ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 194, 2005

*Anna Szymańska**

# APPLICATION OF SELECTED STATISTICAL METHODS IN ASSESSING HOMOGENEITY OF INSURANCE PORTFOLIO

**Abstract**

The foundation of insurance company activity is proper adjustment of premium level to the risk level of the insured. The insurer usually groups policies in portfolios characterized with similar risk.

However, there exist risk factors not observable directly, having impact on the claim size and frequency. An important issue, therefore is the assessment of portfolio homogeneity.

The purpose of this work is the assessment of selected methods of testing portfolio homogeneity illustrated with an example of motor insurance.

**Key words:** homogeneity, portfolio, risk factors.

## I. INTRODUCTION

A set of insurance policies in particular kind of insurance is called a portfolio. The policies of certain insurance portfolio are grouped into sets called tariff classes. A kind of risk represented by particular policy is a criterion of that division into classes and is understood as an expected loss of an insurer.

Basic assumptions of portfolio construction are:

1. An insurance policy is located in particular tariff class (which is called a sub-portfolio or a group) on the basis of known risk factors.

2. The classes should be characterized by similar level of risk and greater homogeneity than the whole portfolio.

3. Within a particular class, similar number and size of losses for individual policies are expected. This implies similar insurance rate.

* Ph.D., Chair of Statistical Methods, University of Łódź.

4. Within a particular class, the policies can be grouped into sub-classes, depending on the number and size of losses in previous years (bonus-malus systems). Depending on the system, there are different models of transition from one class to another.

In case of drivers civil responsibility insurances (OC), an insurer can observe only part of factors that decide about the level of risk, i.e.: production year and type of car, the aim of car usage, engine capacity, driver's age and gender. However, there are also factors that cannot be directly observed but which considerably influence the risk level of particular driver. Therefore the issue of sub-portfolio homogeneity assessment is essential. The majority of the methods of insurance rates assessment need the assumption of homogeneity in portfolio classes as well.

## II. SELECTED METHODS

This paper is an attempt of indicating methods, which may be used to assess the homogeneity in insurance portfolio, and directs special attention to statistical tests.

Let us assume that insurer registers only occurrence of a loss or lack of it (assuming the occurrence of one loss once a year only). Then the random variable (i.e. the number of losses in portfolio) follows the binomial distribution. From portfolio we randomly pick $p$ policies. The random variable $Y_j$, $j = 1, 2, ...., p$, is the number of losses for $j$-th policy in $n_j$ of years.

Then the probability estimator of loss occurrence for the pooled sample (i.e. for whole portfolio) has the following form (Niemiro, 1997):

$$q = \frac{\sum_{j=1}^{p} Y_j}{\sum_{j=1}^{p} n_j} \tag{1}$$

For testing the hypothesis that all of random variables, $Y_i$, follow the same distribution one may apply the chi-square goodness-of-fit test (Domański, 2000). Then the null hypothesis has the form: $H_0$: $\theta_1 = ... = \theta_j = ... = \theta_p$ where $\theta_j$ is a structural parameter of the distribution of the number of losses. We reject $H_0$, when $\chi^2 \geqslant \chi^2_\alpha$.

In case of rejecting the null hypothesis, the method presented above may be considered effective. A portfolio may be treated as heterogeneous with regard to the number of losses. If there is no ground for rejecting the null hypothesis one shall look for other methods of portfolio homogeneity assessment.

In automobile insurances it is assumed that number of losses, $X$, in homogeneous portfolio is a random variable following the Poisson distribution and with the parameter of loss intensity $\lambda$:

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad (x = 0, 1, 2, ...). \tag{2}$$

Then the question of the examination of portfolio or portfolio classes homogeneities is reduced to the verification of fit between the number of losses and the Poisson distribution with $\chi^2$ test (Domański, 2001).

If portfolio is non-homogenous then the parameter of loss intensity has usually the gamma distribution with parameters $\alpha$ and $\beta$ while the number of losses has the negative binomial distribution with parameters $p$ and $k$ (Hossack et al., 1999). It means that its probability function has the following form:

$$P(X = x) = \binom{k + x - 1}{x} p^k(1 - p)^x, \quad (x = 0, 1, 2, ...), \tag{3}$$

where

$$k = \alpha \quad \text{and} \quad p = \beta/(1 + \beta). \tag{4}$$

Therefore, if the distribution of the number of losses fits ($\chi^2$ fit-test) the negative binomial distribution, then there is no ground for rejecting the hypothesis of portfolio heterogeneity.

Another method of portfolio homogeneity assessment with respect to the number of losses may be the graphical method proposed by Hossack. This method it is assumes that the number of losses in portfolio follows the negative binomial distribution. Therefore we can say that non-homogeneity of portfolio is assumed. Then, the parameters of the gamma distribution of random variable of losses intensity are calculated. The next step is to draw a graph of the density of the probability distribution of the number of losses. If the graph is similar to the graph of the probability density of the gamma distribution we infer that the portfolio is heterogeneous.

## III. APPLICATIONS

The study conducted was based on data from one insurance company (in the city of Łódź), for OC automobile insurances for the year 2000. From the whole portfolio, containing 31 734 policies, 15 867 of them were drawn independently and grouped according to driver's age. The data are presented in Table 1.

This study aims to assess the portfolio homogeneity with the methods presented above. Application of statistical inference methods needs one sample. That gives adequate numbers of observation in specified (according to the number of losses) classes.

**Table 1.** Number of losses in two portfolio groups and in the whole portfolio

| Group | | I | II | Whole portfolio |
|---|---|---|---|---|
| Age of a driver (in years) | | less than 25 | 25 or more | |
| No. of claims | 0 | 2 907 | 10 221 | 13 128 |
| | 1 | 592 | 1 843 | 2 435 |
| | 2 | 66 | 210 | 276 |
| | 3 | 5 | 18 | 23 |
| | 4 | 0 | 5 | 5 |
| No. of observations | | 3 570 | 12 292 | 15 867 |

**Source:** Insurer's company data.

For portfolio groups and for the whole portfolio (from Table 1), the fit between the distribution of the number of claims and the Poisson distribution and the negative binomial distribution was examined using the chi-square test.

Let $r$ be the number of classes, $n_i$ empirical amount in $i$-th class, $np_i$ theoretical (expected) amount in $i$-th class. The chi-square goodness-of-fit test statistic has the following form:

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i}. \tag{5}$$

On the basis of data from Table 1, the parameter of the Poisson distribution was assessed assuming

$$\hat{\lambda} = \overline{x} \tag{6}$$

and the parameters of the negative binomial distribution were assessed assuming

$$\hat{p} = \frac{\overline{x}}{s^2}, \tag{7}$$

$$\hat{k} = \frac{s^2 \hat{p}^2}{1 - \hat{p}}, \tag{8}$$

where $x$ is the value of the sample mean from the sample, and $s^2$ is the sample variance. The parameters of the gamma distribution were calculated according to formula (4).

**Table 2.** Parameters of the distribution of frequency of claims, based on data from Table 1

| Group | Average number of losses | Variance of number of losses | Parameters of negative binomial distribution | | Parameters of Gamma distribution | |
|---|---|---|---|---|---|---|
| | | | $p$ | $k$ | $\alpha$ | $\beta$ |
| I | 0.207 | 0.209 | 0.98 | 16.95 | 16.95 | 81.88 |
| II | 0.19 | 0.2 | 0.95 | 3.61 | 3.61 | 19 |
| Whole portfolio | 0.2 | 0.21 | 0.952 | 3.96 | 3.96 | 19.8 |

**Source:** Own research.

Basing on data from Table 2 the theoretical amounts of the number of losses were calculated. Results are presented in Tables 3–5.

**Table 3.** Theoretical and empirical amounts for **group I** of insurance portfolio

| No. of claims | Empirical amounts | Theoretical amounts of Poisson distribution | Theoretical amounts of negative binomial distribution |
|---|---|---|---|
| 0 | 2 907 | 2 903 | 2 962 |
| 1 | 592 | 600 | 54 |
| 2 | 66 | 63 | 75 |
| 3 | 5 | 4 | 9 |
| 4 | 0 | 0 | 0 |
| Whole | 3 570 | 3 570 | 3 570 |

**Source:** Own calculations.

**Table 4.** Empirical and theoretical amounts for **group II** of insurance portfolio

| No. of claims | Empirical amounts | Theoretical amounts of Poisson distribution | Theoretical amounts of negative binomial distribution |
|---|---|---|---|
| 0 | 10 221 | 10 169 | 10 217 |
| 1 | 1 843 | 1 932 | 1 844 |
| 2 | 210 | 183 | 212 |
| 3 | 18 | 12 | 19 |
| 4 | 5 | 1 | 5 |
| Whole | 12 297 | 12 297 | 12 297 |

**Source:** Own calculations.

**Table 5.** Empirical and theoretical amounts for the whole insurance portfolio

| No. of claims | Empirical amounts | Theoretical amounts of Poisson distribution | Theoretical amounts of negative binomial distribution |
|---|---|---|---|
| 0 | 13 128 | 13 121 | 13 184 |
| 1 | 2 435 | 2 493 | 2 379 |
| 2 | 276 | 236 | 275 |
| 3 | 23 | 15 | 25 |
| 4 | 5 | 2 | 3 |
| Whole | 15 867 | 15 867 | 15 867 |

**Source:** Own calculations.

**For the group I.** We shall verify the null hypothesis, $H_0$, that the distribution of the number of losses in group I is the Poisson distribution against alternative hypothesis, $H_1$, that the distribution of number of losses in group I is not the Poisson distribution. The value of statistics $\chi^2 = 0.5049$. For $\alpha = 0.05$ there is no ground for rejecting the null hypothesis.

We shall verify the null hypothesis, $H_0$, that the distribution of the number of losses in group I is the negative binomial distribution against alternative hypothesis, $H_1$, that the distribution of the number of losses in group I is not the negative binomial distribution. The value of statistics $\chi^2 = 11.34$. For $\alpha = 0.05$ there is no ground for rejecting the null hypothesis.

Hence, one may assume that the distribution of the number of losses in group I is homogenous.

**For the group II.** We shall verify the null hypothesis, $H_0$, that the distribution of number of losses in group II is the Poisson distribution against alternative hypothesis, $H_1$, that distribution of number of losses in group II is not the Poisson distribution. The value of statistics $\chi^2 = 27.349$. For $\alpha = 0.05$ we reject the null hypothesis to the advantage of the alternative hypothesis.

We shall verify the null hypothesis $H_0$, that the distribution of losses in group II is the negative binomial distribution against alternative hypothesis, $H_1$, that distribution of number of losses in group II is not the negative binomial distribution. The value of statistics $\chi^2 = 0.07$. For $\alpha = 0.05$ there is no ground for rejecting the null hypothesis.

Hence, one may assume that the distribution of number of losses in group II is not homogenous.

**For the whole portfolio.** We shall verify the null hypothesis, $H_0$, that the distribution of the number of losses in portfolio is the Poisson distribution against alternative hypothesis, $H_1$, that distribution of number of losses in portfolio is not the Poisson distribution. The value of statistics $\chi^2 = 16.899$. For $\alpha = 0.05$ there is no ground for rejecting the null hypothesis.

We shall verify the null hypothesis, $H_0$, that the distribution of the number of losses in portfolio is the negative binomial distribution against alternative hypothesis, $H_1$, that distribution of number of losses in portfolio is not the negative binomial distribution. The value of statistics $\chi^2 = 3.053$. For $\alpha = 0.5$ there is no ground for rejecting the null hypothesis.

Hence, one may assume that the distribution of the number of losses in the total portfolio is not homogenous.

For the purpose of results comparison let us apply the graphical method of insurance portfolio valuation that was proposed by Hossack (1999).

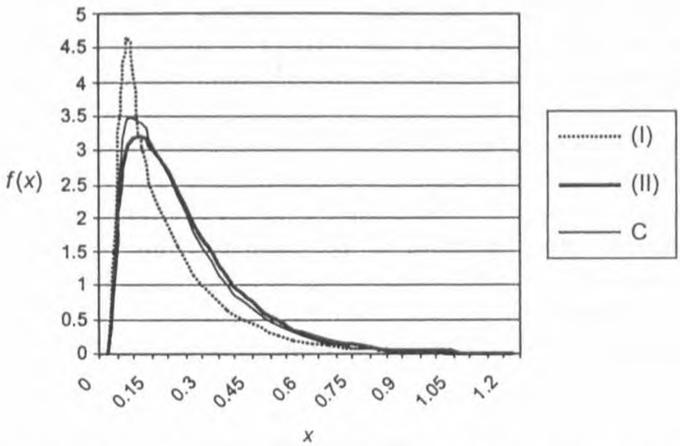


**Figure 1.** Density function of the gamma distribution in two groups of portfolio and in the whole portfolio (see data from Table 1)
**Source:** Own calculations

On the basis of the graphs obtained, group I is homogenous (the graph of density function is high and slender with small standard deviation) and group II and the whole portfolio are not homogenous. The biggest non-homogeneity occurs in group II.

## V. SUMMARY

The results obtained with the graphical method confirm the conclusions drawn from the application of the chi-square goodness-of-fit test. However, both graphical method and chi-square test for conformity, require large amount of data to estimate the parameters of the negative binomial distribution. Grouping the portfolio according to large number of factors may cause a reduction of the number of observations in particular groups.

In consequence, it can preclude the application of these methods. Moreover, practice often shows that both for the Poisson distribution and for the negative binomial distribution there is no ground for rejecting the null hypothesis of conformity of the number of losses in portfolio in comparison with examined distribution. Some cases were also noticed when the chi-square goodness-of-fit test rejects the null hypothesis both for the Poisson distribution and for the negative binomial distribution. In those cases, searching for the form of distribution of number of losses seems to be reasonable. Then, the foregoing methods can not be applied.

One of the methods of portfolio homogeneity examination is variance analysis (Domański 2001). However, application of this method needs the assumption that distributions of the number of losses for particular policies are normal, with equal variances. In such case the test statistics has the $F$-Snedecor distribution.

The issue of homogeneity assessment of portfolio groups requires further researches, because homogeneity is the fundamental assumption in estimation of future losses, and – in consequence – in calculations of insurance rates.

Searching for other methods of portfolio division into tariff groups to achieve homogeneity of tariff classes may be a solution. In practice, however insurance companies do not look for such methods, despite the fact that losses in automobile insurances confirm incorrect construction of portfolios.

## REFERENCES

Domański Cz. (red.) (2000), *Nieklasyczne metody statystyczne*, PWE, Warszawa.

Domański Cz. (red.) (2001), *Metody statystyczne*, Wyd. UŁ, Łódź.

Hossack I.B., Pollard J.H., Zehnwirth B. (1999), *Introductory Statistics with Applications in General Insurance*, Cambridge.

Niemiro W. (1997), *Rachunek prawdopodobieństwa*, WNE UW, Warszawa.

*Anna Szymańska*

## ZASTOSOWANIE TESTÓW STATYSTYCZNYCH DO BADANIA JEDNORODNOŚCI PORTFELA UBEZPIECZEŃ

Streszczenie

Podstawą prawidłowego funkcjonowania towarzystwa ubezpieczeniowego jest odpowiednie dopasowanie wysokości składek do poziomu ryzyka, jakie reprezentują ubezpieczani. Ubezpieczyciel najczęściej grupuje kontrakty ubezpieczeniowe w portfele charakteryzujące się zbliżonym poziomem ryzyka.

Istnieją jednak czynniki bezpośrednio nieobserwowalne, wpływające na wielkość i częstość szkód. Dlatego istotnym zagadnieniem jest ocena jednorodności portfela ubezpieczeniowego.

Celem referatu jest ocena wybranych metod, służących do sprawdzania jednorodności portfeli ubezpieczeniowych na przykładzie danych ubezpieczeń komunikacyjnych.