Problems of Demand and Effects in Portfolio Based Selection of Utility Functions

Riaman, Sukono, Eman Lesmana
Department of Mathematics, Faculty of Mathematics and Natural Sciences,
Universitas Padjadjaran, Indonesia
riaman@unpad.ac.id, sukono@unpad.ac.id, man.msie@gmail.com, marbunellen@gmail.com

Abdul Talib Bon
[3]Department of Production and Operations,
University Tun Hussein Onn Malaysia, Malaysia
talibon@gmail.com

Abstract

Stochastic rules and their inequality are very useful tools in various economic fields and financial problems. The purpose of this paper is to determine the main results obtained from the use of stochastic rules on financial optimization. Emphasis is placed on demand problems and the effects of changes in portfolio selection problems. Some other examples are not directly related to the optimization problem, this is given to show a broad view on the application of stochastic rules in financial matters. Stochastic orders and their inequalities have been used in various fields on problems related to opportunities (probability) and statistics in general. For example, a queuing system with a single service with Increasing Failure Rate (IRF) service time and service discipline that prioritizes job savings. Suppose T_i states the ergodic arrival time of a customer when service discipline is used.

Keywords:
Net premium, Forward Selection, Number of Claims, Generalized Linear Models

1. Introduction
Stochastic orders and their inequalities have been used in various fields on problems related to opportunities (probability) and statistics in general. For example, a queuing system with a single service with Increasing Failure Rate (IRF) service time and service discipline that prioritizes job savings. Suppose T_i states the ergodic arrival time of a customer when service discipline is used.

In the financial sector, the future price (or rate of return) of financial assets is shown by random variables that are uncertain. Therefore, a tool is needed to solve the optimization problem which cannot be an exception for the field, where many of them use stochastic rules. Likewise in many portfolio selection issues many use stochastic rules.

Thus, stochastic rules are very useful tools in the problem of portfolio selection. Investors are considered as a utility function for u, that want to invest wealth into one or two assets or even more. In the case of two assets, call A and B, take X and Y as random variables that show the results of assets A and B.

Furthermore, according to the principle of utility expectations, these investors overestimate asset A than B, if and only if

$E[u(X)] \geq E[u(Y)]$	1

The above inequality shows, that for all investors, the utility functions are included in several function classes, which are part of stochastic rules for X and Y.

The purpose of this paper is to introduce stochastic rules on economic problems, namely more specifically on the problem of demand and the problem of the effects of changes in portfolio selection. The next section discusses the

problem of formal portfolio selection, demand problems, and their comparative applications regarding market equilibrium price systems.

2. Methodology
The object to be studied in this research is the net premium of private car insurance. Dependent variables imposed on the object of this study there are two kinds, namely the variable number of claims and variable magnitude of claims.

2.1 Stacked Linear Model
The Stacked Linear Model is the development of a classical linear model that accommodates two main things, namely the distribution of non-normal bound variables and the transformation for linearity. The Stacked Linear Model has three components in it, namely random components, systematic components, and link functions. The random component of the Linear Model consists of identification of the dependent variable Y and the selection of the probability distribution. Variables bound to Linear Model Followed by exponential family distribution. The systematic component of the Linear Model Stamped determines the independent variable by including linearity as the equation estimator of the model. The link function is a connecting function between the systematic component and the expected value (average) of the random component. In the modeling of insurance with independent variables that are categorical, the link function used is a log link so the model equation as follows. [7], [9]:

$$\ln(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_P X_{ip} \tag{1}$$

$$\mu_i = \exp(X_i^T \boldsymbol{\beta}) \tag{2}$$

where $X_i^T = \begin{bmatrix} 1 & x_{i1} & x_{i2} & \dots & x_{ip} \end{bmatrix}$ and $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_{i1} & \beta_{i2} & \dots & \beta_{ip} \end{bmatrix}^T$.

2.2 Kolmogorov-Smirnov Test
The test of the distribution used in this study is Kolmogorov-Smirnov test of one sample. The null hypothesis of this test is that variable Y follows the selected theoretical distribution, in contrast to the alternative hypothesis. Reject the null hypothesis if the value $D_h = maks|F_0(x) - F^*(x)| \geq D_{\alpha,df=n}$ or $pvalue \leq \alpha$. [9].

2.3 Forward Selection
The forward step regression process begins by regressing the variable magnitude of the claim with each independent variable separately (individual regression). The insignificant independent variables will be set aside, while the significant independent variables are sorted by AIC values for each model. Further, the independent variables in the model with the smallest AIC values are regressed simultaneously with the independent variables in the model with the second smallest AIC value. The insignificant independent variable is set aside, followed by entering the independent variable from the model with the next smallest AIC value. And so on until there are no more variables left. [9].

2.4 Likelihood Ratio Test
This test is conducted to determine the significance of independent variables contained in a model. The null hypothesis of this test is that no parameter contributes to the dependent variable, in contrast to the alternative hypothesis. Reject the null hypothesis if the value of $G_h = -2\ln(\hat{\ell} - \breve{\ell}) \geq \chi^2_{\alpha,df=p-1}$ or $pvalue \leq \alpha$.

2.5 Akaike's Information Criterion (*AIC*)
The selection of the best model in this research is done by considering the value of AIC owned by model every time forward selection stage is done. The value of AIC is obtained through the following equation.

$$AIC = -2\ell + 2p \tag{3}$$

where ℓ is the value of the log-likelihood function of the model formed and p is the number of parameters used in the model. Smaller *AIC* values indicate that the model is getting better.

2.6 Maximum Likelihood Estimation (MLE) Method

To perform the β parameter estimation MLE method is used. Let Y_1, Y_2, ..., Y_n a random sample having a probability density function derived from the Poisson distribution. The maximum likelihood estimator for the β parameter is expressed by $\hat{\boldsymbol{\beta}}$ which is the completion of the first derivative of the likelihood function.

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} p(y_i) = \prod_{i=1}^{n} \frac{\exp(-\mu_i)(\mu_i)^{y_i}}{y_i!} \tag{4}$$

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} [-\exp(X_i^T \boldsymbol{\beta}) + y_i X_i^T \boldsymbol{\beta} - \ln(y_i!)] \tag{5}$$

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = \sum_{i=1}^{n} [-X_i \exp(X_i^T \boldsymbol{\beta}) + y_i X_i] = 0 \tag{6}$$

Because the function in equation 6 is implicit then to get the solution used a numerical iteration procedure, one of which is the Newton-Raphson Method.

2.7 Wald Test

Unlike the Likelihood Ratio test that can test the significance of several variables simultaneously, this test can only be used to test the significance of a single independent variable only. The initial hypothesis of this test is that the independent variable with the β_j parameter does not contribute to the dependent variable, in contrast to the alternative hypothesis. Reject the initial hypothesis if the value of $z_h \geq \chi^2_{\alpha,df=1}$ or $pvalue \leq \alpha$.

2.8 Collective Model

Let X be a random variable denoting the magnitude of the claim and the random variable N denoting the number of claims. To determine the net premium on the collective model can be used the following formula.

$$\text{Net premium} = E[S] = E[N] . E[X] \tag{7}$$

3. Illustration Analysis

The study used secondary data that is car insurance claim data approved by ABC company during 2017 calendar year. The data of claims obtained are grouped into two, namely data for model amount of claim and data for model number of claims.

3.1 Selection of Distribution

The data used in determining the model magnitude of claims is the first group of data. Within this first group there is the dependent variable "Claim Amount" and the variables to be used as independent variables, namely "Car Brand", "Car Year", "Territory", "Cause of Loss", and "Nature of Losses". Each of these independent variables is a categorical variable. The results of Kolmogorov test calculations obtained, shows Gaussian Inversion distributed data. Dependent variable magnitude claims have met the assumption of the Linear Model Stamped because Gaussian Inverse is a member of the exponential family distribution. The link function that is used, is the log link function.

3.1.1 Forward Selection

Linear Model Regression Process Regarded in SAS program using GENMOD procedure. The order of independent variables of the model with the smallest to largest AIC is "Cause of Loss", "Territory", "Car Brand", "Car Year", and "Nature of Losses". That is, the next regression involves the variables of "Claim Amount", "Cause of Loss", and "Territory". The result variable "Cause Loss" is significant, while variable "Territory" is not significant. The next regression involves "Cause Losses" and "Car Brand". The result of the "Car Brand" makes no significant contribution to the "Claim Amount". The "Cause Loss" variable can still be incorporated into the next process. Both the "Cause Loss" and "Year of Car" variables provide significantly. Finally, the "Causes of Loss", "Car Year", and "Nature of Losses" variables are simultaneously simultaneous. both the "Cause Loss" and "Car Year" variables are significant, but the "Nature of Losses" is not. Thus, the "Disadvantages" are set aside from the model. In addition, this model has an AIC smaller than the previous AIC model value. That is, this model is the best model for the magnitude of claims. After obtaining the best model for personal car insurance claims using forward selection, systematic system component parameters were estimated using the Maximum Likelihood Estimator (MLE) method with the help of the SAS program. Significant variables are only "Cause Losses" level 1 and 3, as well as "Year Car" level 12. Therefore, the model of magnitude of the claim can be written as follows.

$$g(\widehat{\mu_1}) = \ln\widehat{\mu_1} = 15{,}2983 - 0{,}6854x_{P1} - 0{,}7116x_{P3} - 0{,}8413x_{T12}$$

where

$\widehat{\mu_1}$: expectations of estimates of the magnitude of private car insurance claims

x_{P1} : variable "Cause Loss" category "hit"

x_{P3} : the "Cause Loss" variable of the category "Dispelled"

x_{T12} : variable "Year Car" category "2015"

3.2 Number of Claims Model

The data used in determining the number of claims model is the second group data. Within this second group there is the dependent variable "Number of Claims" and the variables that will be used as independent variables, namely "Gender", "Car Brand", "Car Year", "CC Car", "Transmission Type" "Territory". Each of these independent variables is a categorical variable. In the insurance field, the policyholder may not file a single claim, or apply only once, or more than once in a given period. Therefore, the variable expressing the number of claims follows one of the discrete distributions. The discrete distributions that belong to the exponential family distribution are Binomial distribution, Poisson distribution, and Negative Binomial distribution. With Microsoft Excel calculated the mean and variance of the variable number of claims. The results show that the mean of this variable is 0.645390071, while the variance is 0.379142374. From these results it can be seen that the mean is greater than the variance, whereas the mean and variance of the Poisson distribution are equal. It is very common to encounter linear modeling for the model of many claims with the Poisson distribution [1]. In SAS, this can be resolved by scale = p in the GENMOD procedure. The link function used is the log link.

3.2.1 Forward Selection

The order of the free variables from the model with the smallest to largest AIC is "Territory", "CC Car", "Transmission Type", "Year of the Car", and "Car Brand". That is, the next regression involves variables "Number of Claims", "Territories", and "CC Car". the two independent variables in this model give meaning to "Number of Claims". In addition, the AIC value of this model is smaller than the AIC model value that involves the "Territory" free variables only. In other words, this model is more suitable than the previous model. Next, the "Transmission Type" variable is inserted into the model to be reordered together with "Number of Claims", "Territories" and "Car Cc". "Transmission Type" must be set aside, while the other two variables in the model indicate a significant effect on the variable "Number of Claims". However, the AIC value of this model is not smaller than the AIC model value involving "Territory" and "Car cc". Therefore, the previous model is still the best model to date. Furthermore, "Year of the Car" is involved in the regression process of "Territory" and "CC Mobil". "Car Year" is not significant. In addition, this model has an AIC value that is higher than the AIC model value that involves "Territory" and "Car cc" only. Therefore, this model is less suitable. Thus, the best model for modeling the number of claims is a model involving only the "Territory" and "cc Car" independent variables. Furthermore, systematic component component parameters are estimated using the Maximum Likelihood Estimator (MLE) method. Significant variables are only category 2 "Territory" and "cc Car" category 2, 4, 6, 7. Therefore, the model number of claims can be written as follows.

$$g(\widehat{\mu_2}) = \ln\widehat{\mu_2} = -0{,}7893 - 0{,}7055x_{W2} + 0{,}4724x_{C2} + 0{,}8823x_{C4} + 0{,}6456x_{C6} + 0{,}6619x_{C7}$$

Where :

$\widehat{\mu_2}$: expectations of estimates of the number of private car insurance claims

x_{W2} : variable "Territory" category "Other Territory"

x_{C2} : variable "cc Car" category "1200cc"

x_{C4} : variable "cc Car" category "1500cc"

x_{C6} : variable "cc Car" category "1800cc"

x_{C7} : variable "cc Car" category "2000cc"

3.3 Estimation of Net Premium Based on Model

Based on each model is obtained expectation of the amount of claims and the number of claims, as follows.

1) Model expectations of personal car insurance claims
$$\widehat{\mu_1} = \exp(15{,}2983 - 0{,}6854x_{P1} - 0{,}7116x_{P3} - 0{,}8413x_{T12})$$

2) Expectation model of the number of private car insurance claims
$$\widehat{\mu_2} = \exp(-0{,}7893 - 0{,}7055x_{W2} + 0{,}4724x_{C2} + 0{,}8823x_{C4} + 0{,}6456x_{C6} + 0{,}6619x_{C7})$$

Case 1

Suppose a man with the initials A resides in Surabaya insured Honda car in 2013 with 1800cc engine capacity at *XYZ* company. The car was hit so he filed a claim to the company. Then the net premium assessment is as follows.

$$\widehat{\mu_1} = \exp(15{,}2983 - 0{,}6854x_{P1} - 0{,}7116x_{P3} - 0{,}8413x_{T12})$$
$$\widehat{\mu_1} = \exp\{15{,}2983 - 0{,}6854(1) - 0{,}7116(0) - 0{,}8413(0)\}$$
$$\widehat{\mu_1} = \exp(14{,}6129)$$
$$\widehat{\mu_1} = 2219738{,}60$$
$$\widehat{\mu_2} = \exp(-0{,}7893 - 0{,}7055x_{W2} + 0{,}4724x_{C2} + 0{,}8823x_{C4} + 0{,}6456x_{C6} + 0{,}6619x_{C7})$$
$$\widehat{\mu_2} = \exp\{-0{,}7893 - 0{,}7055(1) + 0{,}4724(0) + 0{,}8823(0) + 0{,}6456(1) + 0{,}6619(0)\}$$
$$\widehat{\mu_2} = \exp(-0{,}8492)$$
$$\widehat{\mu_2} = 0{,}427$$

$$Net\ Premiums = \widehat{\mu_1} \times \widehat{\mu_2} = 2219738{,}60 \times 0{,}427 = 947828{,}38$$

Thus, the net premium to be paid by insurance participants in this case is IDR 947.828,00.

Case 2

Suppose a woman with the initials B residing in Bandung insured the Toyota car manual transmission in 2015 with a capacity of 1200cc ABC company. His car was hissed so he filed a claim to the company. Then the net premium assessment is as follows.

$$\widehat{\mu_1} = \exp(15{,}2983 - 0{,}6854x_{P1} - 0{,}7116x_{P3} - 0{,}8413x_{T12})$$
$$\widehat{\mu_1} = \exp\{15{,}2983 - 0{,}6854(0) - 0{,}7116(1) - 0{,}8413(1)\}$$
$$\widehat{\mu_1} = \exp(13{,}7454)$$
$$\widehat{\mu_1} = 932290{,}74$$
$$\widehat{\mu_2} = \exp(-0{,}7893 - 0{,}7055x_{W2} + 0{,}4724x_{C2} + 0{,}8823x_{C4} + 0{,}6456x_{C6} + 0{,}6619x_{C7})$$
$$\widehat{\mu_2} = \exp\{-0{,}7893 - 0{,}7055(0) + 0{,}4724(1) + 0{,}8823(0) + 0{,}6456(0) + 0{,}6619(0)\}$$
$$\widehat{\mu_2} = \exp(-0{,}3169)$$
$$\widehat{\mu_2} = 0{,}728$$
$$Nett\ Premiums = \widehat{\mu_1} \times \widehat{\mu_2} = 932290{,}74 \times 0{,}728 = 678707{,}65$$

Thus, the net premium to be paid by insurance participants in this case is IDR 678,707.65. Keep in mind that two examples of such cases are some of the possible cases that are expected to occur within one calendar year of insurance. In both cases researchers suggested the company to apply a net premium of IDR 947,282.00 to avoid future losses.

4. Conclusions

Based on the analysis and discussion that has been done then it can be concluded that: The modeling in this study begins with the selection of distribution and link functions of each dependent variable. Furthermore, the regression process is done by forward selection method. The tests used to test the significance of the independent variables are the Likelihood Ratio test and the Wald test. The best model selection is done by looking at the value of AIC owned by model in every regression process. The independent variables that contribute significantly to the variable magnitude of claims are the cause of the loss and the year of the car, while for the number of claims are the region and the car cc. After going

through these processes, the model for the number of private car insurance claims, namely $\ln \widehat{\mu_1} = 15{,}2983 - 0{,}6854 x_{P1} - 0{,}7116 x_{P3} - 0{,}8413 x_{T12}$, while the model for the number of private car insurance claims is $\ln \widehat{\mu_2} = -0{,}7893 - 0{,}7055 x_{W2} + 0{,}4724 x_{C2} + 0{,}8823 x_{C4} + 0{,}6456 x_{C6} + 0{,}6619 x_{C7}$.

Based on the two case studies in this study, the estimated net premium of ABC corporate private car insurance is IDR 947.282,00. This price is the highest price among the premiums generated through case studies. This price is chosen to anticipate future company losses.

Acknowledgments

References

Agresti, A. 2007. An Introduction to Categorial Data Analysis (2nd ed.). New Jersey: John Wiley & Sons, Inc.

David, M., Jemna, D. 2015. Modeling The Frequency of Auto Insurance Claims by Means of Poisson and Negative Binomial Models. Scientific Annals of the "Alexandru Ioan Cazu" University of Iaşi Economic Sciences, (Online), Vol. 62(2), hlm. 151-167, (http://www.degruyter.com, diakses 10 Juli 2017).

Department of Statistics Online Programs. 2017. Lesson 7.2.2 - Overdispersion, (Online), http://onlinecourses.science.psu.edu/stat504, diakses 8 Juni 2017).

Goldburd, M., Khare, A., dan Tevet, D. 2016. Generalized Linear Models for Insurance Rating. Virginia: Casualty Actuarial Society.

Hogg, R. V. dan Craig, A. T. 1995. Introduction to Mathematical Statistics (5th ed.). New Jersey: Prentice Hall.

Hogg, R. V. dan Craig, A. T. 2005. Introduction to Mathematical Statistics (6th ed.). Delhi: Dorling Kindersley (India) Pvt. Ltd.

Iskandar, K., Fuad, N., Wirasadi, F., dan Sendra, K. 2011. Dasar-dasar Asuransi: Jiwa, Kesehatan, dan Anuitas. Jakarta: Asosiasi Ahli Manajemen Asuransi Indonesia.

Klugman, S. A. 2004. Loss Models: From Data to Decisions (2nd ed.). New Jersey: John Wiley & Sons, Inc.

McCullagh, P. dan Nelder, J.A. 1989. Generalized Linear Models in Monograph on Statistics and Applied Probability (2nd ed.). New York: Chapman and Hall.

Meyricke, R. dan Sherris, M. 2013. The Determinants of Mortality Heterogeneity and Implications for Pricing Annuities. Insurance: Mathematics and Economics, (Online), Vol. 53, hlm. 379-387, (http://sciencedirect.com, diakses 10 September 2016).

Sari, D. P. 2016. Prediksi Premi Murni dengan Generalized Linear Models (GLM) pada Asuransi Kendaraan Bermotor. Skripsi tidak diterbitkan. Jatinangor: Universitas Padjadjaran.

Siegel, S. dan Castellan, N. J. 1988. Nonparemetric Statistical for the Behavioral Sciences. New York: McGraw-Hill.

Wood, S. N. 2006. Generalized Additive Models. Boca Raton: Chapman and Hall.

Biographies

Riaman is the staff of the Department of Mathematics, University of Padjadjaran, with the field of research are : actuarial mathematics, financial mathematics, survival model analysis and reliability.