Segment-specific Bayesian analysis of health insurance costs

Kateřina Seinerová [1]

Abstract

The paper presents an application of an empirical Bayesian credibility model onto real data. The credibility costs combine individual regional experience and collective experience through the credibility factor. The credibility factor expresses the level of confidence in individual risk experience and takes into account heterogeneity of portfolio and amount of experience within the portfolio. This paper presents the differences of the credibility factors and the health care costs in several health care segments among the Czech regions. Estimated values of health care costs for the following time period are gained using the Bühlmann Straub model.

Key words

Bayesian estimation, Credibility factor, Health care

JEL Classification: C11, C31, I11, I31

1. Introduction

Every health insurance company needs to appraise health care expenses for subsequent time periods. For this purpose, it is necessary to estimate these expenses by applying mathematical and probabilistic models, for example those based on Bayesian analysis. In this paper, the Bayesian analysis is applied for such estimation in case of the main health insurance company in the Czech Republic.

The health insurance in the Czech Republic is provided through the country's Social Health Insurance system. It requires all businesses to provide workers membership in one of several health insurance funds, to which both employers and employees contribute. The Czech government provides contribution for the unemployed so that essentially the population is universally insured. The health care in the Czech Republic is funded from public health insurance, direct payments, the national budget and regional budgets. The health care system strives to create conditions in which there are no differences in the availability of health care. Solidarity between healthy people and the sick is fostered in health care systems by separation between the provision of health care and its financing. Solidarity of the economically active with the economically inactive people means that every insured person pays an insurance premium as a percentage of their income regardless of what health care they receive or will receive.

Under these specific conditions, the data from the major health insurance company, which is the Všeobecná zdravotní pojišťovna (VZP), are representative for the whole insurance system, and their credibility analysis results may reflect the actual tendencies in this area.

[1] Ing. Kateřina Seinerová,Ústav matematiky a kvantitativních metod, Fakulta ekonomicko-správní, Univerzita Pardubice, katerina.seinerova@upce.cz.

2. Data

In the Czech Republic, public health system is provided by seven health insurance companies. The main insurer, that insures majority of the Czech population, is Všeobecná zdravotní pojišťovna (VZP). For this reason, data from VZP yearbooks was used for computation in this work. Expenses in several segments of the health care cover the time period 2008 – 2013.

The hospitalization segment involves expenses by inpatient care facilities, such as hospitals, hospices or eventide homes, including the acute care provided in these facilities. The day surgery segment involves expenses by outpatient care facilities. The medication segment involves expenses for medicaments administered in both inpatient and outpatient care facilities.

The dentistry segment involves expenses related to all types of dental care. The transportation segment involves expenses for transport of patients, including emergency and bone marrow transport. The equipment segment involves expenses for common medical equipment used in all types of health care facilities. [5]

Different segments exhibit different evolutions in time. For example, while expenses for dentistry decrease moderately in all the regions, the expenses for transportation exhibit oscillations in some regions and significant increases in the others.

3. Bühlmann Straub model

The Bühlmann Straub model uses a Bayesian approach and belongs amongst free distribution models, so it needs to establish hypothesis neither on the distribution of the individual risks, nor on the prior distribution of the risk parameters.

The key difference between classical statistics and Bayesian analysis consists in the perception of an unknown parameter θ. This parameter is supposed to be a random variable instead of an unknown constant. This random variable has a probability distribution $f(\theta)$, called prior distribution. In the insurance practice, information on insurance payment amount and distribution can be used to determine the prior distribution of the parameter θ. Health insurance companies are then able to specify posterior distribution of the parameter θ using inner information from regions.

The Bühlmann Straub model combines individual risk experience with experience from collective of similar risks. The result of the application of the Bühlmann Straub model is the linear credibility premium (or claim frequency or claim size) which uses both the individual experience as well as the collective experience. In other words, the credibility premium combines individual experience and collective experience through the credibility factor. The credibility factor expresses the level of confidence in individual risk experience and takes into account heterogeneity of portfolio and amount of experience with the portfolio. Therefore the Bühlmann Straub models as well as other empirical Bayes credibility models are useful for situation when insurer have to face the problem with a relatively large heterogeneity of portfolio. [3]

Let Y_{ij} is a variable, describing segment health care costs in the i-th region of the Czech Republic ($i = 1, \ldots, N = 14$) in the j-th year ($j = 1, \ldots, n = 6$).

Let P_{ij} (for $i = 1, \ldots, N$ and $j = 1, \ldots, n$) stands for the number of VZP insurers (in the i-th region of the Czech Republic and in the j-th year).

Standardized health care costs, i. e. average health care costs per person $X_{ij} = \frac{Y_{ij}}{P_{ij}}$, satisfy the following conditions:

- For every i = 1, …, N the distribution of variable X_{ij} depends on an unknown parameter θ_i, equal for all the years j = 1, …, n.
- For every j = 1, …, n the variables $X_{i1}/\theta_i, X_{i2}/\theta_i, ..., X_{in}/\theta_i$ are independent, but not necessarily identically distributed.

Then, two functions depending on j can be defined as:

$$E(X_{ij}/\theta_i) = m(\theta_i) \tag{1}$$

$$D(X_{ij}/\theta_i) = \frac{s^2(\theta_i)}{P_{ij}} \tag{2}$$

The above mentioned relations express conditions satisfied for every i = 1, …, N. Relationships between regions are described by the following conditions:

- Parameters of risks $\theta_1, ..., \theta_N$ are random variables, which are independent and identically distributed.
- For $i \neq k$ are (θ_i, X_{ij}) and (θ_k, X_{km}) independent.

Because parameters of risks $\theta_1, ..., \theta_N$ are identically distributed, the values $E(m(\theta_i)), E(s^2(\theta_i)), D(m(\theta_i))$ are independent on i and we can denote them as $E(m(\theta)), E(s^2(\theta)), D(m(\theta))$.

For our calculations the following formulas were used:

$$P_i = \sum_{j=1}^{n} P_{ij} \tag{3}$$

$$P = \sum_{i=1}^{N} P_i \tag{4}$$

$$\overline{X}_i = \frac{1}{P_i}\sum_{j=1}^{n} P_{ij}X_{ij} = \frac{1}{P_i}\sum_{i=1}^{n} Y_{ij} \tag{5}$$

$$\overline{X} = \frac{1}{P}\sum_{i=1}^{N}\sum_{j=1}^{n} P_{ij}X_{ij} = \frac{1}{P}\sum_{i=1}^{N} P_i\overline{X}_i \tag{6}$$

$$P^* = \frac{1}{Nn-1}\sum_{i=1}^{N} P_i \cdot \left(1 - \frac{P_i}{P}\right) \tag{7}$$

Then according to [2], the rules for the estimates of the parameters $E(m(\theta)), E(s^2(\theta)), D(m(\theta))$ are

$$estE(m(\theta)) = \overline{X} \tag{8}$$

$$estE(s^2(\theta)) = \frac{1}{N(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{n} P_{ij}\left(X_{ij} - \overline{X}_i\right)^2 \tag{9}$$

$$estD(m(\theta)) = \frac{1}{P^*}\left\{\frac{1}{Nn-1}\sum_{i=1}^{N}\sum_{j=1}^{n} P_{ij}\left(X_{ij} - \overline{X}\right)^2 - \frac{1}{N(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{n} P_{ij}\left(X_{ij} - \overline{X}_i\right)^2\right\} \tag{10}$$

Credibility factor for the i-th region, according to [1], is calculated in form

$$Z_i = \frac{P_i}{P_i + \frac{E(s^2(\theta))}{D(m(\theta))}} \tag{11}$$

Estimates of the parameters $E(m(\theta)), E(s^2(\theta)), D(m(\theta))$ are the same for all the regions, but the credibility factor Z_i differs from region to region. The higher is value of credibility factor Z_i, the higher is the value P_i which characterizes the extent of the risk.

Then according to [1] and [4], for the estimation of credible health care costs the formula

$$E(m(\theta)/X) = Z_i\overline{X}_i + (1-Z_i)E(m(\theta)) = Z_i\overline{X}_i + (1-Z_i)\overline{X} \quad (12)$$

is used.

4. Results

Results of the computations applying Bühlmann Straub model are presented in this section. Computations are based on a set of real data extracted from VZP yearbook, namely the amount of persons insured by VZP and health care costs paid by VZP in main health care segments in particular regions of the Czech Republic in years 2008-2013. Bühlmann Straub model is applied to compute credible regional health care costs for the following time period.

Table 1 presents estimation of parameters of Bühlmann Straub model $estE(m(\theta))$, $estE(s^2(\theta))$ and $estD(m(\theta))$ computed according to (8-10). These values show different pattern in each health care segment.

Table 1: Bayesian estimates of parameters

Estimation of parameters	Hospitalization	Day surgery	Medication	Dentistry	Transportation	Equipment
$estE(m(\theta))$	11,9061	4,415	3,61866	0,91956	0,39008	0,65793
$estE(s^2(\theta))$	88176,6	46144,7	10505,2	213,409	2845,7	978,475
$estD(m(\theta))$	33,2866	1,70322	0,61971	0,0102	0,00391	0,10424

Table 2: Credibility factors for main health care segments in the considered regions

Region	Hospitalization	Day surgery	Medication	Dentistry	Transportation	Equipment
H. m. Praha	0,99941	0,99404	0,99626	0,99539	0,86106	0,99793
Středočeský	0,99938	0,99369	0,99604	0,99512	0,85416	0,99781
Jihočeský	0,99891	0,98897	0,99307	0,99146	0,76917	0,99615
Plzeňský	0,99872	0,98708	0,99187	0,98999	0,73955	0,99548
Karlovarský	0,99791	0,97903	0,98677	0,98372	0,63443	0,99263
Ústecký	0,99924	0,99228	0,99516	0,99403	0,82697	0,99731
Liberecký	0,99868	0,98661	0,99158	0,98963	0,73266	0,99532
Královéhradecký	0,99870	0,98685	0,99173	0,98982	0,73624	0,99541
Pardubický	0,99881	0,98794	0,99242	0,99066	0,75280	0,99579
Vysočina	0,99885	0,98834	0,99267	0,99097	0,75908	0,99593
Jihomoravský	0,99937	0,99355	0,99595	0,99501	0,85131	0,99776
Olomoucký	0,99843	0,98413	0,99001	0,98770	0,69750	0,99444
Moravskoslezský	0,99902	0,99007	0,99376	0,99231	0,78757	0,99654
Zlínský	0,99886	0,98847	0,99275	0,99107	0,76113	0,99597

Table 2 presents values of credibility factors Z_i, computed according to (11) for each region in considered segments. The value of credibility factor Z_i shows the effect of the regional data on the value of the credible health care costs. $(1 - Z_i)$ shows the same effect for the national data. The values of credibility factor are very high in all the Czech regions for most of the segments. The exception is the transportation segment, where the values of credibility factor show, that the effect of regional data is not high enough and it is necessary to consider national data.

Table 3 presents values of credible health care costs for main health care segments, computed according to (12). The values of the credible health care costs can be used to estimate the real regional health care costs in the following time period. It means, based on data 2008-2013, the computed values of credible health care costs in each health care segment estimate real regional health care costs in 2014. Whereas the values of credible health care costs are comparable in almost all the regions in the dentistry segment, these values differ from region to region in other segments. We can observe an outstanding position of region Prague, where are the highest credible health care costs per person in hospitalization, day surgery, medication, and equipment segments, but (according to small area) the lowest credible costs for transportation.

Table 3: Credible costs (in CZK) for main health care segments in the considered regions

Region	Hospitalization	Day surgery	Medication	Dentistry	Transportation	Equipment
H. m. Praha	25056	7257	5277	999	311	1342
Středočeský	6754	3254	2728	690	398	432
Jihočeský	9737	3692	3149	950	389	588
Plzeňský	13638	4314	3589	966	492	892
Karlovarský	6649	4757	2929	827	481	311
Ústecký	8327	4258	3055	854	365	425
Liberecký	9312	3223	2963	898	405	535
Královéhradecký	13480	3618	3797	955	382	683
Pardubický	6946	5934	3354	972	364	406
Vysočina	8749	3473	3168	983	371	308
Jihomoravský	13382	3806	3934	940	479	809
Olomoucký	13674	4006	4186	1046	380	625
Moravskoslezský	13353	4941	4081	923	334	813
Zlínský	7868	3915	3369	1004	362	390

5. Conclusion

Computations in this paper are based on a set of real data, in particular health insurance data from the regions of the Czech Republic in years 2008-2013. Bühlmann Straub model is applied to compute credible regional health care costs in main health care segments. Health care costs per person vary from region to region. The computed values of credibility factors Z_i show that the health insurance companies cannot rely on the national data, because the majority effect on the results relates very often to the regional information. The computed values of the regional credible health care costs can be used to estimate the real health care costs for the following time period.

Acknowledgement

This paper was supported by the project No.SGSFES_2015001 Ekonomický a sociální rozvoj v soukromém a veřejném sektoru (Economical and social development in private and public sector), financed from Czech Republic funds.

References

[1] Boland, P. J. (2007). Statistical and Probabilistic Methods in Actuarial Science. New York: Taylor & Francis Group.

[2] Bühlmann, H. and Straub, E. (1970). *Glaubwürdigkeit für Schadensätze.* Bern: Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker.

[3] Jindrová, P. (2014). Credibility Risk Models in Accident Insurance. In: *Proceedings of the 7th International Scientific Conference: Managing and Modelling of Financial risk.* Ostrava: VŠB – TU Ostrava, Faculty of Economics, Finance department. pp. 307-316

[4] Pacáková, V. (2004). *Aplikovaná poistná štatistika*, 3th ed. Bratislava: EKONÓMIA.

[5] VZP – Všeobecná zdravotní pojišťovna České republiky (2015). Retrieved from: http://vzp.cz/