses the Manchester United method to calculate the pricing of only three categories of major disease insurance and compares the pricing with the calculation results of the generalized linear model under the HAC dependence structure. Substituting formula (10) into formula (8) and then formula (5), we can theoretically obtain the explicit solution. While taking into account the fact that the formula of the solution is very reliable, this paper uses the numerical integration to calculate the premiums directly. According to our calculations, in the group of age 0-2 and whatever the gender, the single premium rate calculated by the Manchester United method without dependency is significantly lower than the rate calculated by the generalized linear model under HAC dependence structure. The incidence of major diseases is very low for newborns and infants. Therefore, once the situation occurs, the reason will often be congenital or hereditary, so it has a characteristic of strong dependency. Limited to the difficulty of data acquisition, we cannot extract the microcrystalline (microarray) information for cause analysis, but the dependence structure embodied by the data basically reflects this trend. With the increase of age, after 50 years, comparing the incidence determined by the generalized linear model under HAC dependence structure with the incidence without considering dependence structure, we observed that the male's major diseases incidence of latter is high, while the rate of female is low.

3.3. Rate Making and Salary Share Analysis. Considering that there are no insurance products that only cover three types of major diseases in the market, as a contrast, we choose the single premiums of major disease insurance products named Care Expert with an insurance period of 20 years in PICC Health for analysis. Take 50-year-old person as an example; according to the experience incidence table of major disease in Chinese life insurance industry (2006-2010), population of males who suffer from three major diseases accounts for 72.9% of that of 25 major diseases, while the population of females accounts for 75.1%; when the age comes to 60 years, population of males who suffer from three major diseases accounts for 76.3% of that of 25 major diseases, while the population of females accounts for 76.5%, as is shown in Figure 12. We calculate the premium and adjust in accordance with this ratio, making the premium comparison on the same basis.

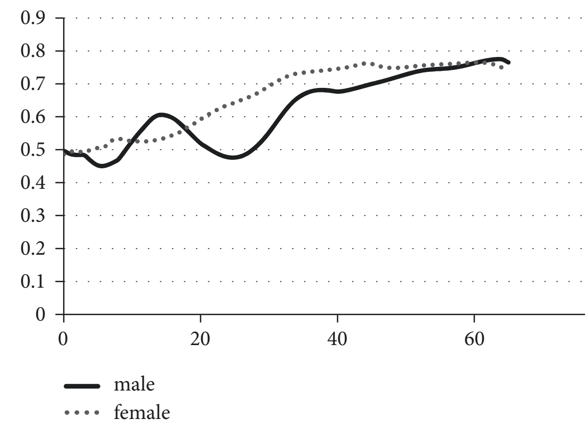

Figure 12: Incidence of 25 diseases.

This paper calculates single pure premium of the independent main type of major disease insurance and uses the commonly used Manchester United method and the nonlife insurance pricing method based on the generalized linear model. What is more, the insurance period is 20 years, the insurance amount is 10000 (that is, as long as the insured suffers from one of the 25 kinds of diseases mentioned above during the insurance period, he or she will receive an indemnity of 10000), and assumed interest rate is 2.5%. This paper is estimated at an overall cost rate of 20% and an additional rate of 5%. The result is shown in Table 3.

The comparison of the rate based on the generalized linear model under HAC dependence structure and the single premium of major disease insurance products named Care Expert in PICC Health by gender is shown in Figure 13; the comparison of the rate based on the Manchester United method without considering the dependence structure and the single premium of major disease insurance products named Care Expert in PICC Health by gender is shown in Figure 14.

We can see that there is a certain difference between the results based on Manchester United method and the single premium of the Care Expert. To be specific, the premium we calculate is the single pure premium of the independent main type of major disease insurance, the insured liability does not contain the risk of death, and there is no further consideration of additional premiums. Therefore, the single pure premium based on Manchester United method is relatively lower than that of the Care Expert.

Table 3: The 20-year payment rate table.

Age	Male	Female
0	4.4466	4.8746
1	4.9486	5.2147
2	5.4206	5.6713
3	5.9123	6.2022
4	6.6847	6.6583
5	7.4954	7.1741
6	8.1856	7.7561
7	8.7609	8.1593
8	9.3008	8.8065
9	9.5803	9.6167
10	9.8928	10.6095
11	10.2808	11.5311
12	10.7391	12.5667
13	11.2804	13.6245
14	12.0878	14.7379
15	13.2119	15.8768
16	14.5824	17.0404
17	16.3400	18.2677
18	18.4221	19.4223
19	20.8096	20.7305
20	23.5276	22.0529
21	26.2721	23.4886
22	29.3548	25.1003
23	32.6189	26.8230
24	35.9430	28.8320
25	39.2126	31.0213
26	42.2645	33.3215
27	45.0219	35.8306
28	47.4293	38.4327
29	49.6161	41.0938
30	51.5560	43.8345
31	53.3132	46.8643
32	55.2493	50.1393
33	57.6480	53.8774
34	60.8336	58.0830
35	64.8031	62.8958
36	69.5367	68.0585
37	74.9776	73.7480
38	81.4499	79.8569
39	88.7428	86.4127
40	96.6820	93.4586
41	104.6395	101.0047
42	112.7835	108.9319
43	121.4145	117.4151
44	130.6169	126.5458
45	140.4890	137.4095
46	151.1845	150.1536
47	162.6015	163.5616
48	174.7599	177.5205
49	187.7563	192.1068
50	201.6729	207.4667

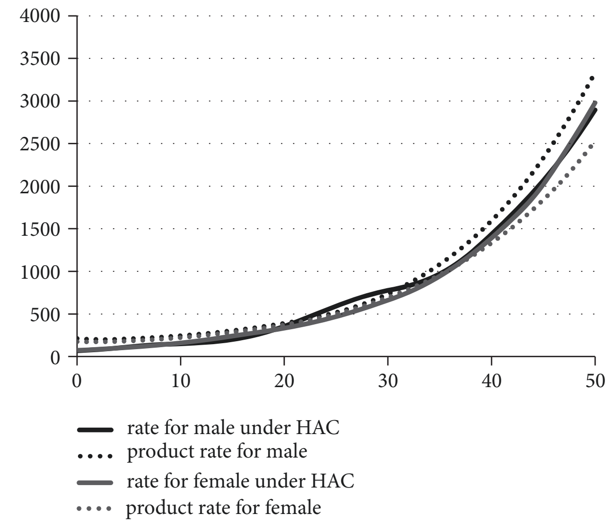

Figure 13: Rate comparison under HAC.

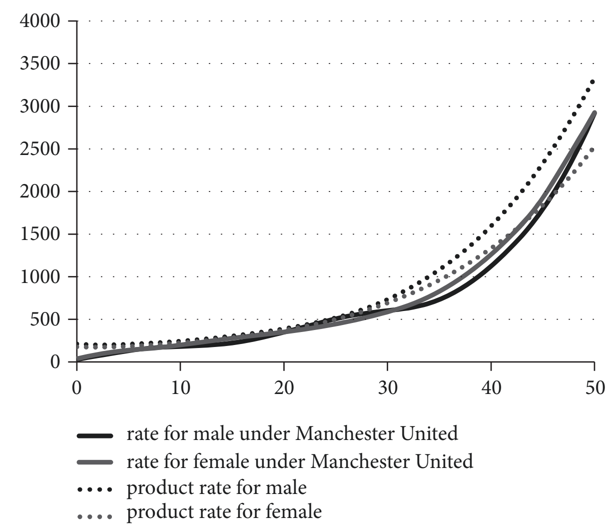

Figure 14: Rate comparison under Manchester United.

Although there are many shortcomings in the method used in this paper, the nonlife insurance pricing method based on the expected loss is more comprehensive than that of the long-term major disease insurance method. Comparing the single pure premium calculated by the generalized linear model and that of the Care Expert, we can see that using the generalized linear model under the HAC dependence structure or using the Manchester United method produces the level of the price as that of PICC Health. The error of parameter setting (for example, the overall rate of fees increases from 20% to 25%) may cause the premium to be the same as in the old age range. However, the focus of this paper is not adjusting the parameters to make the values equal in total but to focus on the premium structure of the rating system. We can find from observing the rating system that the ratio of the premium under HAC dependence structure and that of products in PICC of males are 31.74% and females are 42.11% at 0 years old. With this age, the ratio is gradually increasing. At the age of 22, males' rate exceeds females for the first time reaching 102.17% and 91.13% for female; at the age of 32, males' rate falls below 1 for the first time reaching 96.78% and 95.50% for female. After that, males' rate steadily declines and females' gradually increases. At 37 years old, females' rate exceeds 1 reaching 100.63%; at 50 years old, males' rate is 86.84% and females' is 117.63%. Thus, without considering the risk dependence, the risk of males at the age of 22-31 has been underestimated, the premium is relatively low, and the risk of other ages has been overestimated and the premium is relatively high. For females, the risk of the young has been overestimated, the premium is relatively high, but the risk between 37 and 50 years old has been underestimated and the premium is relatively low. On the whole, there are male-to-female premium subsidy effects and child-to-adult premium subsidy effects; for females, there are young-to-old premium subsidy effects, and for males, there are children and old-to-young premium subsidy effects. This is in general consistent with the industry's perception of health insurance.

Regardless of the dependency of diseases can easily distort the premiums. It is known from the analysis above that an applicant who has adverse selection will choose to purchase health insurance after 18 years old if the information is transparency. Males will purchase between 22 and 31, while females will purchase after 37 years old. This is obviously not conducive to premium fairness. There are two reasons for the distortion of the premium system. One is that the health insurance is still in the early stages of development, the main purpose of each insurance company is to improve market share and does not intend to profit in the health insurance business; the other is that young group purchase health insurance for either having a high level of awareness of health or the welfare of the work unit is not comprehensive. For the first group of young people, their work units have already paid for their medical insurance and some even could withstand the risk of major disease with their own economic strength; if the insurance company has a higher price, it will combat the purchase enthusiasm of such quality customers. For the second group of young people, their income level and the education level generally cannot be compared with those of the first group, and they generally do not take much care of themselves because of the working environment or the survival pressure. As the main source of income in the family, such kind of group shoulders most of the responsibilities of the family and has a strong enthusiasm to buy health insurance to pass on risks of major disease. However, they have a relatively large risk to suffer from diseases.

Base on the life annuity and the discount rate of 2.5%, we calculate the conversion function and construct the life annuity coefficient table by age and gender and combine with the 20-year payment rate table based on the generalized linear model under the HAC dependence structure under per million of insurance amount.

In the year of 2016, the average wage of workers in Beijing has reached 7086 yuan per month. According to the ratio recommended by CIRC "Notice", major disease insurance premiums should be set to 5 to 10 times the annual income. Assuming the individual monthly income is M (unit: yuan), then the annual income should be $12M$, and in accordance with the ratio recommended by CIRC "Notice", the proposed total amount of protection should be $[60M, 120M]$ (unit: yuan). Assuming that the 20-year payment rate of gender i and age t under per million of insurance amount is p_{it}, then the premium should reach $[(60 \cdot p_{it} \cdot M)/10000, (120 \cdot p_{it} \cdot M)/10000]$ (unit: yuan), and the proportion of the premium payable to the wage level should reach $[6p_{it}/1000, 12p_{it}/1000]$. Thus, the ratio of the minimum guaranteed level of premiums to the wage is shown in Table 4.

Taking the 30-year-old working population, for example, males should take 30.93% of their own monthly salary for major disease insurance, and females should take 26.30%, so that it could cover the medical expenses that are 5 times the income level if they suffer from one of the 25 major diseases at the age between 30 and 50. This ratio is even higher than its medical insurance premiums. In fact, with age, this proportion will gradually increase, and at the age of 50, the proportion of premiums spent on major diseases is even higher than the total amount of wages. This result is obviously unreasonable. On the other hand, infants and young children do not have income, so the provision of a certain percentage does not make any sense. Therefore, the correct interpretation of the table is to measure based on the family. For example, a family of four (the elderly was covered by their own premiums), where both the husband and the wife are 30 years old and have equal wages, and they have a 2-year-old boy and a newborn girl, then the major disease premium should account for 31.8% of household income. Overall, the burden on the family is still very heavy. Part of the coordination areas tries to use 1% of the wage level as a major disease payment base. According to the above analysis, its coverage and security level will be seriously inadequate.

Therefore, it is recommended that the government should control the cost of medical care and appropriately reduce the level of protection.

4. Discussion

4.1. Main Finding of This Study. If we omit the dependence structure before setting a premium for a health insurance that is covering many different disease, there might be some distortions in the rating for different individuals, different risk groups, and especially different age period. These distortions will result in moral hazard or adverse selection, which may be dangerous to the aim of insurance.

4.2. What Is Already Known on This Topic. Considering the dependencies between multivariate, there are already relevant applied research results in China. Du and Gao (2013) [26] used the HAC model to study the relationship between risk and infection, and the applied research is distinctive. Luo and Wu (2016) [27] combined the Co VaR model with the HAC model to measure spillover effects of systemic risk in the stock market. There are also some studies that have characteristics at the technical level. For example, He (2010) [28] discussed the high-dimensional HAC model and its dynamics and

Table 4: The ratio of the minimum guaranteed level of premiums to the wage.

Age	Male	Female
0	2.67%	2.92%
1	2.97%	3.13%
2	3.25%	3.40%
3	3.55%	3.72%
4	4.01%	3.99%
5	4.50%	4.30%
6	4.91%	4.65%
7	5.26%	4.90%
8	5.58%	5.28%
9	5.75%	5.77%
10	5.94%	6.37%
11	6.17%	6.92%
12	6.44%	7.54%
13	6.77%	8.17%
14	7.25%	8.84%
15	7.93%	9.53%
16	8.75%	10.22%
17	9.80%	10.96%
18	11.05%	11.65%
19	12.49%	12.44%
20	14.12%	13.23%
21	15.76%	14.09%
22	17.61%	15.06%
23	19.57%	16.09%
24	21.57%	17.30%
25	23.53%	18.61%
26	25.36%	19.99%
27	27.01%	21.50%
28	28.46%	23.06%
29	29.77%	24.66%
30	30.93%	26.30%
31	31.99%	28.12%
32	33.15%	30.08%
33	34.59%	32.33%
34	36.50%	34.85%
35	38.88%	37.74%
36	41.72%	40.84%
37	44.99%	44.25%
38	48.87%	47.91%
39	53.25%	51.85%
40	58.01%	56.08%
41	62.78%	60.60%
42	67.67%	65.36%
43	72.85%	70.45%
44	78.37%	75.93%

introduced the covariate into the hidden Markov model. He also used the EM algorithm on parameter estimation. Zhang and Hu (2014) [29] used the ARMA-GARCH process to eliminate the autocorrelation and conditional heteroscedasticity of the sequence and used the two-stage maximum likelihood method to estimate the HAC function. Examples of empirical analysis can see Xie et al. (2015) [30].

4.3. What This Study Adds. (1) The main idea of traditional health insurance pricing is as follows: premiums at insurance level are equal to the sum of premiums at a responsible level. In fact, the premiums at all levels of responsibility are priced separately to address heterogeneity. Such a split pricing will result in an underestimation or overestimation of systematic risks; also it will distort the rating system, make the pricing of high-risk individuals too high or too low, and lead to an excessive rewards and punishments effect. This paper makes prices in a system so that it can improve the rate of compensation and comprehensive rate of fees and ensure a better health.

(2) Extending the hypothesis of independence in the actuarial studies, considering the dependence structure of risk variables, measuring risks accurately, and providing a reference for the analysis of the competitive risk and multirisk. This paper takes full account of the dependence structure between disease types to make the price. For example, acute myocardial infarction and stroke sequelae show a very significant dependency, but the relationship with the malignant tumor is much weaker; if we just put them together to extract the dependency, there will be a serious distortion in the rating system.

(3) In terms of the design and pricing of insurance products, this paper takes into account the classification of rate making technology and survival analysis ideas in the traditional actuarial pricing models. What is more, the generalized linear model that is commonly used in nonlife insurance pricing is applied to the empirical study to discuss a new method of determining health insurance rates.

(4) Through the comparative analysis of different calculation, the result of this paper analyzes the characteristics of model pricing and gives some suggestions on the pricing of health insurance products.

(5) The analysis and modeling adopted in this paper help to calculate the price of any insurance age without the rate table and any interpolations for noninteger age insurance premiums. This effectively solves the problem of volume purchase of health insurance before birthdays.

(6) In order to link the proposed level of protection, this paper analyzes in detail the proportion of premiums required to meet the preset level of protection to the income and provides a reference for the study of guarantee level.

4.4. Limitations of This Study. This paper only considers the incidence of major diseases in pricing process, and the data used are the insurance industry experience data, which will be different from the natural incidence in China. The lack of data and the limit of authors' ability may make the rates proposed in this paper to have a certain insignificant deviation. Besides, this paper proposes to develop a classification rate determination system based on different situations of the insured. It is necessary to model the variables such as age, insurance period, genetic history, living area, and so on; to explore the experience rate combined with history; even to integrate classification and experience rate system to develop a comprehensive rate system to further reduce premium distortions and excessive rewards and punishments effect (Xie Yuantao and Li Zhengxiao, 2015 [24]). Because of the lack of data, this paper does not take these variables into account in the empirical study. And the estimation of the transition probability is not sufficient, which is the possible direction of further study.

5. Conclusions and Outlooks

5.1. Conclusions. Health insurance consumption is accompanied by two contradictions of uncertainty. The first pair of contradictions is to protect the future loss, which may not occur with a certain income. According to the loss aversion effect of prospect theory, it is not equal in the effectiveness of the sensitivity, so even if the health insurance makes a loss, there is still an illusion that the premiums are expensive. The second pair of contradictions is that when the serious illness occurs, the insured person is facing a disgusting effect on health loss, especially in the face of uncertain treatment results, which will lead to overtreatment problems. According to the previous analysis, the existing wage level is unlikely to meet the protection needs of major diseases, so the government needs to reduce the level of protection, reduce medical costs, and control the problem of overtreatment.

The problems in the development of health insurance are mainly as follows: the rate is unscientific, the role of positioning is not clear, health insurance products are seriously homogenized, the business model has defects, and professionals are scarce. This paper analyzes focusing on these perspectives.

Premiums of domestic health insurance market grew rapidly, but health insurance businesses generally suffer losses. Low prices and high costs may be just one aspect, the rate structure is more important. For example, the inequality of rate distributed in gender, age, and diseases, on the one hand, will affect the overall cost and overall rate of insurance operations and on the other hand will lead to an adverse selection. This paper believes that the rate factors considered by the domestic rate determination system are too less, the independence test of rate factors is not enough, the loss estimation is not sufficient, the rating system is easy to distort, and it is likely to make excessive reward and punishment effect and the premium subsidy effect. A typical example is that men pay for women and young people pay for the elderly.

The life insurance company mainly operates Chinese health insurance business, and the health insurance products launched by these companies are mostly tied with participating investment type of insurance or deposit making the life insurance policies relatively lacking a protective function. The investment or savings functions of a health insurance are not better than that of the bank or fund companies. Hence, the capital income is relatively low for high-income groups, and the cost is relatively high for middle-low income groups who really need health protection. Consequently, the health insurance cannot play a role to improve the health level of the whole citizens in the health care system.

The Chinese type of health insurance products are single and most insurance companies offer major disease insurances with a fixed payment, while almost no companies provide nonfixed medical expenses compensation insurances. Although it is the global development trend, the insurer still has to think about other institutional arrangements. To some extent, fixed payment makes the amount of compensation much higher than the medical expenses, so that it is easy to mobilize the enthusiasm of the insured. And once the reverse selection occurs, it will undoubtedly lead to a large waste of insurance funds and cannot effectively solve the poverty problem caused by diseases. So, at this stage, we should develop some products, and position health insurance needs in basic medical expenses compensation so that the system could resolve the risk of serious illness through the purchase of suitable health insurance products; then the health insurance could play a role in the Chinese health care system.

Generally, if any country sets a strategy for health insurance's development, there is no unified point of view about whether using commercial insurance model or social security model is best appropriate. Usually, it dependents on the decision of the government to decide which model to use. The academic and the industry fields have being exploring on how to stimulate the development of health insurance industries such as the introduction of tax-deferred products and preferential tax products and its benefits to the people but it is not compatible with the Chinese tax system. China's Income Tax classification system implements the model of withholding and self-tax declaration, collection, and management, while the United States implements a unified excess progressive tax rate in personal income tax and takes a comprehensive tax levying on net income and uses a collection and management pattern of the source withholding, self-prepaid, and comprehensive declaration of the annual collection based on the family unit. Therefore, preferential tax products are very difficult to implement in China and are only effective for a small number of new products after reporting. Insurance products must be innovative, covering different levels of family needs, given the system an edge, to develop vigorously.

Insurance innovation still has room for improvement. Companies that carry out health insurance business should develop many types of policies to meet the needs of the community. The insurance companies should have a long-term strategic vision in the course of business and at the same time play a role to protect people's livelihood and maintain social stability. Essentially, the company should not decide the product structure of the business based on short-term profit situation but to analyze the market demands and provide policies that meet the needs of the market by accumulating market experience. And only in this way, could profits be sustained. In terms of company operations, the asset share model should be introduced and discussions on health insurance business should be carried out in longer term.

5.2. Outlooks. We set up a "new triangle quadrilateral" paradigm of "integrated health insurance and health maintenance" management and decision-making driven by large data in the founding assembly of Health Insurance Professional Committee of Chinese Preventive Medicine Association. The idea of traditional insurance is that individuals pay a certain premium, if the disease occurs and then takes part of the fund for medical expenses or pay a fixed amount. This insurance model is passive and inefficient to achieve fund balance through actuarial modeling. Should China introduce the new insurance model of health management, will help optimize health management and achieve early effective intervention on health risk, that is, taking out part of the money from the fund for routine examination to achieve early detection, treatment, and intervention. By this way, the real incidence and payment can be greatly reduced, the insured will be healthy, the hospital will avoid waste of medical resources, the guarantor will reduce the cost, and the fund manager will realize a full optimization in the whole system. It is much good than harm for all parties to change the passive insurance model to an active health insurance model. The traditional insurance analysis focuses on the asset and liability management from the cash flow, duration, convex, and other aspects that the claims are passive. After introducing health management, the health intervention will be that the claims are significantly reduced, thereby improving the rate of compensation and comprehensive rate of fees to protect the health of people.

The health insurance rate is determined by the incidence rate and that is similar to the life table for life insurance pricing which are both very important. The government has gradually defined the development direction that professional health insurance companies need to operate the health insurance business at the regulatory level. This is conducive for the accumulation and sharing of the industrial data and improving the professional level of the health insurance industry.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest, and the received funding, grants, or scholarships do not lead to any conflicts of interest regarding the publication of this manuscript.

Acknowledgments

The paper was financially supported by the National Social Science Foundation of China "Research on the Redistribution Function of the Social Security System" under Grant no. 18BJY212 and "the Fundamental Research Funds for the Central Universities" in UIBE (CXTD9-04) and joint education program of UIBE. Thanks are due to Jing-hua Cui for her help in the writing of the paper.

References

[1] C. Qiu and T. Chen, "A study on the rate of compensated hospitalization medicare insurance—based on the analysis of claim cost hypothesis," *Shanghai Economic Review*, vol. 1, pp. 111–121, 2014.

[2] A. Olivieri and E. Pitacco, *Life Insurance: Pricing. Introduction to Insurance Mathematics*, Springer, Heidelberg, 2011.

[3] D. Powell, *Optimal health insurance and the distortionary effects of the tax subsidy*, Social Science Electronic Publishing, 2013.

[4] L. R. Carter and R. D. Lee, "Modeling and forecasting US sex differentials in mortality," *International Journal of Forecasting*, vol. 8, no. 3, pp. 393–411, 1992.

[5] A. E. Renshaw and S. Haberman, "A cohort-based extension to the Lee-Carter model for mortality reduction factors," *Insurance: Mathematics and Economics*, vol. 38, no. 3, pp. 556–570, 2012.

[6] E. N. Reither, R. K. Masters, Y. C. Yang, D. A. Powers, H. Zheng, and K. C. Land, "Should age-period-cohort studies return to the methodologies of the 1970s?" *Social Science & Medicine*, vol. 128, pp. 356–365, 2015.

[7] A. Bell, "Life-course and cohort trajectories of mental health in the UK, 1991-2008—A multilevel age-period-cohort analysis," *Social Science & Medicine*, vol. 120, pp. 21–30, 2014.

[8] A. J. Cairns, D. Blake, and K. Dowd, "A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration," *Journal of Risk and Insurance*, vol. 73, no. 4, pp. 687–718, 2006.

[9] A. J. G. Cairns, D. Blake, K. Dowd et al., "A quantitative comparison of stochastic mortality models using data From England and wales and the United States," *North American Actuarial Journal*, vol. 13, no. 1, pp. 1–35, 2009.

[10] A. J. G. Cairns, D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, and M. Khalaf-Allah, "Mortality density forecasts: an analysis of six stochastic mortality models," *Insurance: Mathematics and Economics*, vol. 48, no. 3, pp. 355–367, 2011.

[11] D. Hainaut, "Multidimensional Lee-Carter model with switching mortality processes," *Insurance: Mathematics and Economics*, vol. 50, no. 2, pp. 236–246, 2012.

[12] R. Giacometti, M. Bertocchi, S. T. Rachev, and F. J. Fabozzi, "A comparison of the Lee-Carter model and AR-ARCH model for forecasting mortality rates," *Insurance: Mathematics and Economics*, vol. 50, no. 1, pp. 85–93, 2012.

[13] X. Wang and S. Huang, "Comparison and selection of stochastic prediction model of population mortality in China," *Population & Economics*, vol. 1, pp. 82–86, 2011.

[14] Y. He and G. Liu, "Two-factor stochastic mortality state space model and longevity risk measure," *Financial Theory and Practice*, vol. 5, pp. 24–28, 2014.

[15] A. Debón, F. Martínez-Ruiz, and F. Montes, "A geostatistical approach for dynamic life tables: the effect of mortality on remaining lifetime and annuities," *Insurance: Mathematics and Economics*, vol. 47, no. 3, pp. 327–336, 2010.

[16] K. G. Jöreskog, U. H. Olsson, and F. Y. Wallentin, "Generalized linear models," in *Multivariate Analysis with LISREL*, Springer Series in Statistics, pp. 135–169, Springer International Publishing, Cham, 2016.

[17] E. L. Ray, K. Sakrejda, S. A. Lauer, M. A. Johansson, and N. G. Reich, "Infectious disease prediction with kernel conditional density estimation," *Statistics in Medicine*, vol. 36, no. 30, pp. 4908–4929, 2017.

[18] C. Savu and M. Trede, "Hierarchies of Archimedean copulas," *Quantitative Finance*, vol. 10, no. 3, pp. 295–304, 2010.

[19] J. Górecki, M. Hofert, and M. Holeňa, "An approach to structure determination and estimation of hierarchical Archimedean Copulas and its application to Bayesian classification," *Journal of Intelligent Information Systems*, vol. 46, no. 1, pp. 21–59, 2016.

[20] D. Zimmer, "Using copulas to estimate the coefficient of a binary endogenous regressor in a Poisson regression: application to the effect of insurance on doctor visits," *Health Economics (United Kingdom)*, vol. 27, no. 3, pp. 545–556, 2018.

[21] R. Li, Y. Cheng, Q. Chen, and J. Fine, "Quantile association for bivariate survival data," *Biometrics*, vol. 73, no. 2, pp. 506–516, 2017.

[22] J. Stöber, H. G. Hong, C. Czado, and P. Ghosh, "Comorbidity of chronic diseases in the elderly: patterns identified by a copula design for mixed responses," *Computational Statistics & Data Analysis*, vol. 88, pp. 28–39, 2015.

[23] A. K. Nikoloulopoulos, "A vine copula mixed effect model for trivariate meta-analysis of diagnostic test accuracy studies accounting for disease prevalence," *Statistical Methods in Medical Research*, vol. 26, no. 5, pp. 2270–2286, 2017.

[24] K. Sukcharoen and D. J. Leatham, "Hedging downside risk of oil refineries: a vine copula approach," *Energy Economics*, vol. 66, pp. 493–507, 2017.

[25] J. Yang, Y. Xie, and Y. Guo, "Panel data clustering analysis based on composite PCC: a parametric approach," *Cluster Computing*, vol. 5, pp. 1–11, 2018.

[26] Z. Du and L. Gao, "Study on financial crisis transmission path based on strata condition copula," *Technoeconomics & Management Research*, vol. 10, pp. 76–80, 2013.

[27] Y. Luo and Y. Wu, "Stock market linkage measure based on hierarchical copula theory," *Journal of Chongqing Institute of Technology (natural science)*, vol. 11, pp. 171–176, 2016.

[28] X. He, "Construction and application of dynamic hierarchical archimedes copula model," *Statistics and Decision*, vol. 13, pp. 4–8, 2010.

[29] L. Zhang and X. Hu, "Correlation analysis of financial time series based on hierarchical archimedes copula," *Statistics & Information Forum*, vol. 6, pp. 34–40, 2014.

[30] Y. Xie and Z. Li, "Expansion and comparison of rewards and punishments factors based on joint pricing model," *Statistics & Information Forum*, vol. 6, pp. 33–39, 2015.