Application of the Ordered Lorenz Curve in the Analysis of a Non-Life Insurance Portfolio

FREDRIK AHLBERG
Application of the Ordered Lorenz Curve in the Analysis of a Non-Life Insurance Portfolio

FREDRIK AHLBERG
Insurance analysts have a great variety of assessment tools at their disposal in order to ensure a healthy insurance portfolio. To describe the financial income and loss distribution of the insurance portfolio one of the more fundamental mathematical instrument is the Lorenz curve. A measure developed in the early 19th century by Max O. Lorenz which intended to describe a population’s income distribution in a macro perspective. By developing further on this method with guidance from the article by Frees, Meyers and Cummings, [5], a link between the Lorenz curve and the insurance portfolio’s risk segment will be investigated.

By constructing an insurance rating function which determine an insurance expected loss, depending on the policyholders characteristics, ordering the premium and loss distributions by its relative loss the intent is to identify profitable blocks along the ordered Lorenz curve. With this insight an analyst can redefine the portfolio structure and highlight the desirable characteristics which define a policyholder. In order to keep up with the competition an insurer has to, in the long run, create a sustainable, profitable portfolio with lowering the risk of occurring greater insurance claims.
SAMMANFATTNING


Med hjälp av klassisk regression kan försäkringarna värderas efter deras potentiella förlust, premie och förlust fördelningen sorteras efter en relativ förlust, för att ta fram en rankad Lorenzkurva. Detta för att koppla den underliggande informationen från försäkringstagaren till dess relativa finansiella prestation. Att hitta segment i portföljen som visar potentiellt större lönsamhet och lägre risk är något en analytiker ständigt bör eftersträva för att hålla jämna steg med konkurrenserna på markanden och utöka sin portfölj med en strategi som bygger på väl underliggande teori.
ACKNOWLEDGEMENTS

First of all, I wish to recognize my supervisor at Söderberg & Partners, Robert Karlsson for being utterly most helpful and encouraging throughout this process. A special thanks to all the colleagues at Söderberg & Partners at the non-life department who has been a great support along this journey and asking genuinely whenever this thesis is available for reading. All of you have also showed what a great work place this is and I am thrilled to continue working along side all of you. Lastly I wish to thank my supervisor Professor Boualem Djehiche of the department of Mathematics at KTH. For the guidance and support throughout this thesis I cannot express my gratitude enough.
# Table of Contents

1 Introduction .......................................................... 1  
   1.1 Background ......................................................... 1  
   1.2 Formulating the Problem .......................................... 2  
   1.3 Restrictions ....................................................... 2  
   1.4 Outline ............................................................ 3  

2 Theory ................................................................. 5  
   2.1 The Generalized Lorenz Curve .................................... 5  
   2.2 Insurance Scoring .................................................. 6  
      2.2.1 Characteristics ............................................... 6  
      2.2.2 Insurance Score .............................................. 7  
      2.2.3 Linear Regression of Insurance Score ....................... 8  
   2.3 Relativity .......................................................... 9  
   2.4 Introducing the Ordered Lorenz Curve ......................... 9  
   2.5 The Gini Index .................................................... 11  

3 The Order Lorenz Curve Method ...................................... 13  
   3.1 Organizing the Data .............................................. 13  
   3.2 The Classic Lorenz Curve Method ............................... 14  
   3.3 Sorting Claims and Premiums .................................... 15  
   3.4 Constructing the Ordered Relative Distributions ............. 16  
   3.5 Calculate the Gini Index ........................................ 17  

4 Results .................................................................. 19  
   4.1 Results of Portfolio A ............................................. 19  
   4.2 Results of Portfolio B ............................................. 21  
   4.3 Results of the Non-Claim Insurances in Portfolio A .......... 22  

5 Discussion ............................................................ 23  
   5.1 Discussing the Method and its Improvements ................. 23
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2 The Scoring Function</td>
<td>24</td>
</tr>
<tr>
<td>Bibliography</td>
<td>27</td>
</tr>
<tr>
<td>A Appendix A</td>
<td>29</td>
</tr>
</tbody>
</table>
1.1 Background

The insurance market is a competitive business and as the counseling and mediatory business gains more attention from customer the importance of understanding the foundation behind selecting and analyze a profitable yet fair insurance portfolio is essential. In today’s high speed driven market, access to the immense amount of data produced is not considered an obstacle, rather the reckoning and understanding of insightful algorithms can yield information of value for analysts and executives. In a traditional setting, the insurance portfolio contains a group of policyholders, in which each pays a premium to cover the potential future losses. The actuary, which determine the price of the insurance contract, has often a set of characteristics available from the client seeking an insurance coverage. Since the potential loss on this contract is stochastic and not deterministic the outcome is a random process. Through statistical methods, primarily performed on historical data, the loss determination and premiums will depend on these given characteristics. For an insurance analyst, the focus would be to investigate the relations between the premium and loss distributions in order to derive information about the profitability or vulnerability of this portfolio.

A classic tool, developed in the early 19th century for studying well-fare inequality [2], is the Lorenz curve. The Lorenz curve features the application of comparing two distributions’ relation to one another and hence proven useful for economist in a variety of other applications than just well-fare statistics. Several prominent mathematicians have developed insightful insurance portfolio theory regarding the Lorenz curve and a numerous of publications can be connected to the Lorenz curve. Onward in this report, the extended theory about the ordered Lorenz curve will be presented supported by the previous work of Edward Frees, Glenn Meyers
and David Cummings and their publications [6] and [5] in the *Journal of Risk and Insurance*, *vol. 81* 2013 and *Journal of the American Statistical Association*, *Vol 106*, 2011. The results from these findings suggests an alternative way to highlight both profitable and vulnerable blocks in an insurance portfolio when observing the corresponding premium and loss distribution functions. The assessment and comparison of the performance of an insurance portfolio is read through the eye’s of the Gini methodology [10], which is another well established concept with distinguishing connections to the Lorenz curve.

### 1.2 Formulating the Problem

The main purpose with this project is to investigate a standard procedure for analyzing a non-life insurance portfolios premium distribution in relation to its corresponding loss distribution. As of today there is no generalized method to analyze the premium distribution based on the policyholder’s characteristics, see 2.2.1, and give an intuitively visual overlook of the Insurance portfolio as a whole. The objective is to derive a procedure in which it is easy to describe a portfolios status, determine a rate based on the policyholders attributes and visually highlight profitable policyholders. In order of priority the project aims to describe an algorithm which hold for following properties.

- The chosen method refers to valid statistical tools supported by a modern bibliography.
- The algorithm should be supported with relevant and retrievable data.
- The theory should be well stipulated an easy for the reader to follow.

### 1.3 Restrictions

Now, since it is often common that all information necessary to perform a complex and more thorough algorithm the restrictions of a strategic yet simple algorithm are essential to stipulate. The Lorenz Curve respond to the distribution of an insurance products losses and its premium holders. Thus it establishes that the data required is the losses and the premiums. The procedure of the algorithm should not strive in this project to comprehend other factors, which very well may be of influence to the distributions, but the following discussion of the algorithm will not take that into consideration.

The research for the algorithm and its effects are also limited to a few specific characteristics for each policy group, that in order to determine a score for each holder, see section 2.2. It is also to be taken into consideration that this project is aimed at non-life insurance policies and that no discussion whether it fits a universal premium policy analysis is of interest for this thesis.
1.4 Outline

The structure of the report contains a background of the insurance business and the role of an insurance analyst in today’s modern society. The following section will focus on the theory and introduce an explanatory background in which the method will be based on.

Building on the established theory the thesis turns to its main focus, which is the methodology of the algorithm process, in which the data will be described and translated into action. The results are then presented objectively in which a discussion upon that will follow. Except from analysing the result, the discussion will also contain further improvement suggestions for the method.
2.1 The Generalized Lorenz Curve

The main tool in this analysis is based on the Lorenz curve, derived by Max Otto Lorenz in 1905 [2]. The generalized Lorenz curve is visualize as a graph with the horizontal axis representing the proportion of the "population" and a corresponding distribution function on its vertical axis. Since it development financial mathematicians and economist has shown interest in it and been a useful tool especially in income, wealth and inequality measurement. The basic deduction of the Lorenz curve is based on the text in [7] and [10].

Let us denote the population and each member’s $i$ earning with a set of random variables $X_i$ where $i = 1, ..., n$. The proportion of the population earning less or equal to a sum $s$ is referred to as the cumulative distribution function $F_X(s)$, in other words, the probability that member $X_i$ earns less or equal to $s$. This is written as

$$F_X(s) = P(X \leq s).$$

If $p$ represent the proportion of $X$ that receives the income $s$, i.e. $p = F(X(p))$, we have that the inverse of the cumulative distribution is $X(p) = F^{-1}(p)$. The cumulative value of the variate $X(p)$ then represents the generalized Lorenz curve. Dividing the generalized Lorenz curve by the expected value of $X$ lead us to the Lorenz curve, given that $E[X] \neq 0$.

$$L(p) = \frac{1}{E[X]} \int_0^p F^{-1}(t) dt = \frac{1}{E[X]} \int_0^p X(t) dt.$$  

This establishes some important properties regarding the curve [10].
1. The curve will pass between the points (0,0) and (1,1).

2. Since the derivative of $L(p)$ is $X(p)$ at $p$, the curve is increasing when $X(p)$ is positive, decreasing for the non-positive $X(p)$.

Furthermore, within the graph there is a 45° straight line between (0,0) and (1,1). This line is referred to as the "Line of Equality" or "Line of Safe Asset" [10]. Since if the case is that our income is equal in distribution, the curve $L(p)$ would follow this line.

Figure 2.1: An example of a Lorenz Curve, source : [8]

2.2 Insurance Scoring

2.2.1 Characteristics

One of the fundamental steps for underwriters when it comes to assign scores or rates for the policyholders, are to consider the essential characteristics of the holders. The characteristics should be of such type that it can be measurable attributes corresponding to each policyholder and taken into consideration when determine the potential losses for the underwriter i.e. determine a premium to cover for such a risk.

Examples of such characteristics are:

- Credit score
- Prior loss information
- Turnover or revenue
- Years in business
- Earnings
- Number of employees
- Company structure
How one translate this information into structured data to fit the algorithm is one of the analyst decision and should be stated clearly to not present any delusional results. Further on in the report, a characteristic will be denoted \( c \) and a set of characteristics represented in vector form will be referred to \( \mathbf{c} \).

### 2.2.2 Insurance Score

For each policyholder \( i \) a set of characteristics are given as a vector \( \mathbf{c}_i = (c_1, c_2, \ldots, c_k)^T \). This set is desirable to use when scoring insurances belonging to the same portfolio. It is suggested by the article [5] that an appropriate method of assigning policyholder a rating is a regression analysis for each loss or risk given the set of characteristics. A suitable proposal is to take the expected value of the loss or risk given \( \mathbf{c}_i \), see equation 2.6.

For further reading we denote the loss or risk as \( y \) and a payment or premium as \( P \). The loss function of a written insurance contract is then said to be

\[
I(\mathbf{c}, y) = y(\mathbf{c}) - P(\mathbf{c}).
\]

For determine a score it is suggested to look at the expected loss associated with the policies corresponding with its set of characteristics and determined premium. Here an indicator function, denoted with \( I \) is introduced, where its value is 1 if the policy is within our selected set \( A \), 0 otherwise.

\[
I = \begin{cases} 
1, i \in A \\
0, i \notin A 
\end{cases}
\]

By following the steps in [5] the regression function can be defined.

\[
E[I(i \in A)l(\mathbf{c}_i, y_i)] = E[I(i \in A)(E[y_i|\mathbf{c}_i] - P(\mathbf{c}_i))]
\]

\[
S(\mathbf{c}) = E[y|\mathbf{c}].
\]

If the set \( A \) of policies follows the condition that

\[
E[y|\mathbf{c}] < P(\mathbf{c})
\]

it will ensure a negative expected loss [5], i.e. a profit. Note that within the article [5] by Frees, Meyers and Cummings the regression function is said to be of exponential type. However, this section’s purpose is to introduce the concept of regression analysis for a general method. For this intent a linear regression is sufficient for a simple scoring algorithm.
2.2.3 Linear Regression of Insurance Score

Consider the case above where the aim is to estimate a scoring function based on the characteristics described by \( c \). Here we follow the steps described in [4] for linear regression. The linear case is described by the set of coefficients \( \beta_0, ..., \beta_k \).

\[
S(c, \beta) = \beta_0 + \beta_1 c_1 + \beta_2 c_2 + ... \beta_k c_k.
\]

The coefficients in equation 2.8 are estimated with the help of an ordinary least square fitting method, where the goal is to minimize the squared error \( \epsilon \).

\[
\min \sum_{i=1}^{n} \epsilon_i^2 = \min \sum_{i=1}^{n} (S_i - \hat{S}_i)^2 = \min \sum_{i=1}^{n} (S_i - (\beta_0 + \beta_1 c_{i,1} + \beta_2 c_{i,2} + ... \beta_k c_{i,k}))^2.
\]

With the steps described in [4] a system of equations can be determined and solved for the coefficients. Step one is to partially differentiate equation 2.9 for each coefficient \( \beta_j \).

\[
2 \sum_{i=1}^{n} -c_{i,j}(S_i - \beta_0 - \beta_1 c_{i,1} - ... - \beta_k c_{i,k}) = 0.
\]

The expression is simplified by dividing both sides with \(-2\).

\[
\sum_{i=1}^{n} c_{i,j}(S_i - \beta_0 - \beta_1 c_{i,1} - ... - \beta_k c_{i,k}) = 0.
\]

This lead to a system of equation on the following form, beginning with the derivative with respect to \( \beta_0 \) and moving on up to \( \beta_k \).

\[
\sum_{i=1}^{n} (S_i - \beta_0 - \beta_1 c_{i,1} - ... - \beta_k c_{i,k}) = 0
\]

\[
\sum_{i=1}^{n} c_{i,1}(S_i - \beta_0 - \beta_1 c_{i,1} - ... - \beta_k c_{i,k}) = 0
\]

\[
\vdots
\]

\[
\sum_{i=1}^{n} c_{i,k}(S_i - \beta_0 - \beta_1 c_{i,1} - ... - \beta_k c_{i,k}) = 0.
\]

Written on matrix form the equation becomes

\[
c^T(S - c\beta) = 0
\]

and solved for the unknown coefficients

\[
\beta = (c^Tc)^{-1}c^TS.
\]

Visually the matrices are displayed as below. We let \( n \) correspond to number of insurances and \( k \) define the set of characteristics the insurances are influenced by.

\[
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_k
\end{pmatrix} = 
\begin{pmatrix}
1 & 1 & \ldots & 1 \\
1 & c_{1,1} & \ldots & c_{1,k} \\
\vdots & \vdots & \ddots & \vdots \\
1 & c_{n,1} & \ldots & c_{n,k}
\end{pmatrix}^{-1}
\begin{pmatrix}
1 & 1 & \ldots & 1 \\
1 & c_{1,1} & \ldots & c_{1,k} \\
\vdots & \vdots & \ddots & \vdots \\
1 & c_{n,1} & \ldots & c_{n,k}
\end{pmatrix}
\begin{pmatrix}
S_1 \\
S_2 \\
\vdots \\
S_n
\end{pmatrix}
\]
2.3 Relativity

In this section the role of the relativity is introduced. This is the variable we later on will use to order both our estimated premium distribution function and its corresponding loss distribution function. Let us begin by define the relativity function \( R \).

\[
R(c) = \frac{S(c)}{P(c)}
\]

Intuitively this can be seen as the relative return based on the scoring function. Recall that the scoring function \( S(c) \) is chosen such that condition 2.7 holds. That is in order to ensure a negative expected loss, hence a relative return. According to Frees, Meyers and Cummings’ article [6] the arguments for using relativity are stated as:

- The relative variable help discriminate between losses and premiums.
- With the sorting order operating through the relativity, unprofitable blocks can be identified within the Lorenz curve.
- Premium structure can be evaluated through different settings of the relativity variable.

2.4 Introducing the Ordered Lorenz Curve

The concept of the Lorenz curve was introduced in 2.1. In this section the method will be extended and with the usage of relativity as our sorting variable we introduce the ordered Lorenz curve. As mentioned in 2.1 a classic Lorenz curve shows the proportion of the population on the horizontal axis and the corresponding distribution of interest of the population on the vertical axis. According to Frees, Meyers and Cummings in [6] there are two ways the ordered Lorenz curve extends from the classical.

1. Ordering of losses and premiums by the relativity variable \( R(c) \).

2. Premiums are allowed to vary for each observation.

To motivate that the distribution functions for the premium and losses can be used and ordered a Lorenz curve, the derivation can be followed below. Beginning by stipulating the theoretical ordered Lorenz curve \( OL(p) \) according to lemma A1 in [5].

\[
OL(p) = F_L(F_P^{-1}(p)) = \frac{\mu_p}{\mu_y} \mu \int_0^p h(F_P^{-1}(z))dz.
\]

and

\[
h(z) = \frac{w_m(z)}{w_P(z)}.
\]
$w_m(z)$ and $w_P(z)$ are defined as weights in our distribution functions, see [5]. By following the steps below and with the set up to use the scoring function $S(c) = m(c)$, our regression function, one can show that 2.19 follows the properties of a Lorenz curve.

\[
(2.21) \quad OL(p) = \frac{\mu_p}{\mu_y} \int_0^p \frac{w_m(F_P^{-1}(z))}{w_P(F_P^{-1}(z))} \, dz = \begin{cases} 
F_P^{-1}(z) = s \\
P_F(F_P^{-1}(z)) = z = F_P(s) \\
dF_P^{-1}(z) = F_P^{-1}(s) \, dz = ds 
\end{cases} = \frac{\mu_p}{\mu_y} \int_0^p w_m(s) \, w_P(s) \, f_P(s) \, ds.
\]

Now, the fact that we use $S(c) = m(c)$ will provide us with

\[
w_m(s) = E[m(c) | R(c) = s] \\
= E[m(c) | S(c) / P(c) = s] \\
= E[m(c) | m(c) = sP(c)] \\
= E[sP(c) | m(c) = sP(c)] \\
= sE[P(c) | S(c) = sP(c)] \\
= sE[P(c) | S(c) / P(c) = s] \\
= sE[P(c) | R(c) = s] \\
= sw_P(s).
\]

With $w_m(s) = s \, w_P(s)$ inserted into 2.21 it follows that

\[
(2.22) \quad \frac{\mu_p}{\mu_y} \int_0^p w_m(s) \, f_P(s) \, ds = \frac{\mu_p}{\mu_y} \int_0^p s \, w_P(s) \, f_P(s) \, ds = \frac{\mu_p}{\mu_y} \int_0^p s \, f_P(s) \, ds = \frac{\mu_p}{\mu_y} \int_0^p F_P^{-1}(z) \, dz
\]

and we have that the $OL(p)$ can be written as a Lorenz curve on the form

\[
(2.23) \quad OL(p) = \frac{\mu_p}{\mu_y} \int_0^p F_P^{-1}(z) \, dz = L(F_P; p),
\]

thus the graph of $(F_P(s), F_L(s))$ ordered by their relativities can be seen as a Lorenz curve.
2.5 The Gini Index

As the ordered Lorenz curve provide a visually explanatory graph between the risk and premium distribution association, it is not sufficient enough to compare the profitability between different insurance portfolios. A more detailed measure is the Gini coefficient, or the Gini index when referring to the value in percentage, which is well known measure of ratio when it comes to comparing distributions related to the Lorenz curve. A Gini index of 0 % suggests an equal distribution in income versus people while 100% describes total inequality. As established in 2.1 the Lorenz curve moves along the horizontal axis with the increase in proportion of policyholders and in vertical direction the loss distribution associated with the proportion, the Gini index measures the ratio of the area under the "line of equality" and the Lorenz curve and the total area covered by the "line of equality". Visually, the shaded area in figure 2.2 can be interpreted as the Gini index. We may define this measure according to [5] as

\[
G(F_P,F_L) = \int_0^\infty F_P(s)dF_P(s) - \int_0^\infty F_L(s)dF_P(s) = \frac{1}{2} - \int_0^\infty F_L(s)dF_P(s) = 1 - \frac{1}{2} \int_0^\infty F_L(s)dF_P(s)
\]

Note that this measure may also be linked to different portfolios relative profitability. Since portfolios varies in size, structure and premium-loss distribution, the Gini index can act like a relative comparison tool in order to rank or distinguish performance between them. An important note one should be aware of when using the Gini index, is that it does not tell anything regarding the different portfolios equality. A top heavy portfolio could very well have the same Gini index as a bottom heavy portfolio. Furthermore, since the method is constructed through empirical observations the Gini index need as well to be computed as an estimate, depending on the data points given by the empirical determined functions \( \hat{F}_P \) and \( \hat{F}_L \).

Figure 2.2: Graphical representation of the Gini index.[1]
3.1 Organizing the Data

The method itself will be conducted on two different insurance products and their respective claim history. Further on the separated insurance portfolios will be referred to portfolio A and portfolio B. Now, each set of data follows the structure of a policy Id, premium, the loss amount and its policy’s characteristics. The insurance which is represented by portfolio A has its premiums based on the policyholders turnover and its chosen insurance amount where as insurance B is combined by several smaller insurance covers and therefore has ten different insurance amounts and the policyholder’s turnover as arguments for determine the premium. Our characteristic vectors are then seen as \( \mathbf{c}_A = (c_1, c_2) \) where \( c_1 \) corresponds to the turnover of the policyholder and \( c_2 \) the insurance amount. For \( \mathbf{c}_B = (c_1, \ldots, c_{10}) \) the first component is the turnover and the following are several insurance amounts. The portfolios’ premium and loss vectors are referred to as \( P(\mathbf{c}) \) and \( y(\mathbf{c}) \). The structure of the data can be seen in the appendix where the two tables, A.1 and A.2, contains the two portfolios. Note that the insurances selected are the ones in the portfolio with claim history.

In the second part of the method and later on brought up for discussion is the application of the method on the entire portfolio. We will use the entire data set of Portfolio A, including the non-claim insurances, and with the help of the scoring function investigate its ordered Lorenz curve as well. The entire portfolio is constructed by 1321 different insurances and their corresponding characteristics. Unfortunately could the entire set of the non-claim insurances with its characteristics from Portfolio B not be retrieved from any database so the final analysis will only be conducted on Portfolio A.
3.2 The Classic Lorenz Curve Method

This section's purpose is formulate the classic Lorenz curve in an insurance portfolio context. We study how the premium is distributed among the policyholders with only using the premium $p$ as our sorting variable, following the structure of $p_i < p_{i+1} < p_n$. The empirical distribution for the premium is calculated according to

$$(3.1) \hat{F}_{\text{Premium}}(p) = \frac{\sum_{i=1}^{n} P_i I(P_i \leq p)}{\sum_{i=1}^{n} P_i}.$$  

The classic Lorenz curve, "income versus people distribution", is then in terms of insurance business the premium distribution over the people. For the both portfolios the curve follows as below. Worth nothing is that portfolio B has a more evenly spread among their insurance premiums than portfolio A. Observing the lower 50th percentile of the policy holders, one sees that in Portfolio A they only cover almost 25% of the total premium while in Portfolio B it is closer to 40%.

Figure 3.1: Premium vs. Policyholder distribution of Portfolio A

Figure 3.2: Premium vs. Policyholder distribution of Portfolio B
3.3 Sorting Claims and Premiums

For each portfolio we let the claim be sorted as $y_i < y_{i+1} \ldots < y_n$ and determine the empirical distribution as

\[
\hat{F}_L(s) = \frac{\sum_{i=1}^{n} y_i I(y_i \leq s)}{\sum_{i=1}^{n} y_i}.
\]

combined with the sorted premiums in equation 3.1 we can then produce the loss and premium distribution in a graph for both portfolios.

![Lorenz Curve](image1.png)

Figure 3.3: Loss versus premium distribution of portfolio A

![Lorenz Curve](image2.png)

Figure 3.4: Loss versus premium distribution of portfolio B
3.4 Constructing the Ordered Relative Distributions

Begin by define the matrix which contains the characteristics of the insurance, \( c_A \) and \( c_B \). The matrix size correspond to the number of insurances included and the number of characteristics plus one. Remember that in 2.8 we also define \( \beta_0 \) and solve for it in 2.16, hence the characteristic matrix is extended with a column of ones, see 2.17. For the sake of not have to deal with large numbers which can affect the matrix calculations, we scale the characteristics by \( 10^6 \). Now, in portfolio A we have 16 insurances and 2 characteristics influencing the premium, for portfolio B the numbers are 14 and 10 making \( n_A = 16, k_A = 2 \) and \( n_B = 14, k_B = 10 \). Once the matrices are defined the next step is to perform the regression analysis and determine the regression coefficients \( \beta_0 \ldots \beta_k \). Before we can determine the coefficients we need to determine the score \( S \) based on our claims. As described in 2.6 we use the expected claim given the set of characteristics \( c \) as our base for scoring. Based on the data given from portfolio A and B we can calculate the expected claim as

\[
S(c) = E[y|c] = \frac{1}{n} \sum_{i=1}^{n} y_i I(c_i = c_j) \quad \forall j = 1 \ldots n.
\]

Once the score is determined the relativity, here interpreted as a relative loss, can be calculated according to 2.18. There are several ways to define the relative variable \( R \), but should be chosen such that it is clear to the user what the variable is intended to measure. In order to construct the ordered distributions for the premium and loss, the step variable \( r \) determine the distribution. Based on the relativity and its value from the characteristics, the empirical distribution functions \( \hat{F}_p(r) \) and \( \hat{F}_L(r) \) are now sorted on the same variable \( R(c) \).

\[
\hat{F}_p(r) = \frac{\sum_{i=1}^{n} P(c_i) I(R(c_i) \leq r)}{\sum_{i=1}^{n} P(c_i)}
\]

\[
\hat{F}_L(r) = \frac{\sum_{i=1}^{n} y_i I(R(c_i) \leq r)}{\sum_{i=1}^{n} y_i}.
\]

If we now plot the two ordered distributions against each other, they produce according to the theory section a valid ordered Lorenz curve which yield, from an analytic perspective, a deeper visual understanding of the portfolio’s performance rather than just the loss versus premium curve. To make distinguished comparisons between portfolios the Lorenz curve is well suited to the mathematics of the well established measure of difference between two distributions, the Gini index.
3.5 Calculate the Gini Index

As mentioned in the theory section, the Gini index can be interpreted as the area of which the Lorenz curve cover between the line of "break even" and the horizontal axis. In general theory the measure is coherent with spanning from (0,1). By the following extended definition of the Gini Coefficient we will let the measurement take the values between (-1,1), since it is a possibility that the portfolio is indeed underachieving. Since we are dealing with empirical functions, we cannot perform the theoretical measure 2.24, therefore we introduce the next step in the algorithm, the discrete Gini Coefficient. The discrete definition follows [5] deduction.

\[
\hat{G}(\hat{F}_P, \hat{F}_L) = 2 \sum_{i=1}^{n-1} \left( \hat{F}_P(i+1) - \hat{F}_P(i) \right) \left( \frac{\hat{F}_P(i+1) + \hat{F}_P(i)}{2} - \frac{\hat{F}_L(i+1) + \hat{F}_L(i)}{2} \right)
\]

(3.6)

\[
\hat{G}(\hat{F}_P, \hat{F}_L) = 1 - \sum_{i=1}^{n-1} \left( \hat{F}_P(i+1) - \hat{F}_P(i) \right) \left( \hat{F}_L(i+1) + \hat{F}_L(i) \right)
\]

(3.7)
The results and analytic value of the algorithm, performed on the two applied portfolios, will be presented in this section. Graphs from all three described Lorenz curves in the chapter regarding the method will be evaluated and the results from ordering the Lorenz curve by its relative will be presented. Furthermore, the results of the Gini index from corresponding graph are summarized and analyzed. If we recognize the stipulated aim of this report, one should remember that it is not the results of the digits which are of interest rather the results of the method’s application and what we intend to describe or measure.

4.1 Results of Portfolio A

Portfolio A contained 16 insurances with claim history and was described by its 2 characteristics, the turnover and the insurance amount. We begin by observing the graph 3.1 where the classical Lorenz curve is displayed with the premium as ordering variable. This figure is intend to give the observer a glance at the premium pricing distribution. What is relevant to notice is that the 50th-percentile of the people only covers roughly 25% of the premium, clearly the top 50% of the policyholders are covering the major part of the portfolios total income. Whether or not one can say that the results indicates that the likeliness of claim being more likely to occur for a policyholder with greater premium, since we only looking at the policies with existing claims in the entire portfolio, is not something that is evaluated by this method. Rather the result provides an easy understanding how this subset of this portfolio’s policyholders premiums are distributed.

The results from comparing the premium distribution with the loss distribution can be seen in the figure 3.3. The curve implies that the lower 50th percentile of the premiums covers around 35% of the total loss. Note that those policies which corresponds to the premiums are not connected
to the losses and therefore it is hard to reach any conclusion more than that the premium per policyholder in graph 3.1 where skewed to be top heavy while the premium versus losses seems to be more equal in distribution, therefore this Portfolio has greater variation in its losses.

Further on, we study the results from the method of ordering the Lorenz curve by our relative variable. That is, our scoring variable the expected loss, divided by the policy’s premium. This graph 4.1 is presented below and results in the following. First, the result at the 50th lower percentile of the premium distribution covers only 20% of the losses. Here we see that, under the ordering of our relative loss, the portfolio is indeed top heavy. Second, the upper 5th-percentile covers the top 20% of the losses and compared to the 3.3 we see a much wider dispersion from the "line of equality". The losses and premiums are now connected through our relative variable which makes it easier to identify such insurances which puts a greater weight on the portfolio, hence we may construct a more profitable portfolio with these results.

Furthermore, the Gini index is calculated on all three graphs for assessment of the portfolio’s performance. As the figures are presented in table 4.1. Now, to compare the result from the classic Lorenz curve, that is the premium distribution versus the policyholders, with the two other does not provide much sense. However, the result of that Gini index will be of interest further on in the discussion between the two portfolio’s performance. When comparing the ordered Lorenz curve's and the non-ordered's Gini index there is a greater distinction between their results.

<table>
<thead>
<tr>
<th>Premium vs. Policyholders</th>
<th>Premium vs. Losses</th>
<th>Ordered Lorenz Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>39.38%</td>
<td>22.12%</td>
<td>46.16%</td>
</tr>
</tbody>
</table>

Table 4.1: The estimated Gini index for each of Portfolio A's graphs
4.2 Results of Portfolio B

The second portfolio which we performed the analysis on contained a set of 14 policies with each a corresponding claim. The policies for this portfolio were described by the ten characteristics $c_B$. With the classical approach and Lorenz curve, that is the policyholder versus the premium distribution, the results in graph 3.2 visualize a portfolio with a flatter curve, closer to the line of even distribution. Comparing with Portfolio A, the result shows that for Portfolio B we see that about 50% of the lower percentiles of the policyholders cover 40% of the premiums. Observing the results from the graph 3.4 there is a larger dispersion from the line where we established even distribution between losses and premiums.

Now, the results show that the lower 50th percentile of the premiums cover 15% of the losses. However due to the sorting of the loss and premiums the result is not of further interest as described in the theory section. We continue by looking at the result that the ordered Lorenz curve yields. The curve is observed in the graph 4.2. The results show that by ordering through the set relative variable $R(c)$ it yields a curve with a further attraction to the south east corner $(1,0)$. The lower 50th percentile of the ordered premiums cover less than 10% of the losses. Furthermore, the Gini indices, which tells us something about the profitability of the portfolio, are evaluated as well for Portfolio B and visualized in table 4.2. As our empirical understanding of the curve tells us, the Gini index of the Ordered Lorenz curve shows a larger profitable segment of policies. With the result from both Portfolio A and B and their Gini indices, an analyst could now go ahead and compare and draw conclusions from their respective digits.

![Ordered Lorenz curve of portfolio B](image)

Table 4.2: The estimated Gini index for each of Portfolio B’s graphs

<table>
<thead>
<tr>
<th>Premium vs. Policyholders</th>
<th>Premium vs. Losses</th>
<th>Ordered Lorenz Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.71%</td>
<td>51.39%</td>
<td>64.57%</td>
</tr>
</tbody>
</table>

Figure 4.2: Ordered Lorenz curve of portfolio B
CHAPTER 4. RESULTS

4.3 Results of the Non-Claim Insurances in Portfolio A

With the regression coefficients determined, the insurances which had no previous claim history could be rated. In the same procedure as before, the ordered Lorenz curve could then be determined based on its relative variable, i.e. 2.18. With its larger sample size, the distribution tends towards less dispersion from the line of equality and its lower 50th-percentile is now covering closer to 40% of the potential losses. While the curve is still below the line of equality this indicates a profitable portfolio and those insurances which are in the upper range could now be easier to locate and for an analyst perhaps re-structure the portfolio. Since we are looking at a portfolio with a sample of 1321 different insurances, this result indicated that refining the portfolio structure with the lower 70th-percentage of the premiums, which are related to 50% of the losses, would yield a profitable portfolio and lowering the risk of insurances with a higher relative loss. The Gini index, which can be seen in table 4.3 would in turn also be even higher with a re-structuring of the portfolio.

![Ordered Lorenz Curve Portfolio A](image)

Figure 4.3: The ordered Lorenz curve of the entire portfolio A

<table>
<thead>
<tr>
<th>Ordered Lorenz Curve</th>
<th>18.26%</th>
</tr>
</thead>
</table>

Table 4.3: The estimated Gini index of the entire Portfolio A's graph
This section intends to evaluate the theory, method and the results which was provided by the data. The underlying aim for this discussion is not to stipulate a formal conclusion for the application of this method, rather the text will be formulated in such way that the reader may comprehend the method and its abilities to lead them into thoughts for their own application and further development in the real world scenarios.

5.1 Discussing the Method and its Improvements

The theory behind the Lorenz curve is well established by economics and forms a basic intuition of any economical distribution relation one wish to examine. For an analyst in the insurance business the relation of premium and risk or losses would always be of interest. Are their relation in line with the strategy an insurer or broker has set up before covering such portfolios. Now, as a simple first step using the general Lorenz curve by observing the premium versus policy holder distributions, seen in graphs 3.1 and 3.2, they surely provides a basic understanding of how the portfolio’s premiums are distributed. One may discover fairly easy how Portfolio B’s premiums follows a more even distribution than Portfolio A. Likewise the following measure of the corresponding Gini index follow the understanding that the premiums in Portfolio B is distributed rather more evenly than Portfolio A with a much smaller figure, 0.1471 compared to 0.3134. Intuitively, the Gini measure for this case can be viewed as risk measure, rather than profitability as described for the Lorenz curve where one observe the premium versus losses. Hence the importance of understanding what one intend to describe. A portfolio structured similar to Portfolio B would be less sensitive to a drop or change of policyholders since their cumulative premium are closer to an even spread than Portfolio A, in which contrary will be more exposed to risk since the upper 50th percentage of policyholders carries a greater portion of the
CHAPTER 5. DISCUSSION

total premium in the portfolio. So the general Lorenz curve method does provide information but only with so much. Regarding relative exposure against losses the ordered Lorenz curve method provide deeper insight.

The method was based purely on observations and empirical distributions. Surely it is an easy and straight forward method but if one would dive deeper in the statistics of the method, a more theoretical approach would be of interest to ensure the statistical validity even further. Since data points are available an approach to assign a distribution to the random variable would be to observe the data points, fit a distribution which seems reasonable and with the use of a quantile-quantile plot, see [9], verify its validity. Once the distribution functions are known, the distributions could be plotted with its continuously curve and the insurance scoring could be better determined and statistically validated through confidence intervals.

Looking back at the intended purpose of this analysis, it was asked for a method which could identify specific block of policyholders based of their characteristics. Based on the order Lorenz curve method an insurance analyst can, based on the distribution be it the empirical or theoretical, identify the blocks of insurances which are above a set mark and re-balance the insurance portfolio with policyholders, having desired characteristics, and increase the profitability and reduce the risk of higher losses. For further analysis it is suggested to observe the characteristics which belongs to the insurances which are, according to the analyst performing well, and try to identify their treats and what separates them from the riskier ones.

5.2 The Scoring Function

Based on the data the theory showed that a scoring function, 2.8 could be determined based on solely a few data points. The scoring function 2.6 could be determined in several ways, but for this reports purpose of explaining the basics of a scoring function, a simple case of a linear function fulfills the criteria. Nevertheless, for the purpose of further develop a discussion of more elaborated approaches will follow in this section.

Now, this method elaborates only with the standard linear case. In some scenarios a rating plan is proposed but one would want re-think the characteristics which are suggested to be used when determine such a scoring function. A well known method to investigate the influence under ones measure is by its variable is the Principle Component Analysis. In short terms, this method seeks a maximize the variance and separation between groups and identify the driving variables, i.e. under which component tends the objects in the group be separated at most. With the use of Principle Component Analysis, an insurance analyst could use it to it advantage and reduce the number of characteristics being used in the scoring function. For further reading and explanation,
see [3].

By evaluate the theoretical distribution of the data set, one could perform a regression analysis with a function closer in coherence with the theoretical distribution function for the losses. With the inclusion of the characteristics in the scoring function as variables. As an example, in [5] they determine the data of insurance losses to be of exponential distribution, hence assigning the scoring function to be exponential.
BIBLIOGRAPHY


Inequalities: Theory of Majorization and Its Applications. 

METHODS OF MULTIVARIATE ANALYSIS Third Edition. 

The SAGE Handbook of Regression Analysis and Causal Inference. 

Summarizing insurance scores using gini index. 

Insurance ratemaking and a gini index. 

A General Definition of the Lorenz Curve. 

The estimation of the lorenz curve and gini index. 

Risk and Portfolio Analysis, Principles and Methods. 

*The Gini Methodology, A Primer on a Statistical Methodology.*
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10082</td>
<td>102003</td>
<td>54000000</td>
<td>5000000</td>
</tr>
<tr>
<td>2</td>
<td>3014</td>
<td>107500</td>
<td>1700000</td>
<td>5000000</td>
</tr>
<tr>
<td>3</td>
<td>3711</td>
<td>735313</td>
<td>2000000</td>
<td>10000000</td>
</tr>
<tr>
<td>4</td>
<td>3761</td>
<td>468433</td>
<td>170000000</td>
<td>10000000</td>
</tr>
<tr>
<td>5</td>
<td>9134</td>
<td>508750</td>
<td>30000000</td>
<td>10000000</td>
</tr>
<tr>
<td>6</td>
<td>7531</td>
<td>100000</td>
<td>24000000</td>
<td>10000000</td>
</tr>
<tr>
<td>7</td>
<td>3087</td>
<td>1480000</td>
<td>2000000</td>
<td>5000000</td>
</tr>
<tr>
<td>8</td>
<td>11360</td>
<td>43750</td>
<td>150000000</td>
<td>10000000</td>
</tr>
<tr>
<td>9</td>
<td>15500</td>
<td>350000</td>
<td>16400000</td>
<td>10000000</td>
</tr>
<tr>
<td>10</td>
<td>12516</td>
<td>826524</td>
<td>347207000</td>
<td>10000000</td>
</tr>
<tr>
<td>11</td>
<td>2983</td>
<td>15000</td>
<td>1156000</td>
<td>5000000</td>
</tr>
<tr>
<td>12</td>
<td>19462</td>
<td>100000</td>
<td>288000000</td>
<td>20000000</td>
</tr>
<tr>
<td>13</td>
<td>16762</td>
<td>200000</td>
<td>94627000</td>
<td>20000000</td>
</tr>
<tr>
<td>14</td>
<td>8619</td>
<td>100000</td>
<td>141172000</td>
<td>10000000</td>
</tr>
<tr>
<td>15</td>
<td>7429</td>
<td>20100</td>
<td>34000000</td>
<td>5000000</td>
</tr>
<tr>
<td>16</td>
<td>7958</td>
<td>42375</td>
<td>12700000</td>
<td>10000000</td>
</tr>
</tbody>
</table>

Table A.1: Insurance portfolio A
Table A.2: Insurance portfolio B

<table>
<thead>
<tr>
<th>PolicyId</th>
<th>Premium [SEK]</th>
<th>Claim [SEK]</th>
<th>Turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6090</td>
<td>25200</td>
<td>2000000</td>
</tr>
<tr>
<td>2</td>
<td>6000</td>
<td>200000</td>
<td>2000000</td>
</tr>
<tr>
<td>3</td>
<td>10700</td>
<td>30001</td>
<td>25000000</td>
</tr>
<tr>
<td>4</td>
<td>13230</td>
<td>1197</td>
<td>3000000</td>
</tr>
<tr>
<td>5</td>
<td>7754</td>
<td>12675</td>
<td>3300000</td>
</tr>
<tr>
<td>6</td>
<td>7000</td>
<td>133310</td>
<td>2100000</td>
</tr>
<tr>
<td>7</td>
<td>5500</td>
<td>30001</td>
<td>1000000</td>
</tr>
<tr>
<td>8</td>
<td>5499</td>
<td>9001</td>
<td>750000</td>
</tr>
<tr>
<td>9</td>
<td>6715</td>
<td>70000</td>
<td>1100000</td>
</tr>
<tr>
<td>10</td>
<td>8925</td>
<td>6000</td>
<td>8000000</td>
</tr>
<tr>
<td>11</td>
<td>6569</td>
<td>1000</td>
<td>10000000</td>
</tr>
<tr>
<td>12</td>
<td>5523</td>
<td>38000</td>
<td>800000</td>
</tr>
<tr>
<td>13</td>
<td>7266</td>
<td>1093</td>
<td>1350000</td>
</tr>
<tr>
<td>14</td>
<td>6825</td>
<td>960</td>
<td>25000000</td>
</tr>
</tbody>
</table>

Table A.3: Insurance portfolio B, insurance amounts in MSEK

<table>
<thead>
<tr>
<th>PolicyId</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.444</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.444</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5,05</td>
<td>17</td>
<td>2</td>
<td>5</td>
<td>0.445</td>
<td>1</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>5,05</td>
<td>17</td>
<td>5</td>
<td>2</td>
<td>0.445</td>
<td>1.2</td>
<td>0.1</td>
<td>6.15</td>
<td>0.09</td>
<td>5.34</td>
</tr>
<tr>
<td>5</td>
<td>5,248</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.443</td>
<td>2.12</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.444</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.444</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>5,248</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.443</td>
<td>1.1</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>5,248</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.443</td>
<td>1.24</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.445</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>0.445</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>5,196</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.445</td>
<td>1.12</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>17</td>
<td>5</td>
<td>2</td>
<td>0.445</td>
<td>0</td>
<td>0</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>5,05</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.445</td>
<td>1</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------</td>
<td>---------------</td>
<td>------------</td>
<td>----------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>20100</td>
<td>72118</td>
<td>0.3</td>
<td>34</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>42375</td>
<td>95624</td>
<td>0.4</td>
<td>12.7</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>102003</td>
<td>155446</td>
<td>0.7</td>
<td>54</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>15000</td>
<td>15365</td>
<td>1</td>
<td>1.156</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>100000</td>
<td>58878</td>
<td>1.7</td>
<td>288</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>43750</td>
<td>24517</td>
<td>1.8</td>
<td>150</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>100000</td>
<td>39505</td>
<td>2.5</td>
<td>24</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>200000</td>
<td>67457</td>
<td>3</td>
<td>94.627</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>826524</td>
<td>231370</td>
<td>3.6</td>
<td>347.207</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>508750</td>
<td>139500</td>
<td>3.6</td>
<td>30</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>100000</td>
<td>26359</td>
<td>3.8</td>
<td>141.172</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>107500</td>
<td>19958</td>
<td>5.4</td>
<td>1.7</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>350000</td>
<td>59254</td>
<td>5.9</td>
<td>16.4</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>468433</td>
<td>76426</td>
<td>6.1</td>
<td>170</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1480000</td>
<td>150756</td>
<td>9.8</td>
<td>2</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>735313</td>
<td>42524</td>
<td>17.3</td>
<td>2</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.4: Portfolio A ordered by the relativity variable

<table>
<thead>
<tr>
<th>PolicyId</th>
<th>Score</th>
<th>Premium</th>
<th>Relativity</th>
<th>Turnover</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1197</td>
<td>13230</td>
<td>0.1</td>
<td>3</td>
<td>17</td>
<td>5</td>
<td>2</td>
<td>0.445</td>
<td>1.2</td>
<td>0.1</td>
<td>6.15</td>
<td>0.090</td>
<td>5.34</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>960</td>
<td>6825</td>
<td>0.1</td>
<td>25</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.445</td>
<td>10</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1093</td>
<td>7266</td>
<td>0.2</td>
<td>1.35</td>
<td>0</td>
<td>17</td>
<td>5</td>
<td>2</td>
<td>0.445</td>
<td>0</td>
<td>0</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>1000</td>
<td>6569</td>
<td>0.2</td>
<td>10</td>
<td>5</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>0.455</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>6000</td>
<td>8925</td>
<td>0.7</td>
<td>8</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.455</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>12675</td>
<td>7754</td>
<td>1.6</td>
<td>3.3</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.443</td>
<td>2.12</td>
<td>0.1</td>
<td>6.15</td>
<td>0.15</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>9001</td>
<td>5499</td>
<td>1.6</td>
<td>0.75</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.443</td>
<td>1.1</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>30001</td>
<td>10700</td>
<td>2.8</td>
<td>25</td>
<td>17</td>
<td>2</td>
<td>5</td>
<td>0.445</td>
<td>10</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>30001</td>
<td>5500</td>
<td>5.5</td>
<td>1</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.444</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>38000</td>
<td>5523</td>
<td>6.9</td>
<td>0.8</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.455</td>
<td>1.12</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>70000</td>
<td>6715</td>
<td>10.4</td>
<td>1.1</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.443</td>
<td>1.24</td>
<td>0.1</td>
<td>6.15</td>
<td>0.015</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>112600</td>
<td>6090</td>
<td>18.5</td>
<td>2</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.444</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>112600</td>
<td>6000</td>
<td>18.8</td>
<td>2</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.444</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>133310</td>
<td>7000</td>
<td>19</td>
<td>2.1</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0.444</td>
<td>0.8</td>
<td>0.1</td>
<td>6.15</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table A.5: Portfolio B ordered by the relativity variable. The insurance amounts and turnovers are in MSEK while the Score and Premium are in SEK