Stockholms universitet

Construction of rating territories for water-damage claims

Patrik Emanuelsson

Masterupppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2011:8
Matematisk statistik
September 2011

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2011:8**
http://www.math.su.se

Construction of rating territories for water-damage claims

Patrik Emanuelsson*

September 2011

Abstract

Sweden is the fifth largest country in Europe and has considerable regional differences with respect to geography, demography and meteorological conditions. From relative flat terrain in the south to larger hills and mountains in the north-west, higher level of precipitation along the west border and larger cities and densely populated areas by the coastal line. In an attempt to explain the increased number of water-damage claims during the past years we want to know how these are related to differences in geography. We investigate how one can create rating territories using generalized linear models, credibility theory, smoothing and clustering techniques.

Under the hypothesis that all residual variation in a generalized linear model for claim frequency is a pure effect of geography we are able to estimate the relative risk of water-damage in each municipality. The estimates are used in order to aggregate the municipalities into larger territories reflecting an elevation and similarity of risk. We can conclude that the best way to group geographical units is using a minimum within territory variance criterion and aggregate by adjacency. Included in a generalized linear model the zone-variable turns out highly significant and there are no remaining detectable differences between the geographical units.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail:patrik.emanuelsson@gmail.com . Supervisor: Dmitrii Silvestrov.

Preface

This report constitutes a thesis of 30 credits in Mathematical Statistics and leads to a Master of Science degree in Actuarial Mathematics from Stockholm University.

I would like to direct my sincerest gratitude to my supervisor Tobias Janstad, whom have guided and supported me in my daily work throughout the whole process of writing this thesis. I would also like to thank my supervisors Peter Møller and Dmitrii Silvestrov for all advice and for sharing your thoughts with me, you have helped me a lot.

Patrik Emanuelsson
Stockholm, 2011-08-30

Contents

List of Figures

List of Tables

1 Preliminaries

1.1 Introduction

Insurance is the business of transferring risk from one party – the policyholder, onto another – the insurer for a predetermined premium, which in turn provides the service of financial reimbursement in case of loss under policy coverage. Determination of a fair premium is of outmost importance for an insurer considering retention of current customers, acquiring new business and gaining a competitive edge. When the policy is targeting a large population and in addition spread over a considerable land areal there are a number of factors generating differences in risk of a claim, whereof some could be derived to or classified as geographical. In most lines of business we could expect that location is important as explanatory of the loss experience since the geographical perils differs a great deal between regions. If an insurer uses rating territories it is wise to keep the definitions up to date as relativities change as a result of competitive market forces [3].

In most non-life insurance companies a private customer can sign a home insurance policy that covers the building, property or both in case of loss. The contract usually covers water-damages incurred as a result of heavy rains, snow melting, rising lakes or rivers and a wide variety of leakages that might occur. Water-damages generally make up a large part of the total claim costs for any insurer. In order to provide the customers with the most fair priced product the loss must be distributed onto the groups that account for it. This is done by including new rating factors or updating the current variable definitions. This thesis will therefore be addressing the question of how geographical classification can be used in order to take into account within collective differences for water-damages related claims in private housing insurance. We will study a couple of smoothing and clustering routines that can be used for construction of rating territories.

1.2 Purpose

The primary objective for this thesis is to establish a methodology for the creation of rating territories and in particular to apply the method on water-damage claims that can be seen as geographically contingent. The questions to be answered are listed below.

1. Where are water claims located spatially?
2. Is there any distinctive pattern, such as if water claims are located close to rivers and lakes?
3. How should a water claim zone variable, for both frequency and severity, optimally be constructed using credibility theory?

4. How does a GLM including a zone variable contribute to a non-spatial GLM?

5. Is non-random pattern in model residual reduced when the zone variable is added?

1.3 General pricing

The premium for an insurance contract is determined by calculating the expected loss that the contract will incur during the policy duration. The premium calculated this way is called the pure premium. Compared to the actual premium charged additional considerations are taken into account, such as administrative expenses. If an insurer has n identical contracts with the same policy duration, whereas each contract will incur a claim cost of $X_1, X_2, ..., X_n$,

$$X_i = \begin{cases} 0, & \text{if there is no claim associated with contract } i \\ > 0, & \text{if contract is aquired at least one claim.} \end{cases}$$

The pure premium for each contract should equal an amount of,

$$E[X_i] = \mu \quad \text{with} \quad Var(X_i) = \sigma^2.$$

Since an insurer has several policies, on average a reduction variance for each contract together with a greater predictability is gained due to the law of large numbers,

$$E[\overline{X}] = \mu \quad \text{and} \quad Var(\overline{X}) = \frac{\sigma^2}{n}.$$

We could thus set the pure premium by calculating the observed average claim cost $\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$ for past years and apply onto new business. In order for the premium to be fair we need to take into account individual differences in risk. These differences can be classified into three general groups.

(1) Policyholder characteristics

- Age

(2) Insured object characteristics

- Value of house
- Type of house
- Building year

(3) Geographical characteristics

- Urban/rural area
- Demographical (population density, education, income etc.)
- Environment

Factors considered in the process of pricing – premium arguments, yield the individual premiums,

$$E[X_i] = \mu_i. \tag{1}$$

Each contract/individual has an associated exposure – the duration w_i, a measure of how long a contract has been in force, expressed in the number of insurance years. In the language of statistics, the collective pure premium for a specific contract could be expressed as having a set of independent random variables $N, X_1, X_2, ...$ where N is the number of claims during a period of e.g. one calendar year. The set of random variables $\{X_i\}$ represent the cost of claim $i = 1, 2, ...$ with aggregated duration $D = \sum_{j \in J} w_j$, where J is the set of all contracts. We want to determine the expected cost per policy and have the following definitions.

Claim severity

$$S = \frac{\text{Claim cost}}{\text{Number of claims}} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Claim frequency

$$F = \frac{\text{Number of claims}}{\text{Duration}} = \frac{N}{D}$$

Pure premium

$$P = \frac{\text{Claim cost}}{\text{Duration}} = \frac{\sum_{i=1}^{N} X_i}{D} = S \times F.$$

Above quantities is referred to a *key ratios* which are of the general form,

$$Y = \frac{X}{w} = \frac{\text{Random variable}}{\text{Exposure}}. \tag{2}$$

Exposure is a measure of how much the insurer is subjected to loss in a particular variable e.g. claim cost or number of claims. These quantities should preferably fulfill the conditions of being practical (easy to obtain and verify) and proportional to loss i.e. two exposures should lead to two times the expected loss of one exposure [13]. Because one usually has an idea of the amount of policies there are for a particular line of business, the duration is considered constant. We are left with the challenge of calculating the expectation of a compound distribution $S_N = \sum_{i=1}^{N} X_i$. Start by assuming that $E[X_i] = \mu_X$ for all i, we get the following result,

$$E[S] = E\left(E[S|N]\right) = E\left(E\left[\sum_{i=1}^{N} \frac{X_i}{N} \middle| N\right]\right) = E\left(\frac{NE[X_i|N]}{N}\right) =$$
$$= E\left(\frac{NE[X_i]}{N}\right) = \mu_X. \quad (3)$$

Using equation (3) the expected pure premium can now be calculated,

$$E[P] = E\left[\frac{\sum_{i=1}^{N} X_i}{D}\right] = E\left(E\left[\frac{\sum_{i=1}^{N} X_i}{D} \middle| N\right]\right) =$$
$$= E\left(\frac{N}{D} E[X_i|N]\right) = E\left[\frac{N}{D}\right] \times \mu_X = E[F] \times E[S]. \quad (4)$$

This means that we can divide the pure premium calculation into two parts, expected claim frequency and claim severity. For a given set of rating variables we can construct a tariff i.e. "pricing chart". Each tariff cell provides us with the amount that in relation to a constant should be charged a policyholder whom hold the given characteristics. Continuing in line with [10] we specify three basic model assumptions needed in order to proceed.

(i) Policy independence – policy events does not affect one another.

(ii) Time independence – policies (response) are independent of which time interval a event occurs.

(iii) Homogeneity – the risks within a tariff cell are similar i.e. homogeneity of risks in tariff cells.

In order for a risk to be insurable the above requisites should be fulfilled as well as a criteria of mass or risk exposure. As a consequence of the law of large numbers, with enought exposure we will be able to make reasonable estimations of the actual risk at hand [15].

1.4 Territorial ratemaking

Geographical location is considered one of the most important factors influencing the claims experience in a home insurance policy and is also heavily correlated with other rating factors such as value of home [13]. The process of creating rating territories is termed territorial ratemaking and is split up into two separate parts.

(1) Establishing boundaries

(2) Determination of relativities

In establishing boundaries for a territory, we have at first to decide which geographic unit that should be used. Geographical units are small distinct areas used as building blocks for aggregation into larger territories reflecting a joint effect of location that cannot be explained by the other rating factors. The territorial effects does not necessarily shed light on physical factors influencing claims experience, but rather that the individuals/objects within a territory has something in common and is intuitively physical in nature. We have to consider that the units should satisfy the two prerequisites.

(1) Each unit should be small enough to reflect homogenous geographical risk.

(2) Each unit should preferably be static i.e. does not change with time.

The most frequently used units in the literature are post codes [1], [3], [6], [14], because it is small enough to reflect homogenous risks and usually readily available for implementation. Unfortunately post codes suffer from the disadvantage of being subject to change over time, the same applies to parishes. At the same time we should keep in mind that these units are constructed for administrative reasons or in order to achieve a effective postal delivery. Two houses close in distance, but in different postcodes, have more in common than houses on opposite sides of the same postcode. There are about 9000 post codes and 1200 parishes in Sweden, which potentially could be used as geographical units. Another option would be using municipalities or counties, where the former are likely to have and the latter might have heterogeneous risks within geographic unit due to their substantial size. Some further guidelines in how we should choose geographical units [13].

(3) Unit should be large enough to produce credible estimates.

(4) Easy to assign company data to unit.

(5) Easy to map external data to unit.

(6) Easy to understand unit construction.

(7) Politically acceptable

(8) Verifiable

The claims process is hypothesized to depend on both geographical as well as non geographical factors and we can also to expect to observe a fair amount of random noise in the actual experience. Geographical factors can be divided into geo-demographic factors, such as demography and population density and geo-physical, such as altitude, access to water within the vicinity or farmed land. One could go about investigating geography with a multivariate approach, using a set of explanatory variables to explain some of the variation in claim experience and examine how the residual variation varies by geographic location. If some locations have something(s) in common in excess of explanatory variables, clusters of higher residual variation in these areas might be observed, i.e. if observed experience is not in line with predicted. In order to take a set of explanatory variables into account we suggest using generalized linear models.

1.5 Generalized linear models

The exponential family is a large class of probability models including some of the most important and frequently used, such as the Poisson-, Gamma-, Bernoulli- and Normal distributions. The following is a general definition of the class of exponential distributions taken from [11].

Definition 1.1. A statistical model for data set $\mathbf{y} = (y_1, ..., y_n)$ is an exponential family(or of exponential type) with canonical parameter vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ and canonical statistic $\mathbf{t}(\mathbf{y}) = (t_1(\mathbf{y}), ..., t_k(\mathbf{y}))$, if f has the structure

$$f(\mathbf{y}; \boldsymbol{\theta}) = a(\boldsymbol{\theta}) h(\mathbf{y}) e^{\boldsymbol{\theta}^T \mathbf{t}(\mathbf{y})}, \tag{5}$$

with the normalizing constant $a(\boldsymbol{\theta}) = 1/C(\boldsymbol{\theta})$,

$$C(\boldsymbol{\theta}) = \int h(\mathbf{y}) e^{\boldsymbol{\theta} \mathbf{t}(\mathbf{y})} d\mathbf{y}.$$

□

The expression in equation (5) can be written for a single observation as,

$$f(y_i; \boldsymbol{\theta}) = \exp\left\{\boldsymbol{\theta}^T \mathbf{t}(y_i) - \log(C(\boldsymbol{\theta})) + \log(h(y_i))\right\}. \tag{6}$$

A generalization when $\mathbf{t}(y_i) = y_i$, including a dispersion parameter ϕ [8]

$$f(y_i, \theta) = \exp\left\{\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i, \phi)\right\} \tag{7}$$

for some functions $a(\cdot), b(\cdot), c(\cdot)$ having the properties,

$$E[Y] = b'(\theta) = \mu \quad Var(Y) = b''(\theta)a(\phi) \quad v(\mu) = b''(b'^{-1}(\mu)). \tag{8}$$

The function $v(\mu)$ – the variance function, fully specifies the probability distribution that has the form as in equation (7) [10]. A common version of the function $a(\phi)$ used e.g. by [10], $a(\phi) = \phi/w_i$ where w_i is a prior weight and implemented in (7) gives rise to the exponential dispersion model (EDM). A subclass of models with restriction that the variance function has the form

$$v(\mu) = \mu^p, \quad p \in (-\infty, \infty) \tag{9}$$

called Tweedie models, include Poisson ($p = 1$) and gamma distribution ($p = 2$) among others. The Tweedie models possess the desirable property of being scale invariant, i.e. Y and cY follows the same family of distributions. This is useful in insurance applications for example if we were to change currency or correct for inflation in the response variable.

2 Data

Data has been obtained from a large insurance company, containing records of individual insurances and events of water-damage claims in Sweden for the period 2004 – 2010. The material consists of over 360'000 policies, exposed in total to an amount of 1'250'000 policy years for which there has occurred 15'492 cases of damage. A water-damage can occur in many forms, e.g. pipes can rupture, machinery or home appliances might break down. In this study we are primarily interested in claims that can be seen as geographically contingent and impose a restriction on the claim definition.

The claims to be investigated are of damages to both personal property and buildings which are classified as one of the following two kinds.

1. Inflowing water, which can be attributed to precipitation, meltwater or rising level of a nearby lake.

2. Drainage system, water flowing from the sewer, usually as a consequence of how efficiently the municipality handles excess precipitation.

Because the policies has multiple insured objects and the causes of damage can be different, one claim is defined as a unique date of damage and for a specific policy id, one claim will hence not be counted twice.

Next we need to determine the geographical units. Sweden is Europe's fourth largest country (excluding Russia) with a land area of 449'946 km^2 and does have varying geographical conditions. The country is fairly flat in the south with higher hills and mountains in the north. To construct a homogeneous geographical unit one should optimally divide the country into a grid of equal size squares. If each square have a side of 500m it would have a corresponding area of 0.25 km^2 i.e. in order to cover the whole country we need about 1.8 million squares, which is more than the total amount of exposure in terms of policy years. Even squares with a side of 1km would be quite unsatisfactory in rural areas, since there are not enough exposure. We have to violate the condition of optimal homogeneity and decide that it is reasonable to use municipalities instead, mainly because the ease of implementation and that the division fulfill requirement (1) – 8) in section 1.4. As previously noted, there are 290 municipalities in Sweden, which correspond to an average size of 1'151km^2, exposure of 4'300 years and 53 claims. The average municipality area corresponds to a square with a side of about 34km. Even after having used a much rougher division than optimal there is still great variation between geographical units. The minimum amount of exposure, 257 years and 4 cases of damage occurs in the municipality of Ydre just east of Lake Vättern, slightly more exposed is Bjurholm, west of Umeå in the north, with 273 policy years and 2 cases of water-damage. The overall average claim frequency amount to 0.0124 per policy year. Since a case is a quite rare event occurring on average once

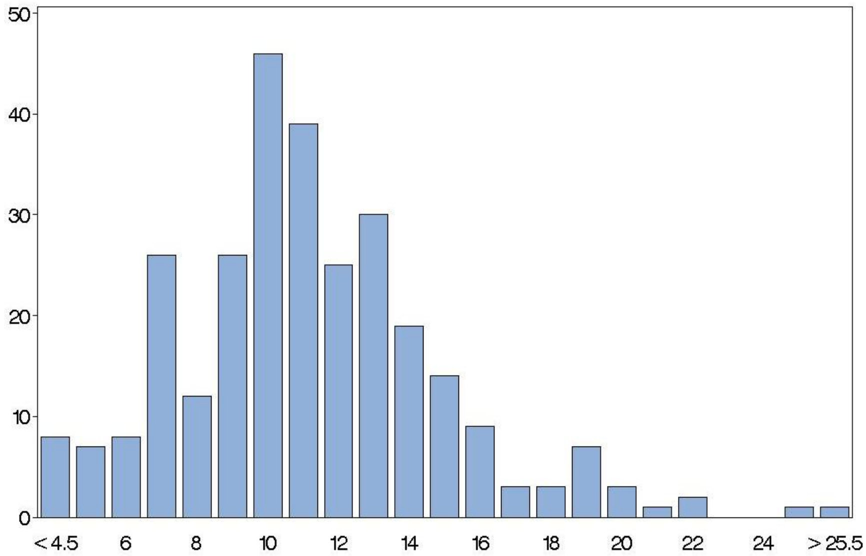

Figure 1: Histogram of the municipality claim frequency per 1000 years of exposure, for the years 2004 – 2010.

every 80th year there are a lot of randomness in a intensity estimate for the municipalities with little exposure. At the other end, the most exposed unit – Stockholm municipality, has 63'082 policy years and 1'085 claims. In order to get a better picture of the variation of estimates amongst our chosen units we depict the empirical frequencies per 1000 policy years in a histogram, seen above in Figure 1. The values on the horizontal axis correspond to the midpoint of a one unit interval except for the largest and smallest values, the vertical axis represent the number of municipalities that fall into each bin. By the same classification levels as in the histogram we plot the claim frequencies on a map, with graduated colors from dark green for the lowest frequencies to fierce red for the ones classified by the highest category, Figure 24. The empirical claim frequencies for municipalities have a median value of 0.0108 and vary between 0.0031 and 0.0268 i.e. with a factor of 8.6 between the smallest and largest values.

2.1 Modeling

We employ the approach of dividing the data into two samples by the amount of exposure. We generate a uniform random number between zero and one, sort in ascending order and choose the first observations which constitute one-seventh of the total exposure, we refer to this sample as the "control sample". All the other observations make up the "model sample". All analysis and modeling is done with the model data sample, resulting estimates are thereafter used in an evaluation of the model fit and appropriateness against the control data sample. The prime motivation for dividing the data set into two samples randomly and not by year is under the assumption that the number of claims is not stationary by year and hence a difficulty to determine the calendar year effect. For each observation we have records of policyholder and insured object characteristics as well as duration and date variables. Along with these, we have also gathered data that reflect general geographical conditions, that are implemented in section 5. The variables we believe have the greatest impact on the size and number of claims are the following.

Table 1: Characteristics that can be attributed to each policy.

Characteristics		
Policyholder	Object	Geographical
• Age	• Building age • Living area size • Ground floor size • Number of buildings • Basement	• None

Building and policyholder age are discrete variables, of which building age adopt values in the range 0 to 80 and policyholder age even include values above. Whereas a value of 80 include all ages greater or equal to 80. The higher age of a building, the greater risk for damage if not properly maintained. The reverse relationship is more plausible for policyholder age since the older one gets the more experience is acquired as well as more time and funds might be allocated to house maintenance tasks and renovations. Since owning a house is uncommon amongst younger policy holders, it is reasonable to group ages in the interval [0,30] into a first category, and for ages above into 10 year bands up to the age of 80. Building age has a quite large group of policies of age 0 (8'122 years) which at a first stage will form one group and values above are divided into 5 year classes.

Basement is a four level categorical variable with values, yes, no, missing and split-level home (swe. souterränghus). If the damp insulation on the outside of the basement walls and floor do not keep the interior of the basement sufficiently dry, the moisture will penetrate the walls and any organic material exposed in the basement will get damaged. For both basements and split-level houses problems with water-damage might arise if rain and melt-water is not properly drained from the construction. A basement is moreover a vulnerable part of a house in case of a flood. Detailed information regarding the characteristics of a particular insured object must be supplied or collected from the policyholder himself. During the studied period, the collection of the basement variable has been disrupted and which in effect has led to a large missing category. Since we consider basement as one of the most important factors an attempt to decrease the amount of exposure in the missing category and hence increase the reliability of the non-missing factor levels estimates is undertaken. It is most likely that a single policyholder only possess one building with a certain sized living area. If the same policyholder has a policies for multiple periods whereof at least one has a non-missing basement variable it is probable that variable level value should apply to both earlier and later periods with missing values. First off, we retain the non-missing values by assuming that given a non-missing basement value, all earlier periods with missing level classification has the same level value. Second, assuming that the reverse relationship applies. The result of our efforts can be seen in Table. 2, which also has been depicted in Figure. 2.

Table 2: The amount of exposure in the new and old basement variable.

	Basement	No basement	Split-level	Missing
New variable	445 831	572 653	36 888	191 394
Old variable	226 011	291 051	16 265	713 440
Category exposure (new)	36%	46%	3%	15%

We manage to decrease the missing category by more than 500'000 policy years, mainly because there are a great number of loyal customers. Another dimension in need to be account for in the basement variable is how much of the exposure that is distributed into the missing category by year. The missing category should be a fairly good mixture of the three non-missing level values. The development of the amount of known exposure together with yearly claim frequency estimates can be seen in Figure. 3. We can conclude that the average claim frequency increases at the same time as the missing category gets relatively larger, which is a result of new business that is signed in periods where specification of the variable is not required.

Size of living area, a discrete variable adopting values in the range $[0, \infty)$, with the largest observed value of 800m^2. However, the exposure is very limited for buildings larger than 300m^2, we consequently group all policies with a larger living area than 300m^2 into one category, all values below are divided into equal sized classes of 50m^2 intervals.

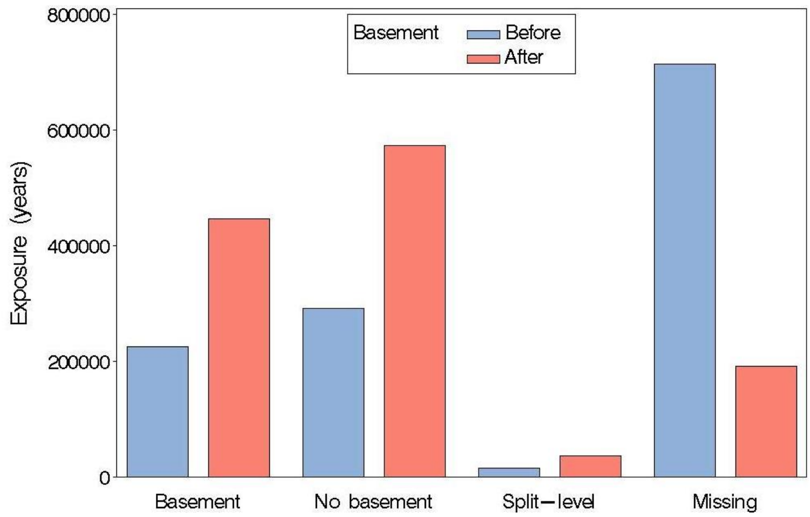

Figure 2: Bar chart comparison of new and old basement variable for the years 2004 – 2010.

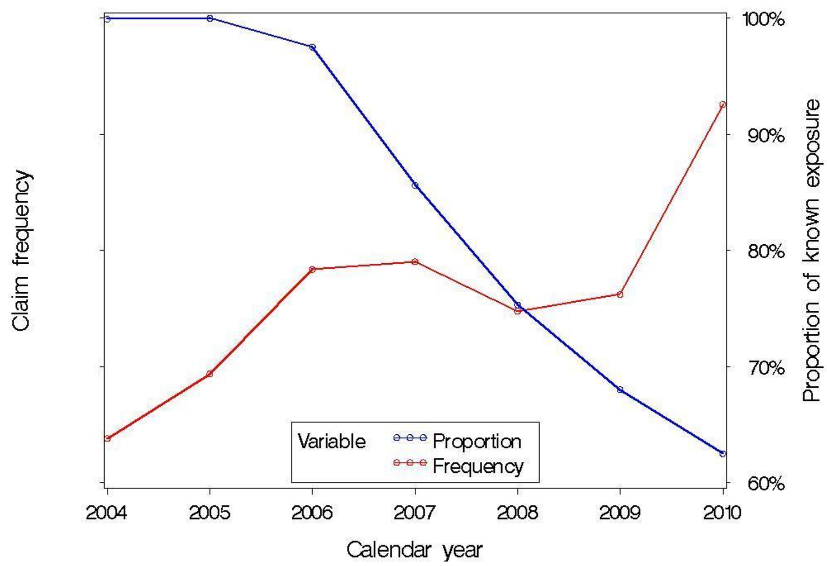

Figure 3: Relationship between claim frequency development and the proportion of known exposure for the new basement variable, 2004 – 2010.

The number of buildings is also a natural factor influencing the risk of obtaining a water-damage of some kind. This factor adopt discrete values between 1 and 7. whereas policies classified into group seven are the ones with the number of buildings ≥ 7. Group seven only constitutes of 860 policy years and is therefore merged together with the sixth buildings class.
The initial classification is quite rough as our variables adopt a wide range of values with varying level of exposure. The parameter dimension needs to be reduced in order for us to obtain reasonably valid estimates.

By using above factors we construct a multiplicative GLM model, at first for claim frequency, assuming a Poisson distribution for the number of claims X_{ij}, and in the terminology of [10] a Relative-Poisson distribution for the key ratio $\frac{X_{ij}}{w_{ij}}$, with a log-link and the dispersion parameter fixed to $\phi = 1$.

$$\mu_i = \mu_0 \gamma_i^1 \dots \gamma_i^m, \quad i = 1, ..., N$$

Here, i is the classification level of a combination of the m explanatory variables and N is the number of levels. For each factor, the class level with the most exposure is appointed as a base level i.e. $\gamma^k_{\text{BASE}} = 1$, since it is the level we can determine with the highest precision. The other estimates will therefore be determined in relation to the base level. We estimate the parameters in the initial model and examine the results by plotting the multiplicative relativities together with their corresponding 95% Wald confidence intervals. The first noticeable discrepancy from what we would expect is a larger estimate for the basement factors missing category than of the others. A probable explanation is that it is an effect of calendar year, since the proportion of known exposure is lower for years with higher claim frequency, which is supported by Figure 3. In the same figure we can observe an apparent increasing trend in the frequencies by year which suggest that a calendar year effect should be included. For both age variables there seem to be a threshold at the ages > 60 and > 45 respectively, which therefore are classified as the highest categories. Building age categories $(5, 10], (10, 15]$ and $(15, 20], (20, 25]$ have overlapping 95% confidence intervals and are grouped into two categories. As for the building size, the estimates also levels out at the highest levels together with a relatively small amount of exposure and overlapping confidence intervals above 200m^2. We reclassify the data and once again re-fit the model including a calendar year effect. By doing so the distribution of exposures amongst classification levels will change and hence also the base levels, as well as the relatively determined estimates. The model is re-fitted and we conduct a likelihood ratio test of type III, we denote the additive factors in the log-link by λ_j^A, where A refers to the factor and j to the level.

Table 3: Likelihood ratio test - Type III

Rating factor	DF	χ^2-statistic	p-value
Basement	3	374.92	<0.0001
Policyholder age	4	715.56	<0.0001
Size of living area	5	446.53	<0.0001
Number of buildings	2	16.61	0.0002
Building age	7	679.72	<0.0001
Calendar year	6	588.99	<0.0001

$$H_0: \quad \lambda_j^A = 0 \quad \text{for all} \quad j$$
$$H_A: \quad \text{At least one} \quad \lambda_j^A \neq 0, j = 1, ..., J$$

All factor effects turns out as significant at a 0.1% level, in fact almost all parameter estimates except for calendar year 2007, number of buildings equal to 1 and building age $[35, 40)$ are significant different from the base level value at a 5% confidence level. The estimates corresponding to the re-fitted model is depicted in Figure 4 – 6 below. The plotted estimates are the multiplicative relativities with corresponding 95% Wald confidence intervals.

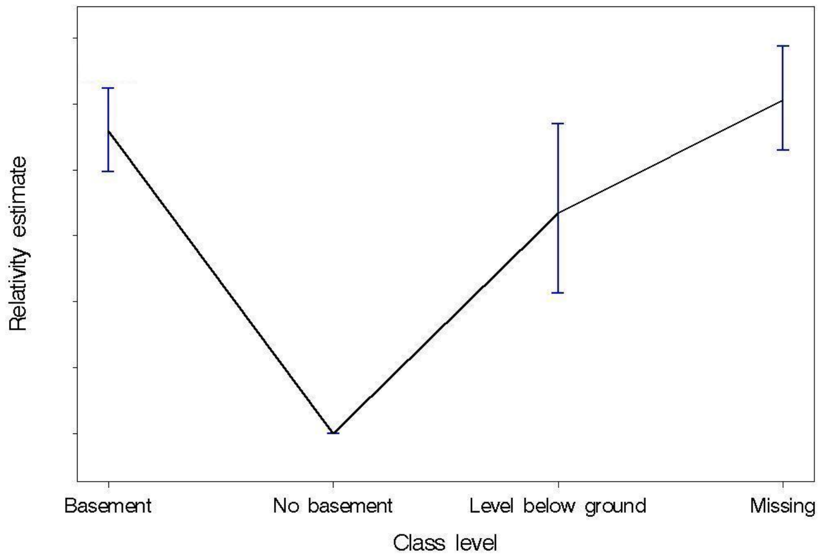

Figure 4: Relativity estimates for the basement factor in the final model.

The inclusion of a calendar year factor seem to have adjusted the missing basement category estimate to a more reasonable value, but it is still a bit larger than the no-basement category. Size of living area is the factor adopting the widest range of relativity estimates ranging from 0.62 to 1.95 which is distributed over six parameters. This might be considered a bit small since the difference between categories are large, on average 0.22 units. We could also consider fitting linear effects for both age variables and living area size in an attempt to reduce the number of parameters. Since the difference between levels is large and there are not a strictly linear relationship between parameter estimates we would be at risk for under-fitting, especially when the last category weighs heavily. The only real surprise occurs in the number of buildings factor, even if small, given the level of all other variables the effect of the number of buildings turn out as decreasing in claim frequency. Turning to the calendar year effect, we can conclude

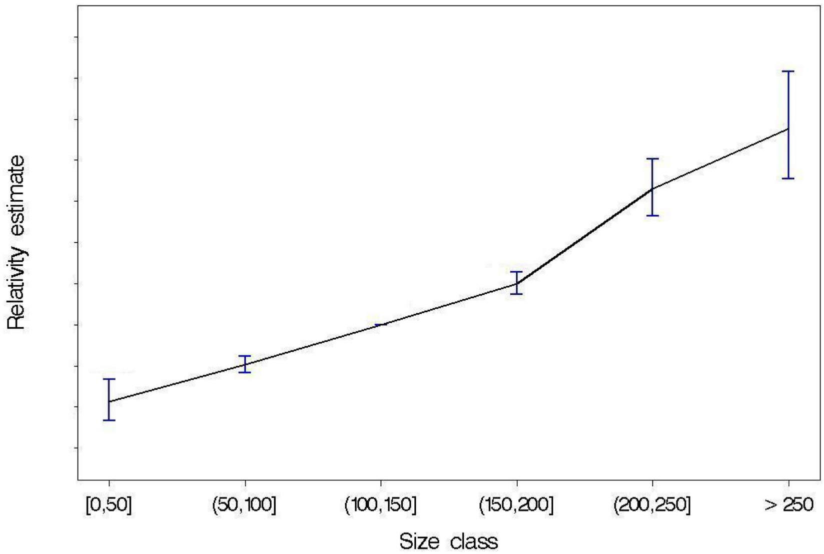

Figure 5: Relativity estimates for size of living area in the final model.

that there has been a large inflation in the number of claims during the past seven years. If we had used the period 2004 – 2009 as model sample and 2010 as model checking period any naive method of estimating the effect of 2010 e.g. fitting a linear trend or assuming that the effect of 2010 would be the same as 2009, would have failed.

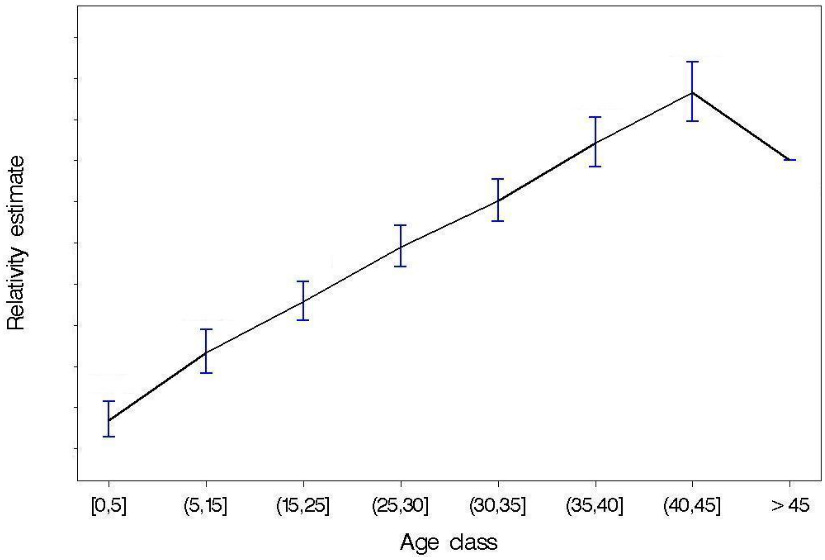

Figure 6: Relativity estimates for building age in the final model.

2.2 Model fit

We investigate the model fit at municipality level by comparing the actual and predicted number of claims. The majority of municipalities do fit quite well as 90% of the observations only deviate between -24.7 to 14.4 cases considering that it is for an amount of exposure corresponding to six years, i.e. averaging about -4.1 to 2.4 cases of yearly deviation. However, there are municipalities that have a substantially larger deviation. The municipality of Stockholm have by far the most number of observed claims compared to predicted, followed by yet another four municipalities in the Stockholm county. In total Stockholm is represented by nine municipalities among the top twelve most deviating. It should be noted that Stockholm municipality also have the highest amount of exposure, which amount to about 54'000 policy years over the average 3'685 years and the entire county having 291'300 years of exposure. Taking into account the amount exposure in each geographical unit the picture changes. The distribution of deviations are more skewed to the right i.e. most municipalities have a higher predicted than observed counts, Figure 7. Skåne county (county code 12) has three out of the seven most deviating municipalities and also the widest range of deviations. Our hypothesis is that the number of claims in a municipality is Poisson distributed, with mean value given by the weighted mean values of the individual policies, i.e. the aggregated expected value of the municipality. Denoting individual k with classification level j in municipality i by

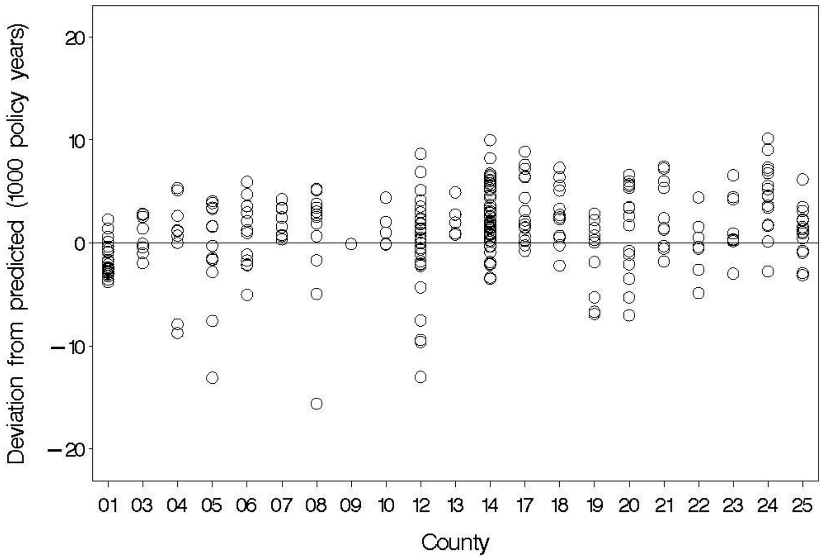

Figure 7: Deviation from predicted in each municipality per 1000 policy years divided into counties by code.

μ_{ijk} and μ_i for the aggregated municipality claim frequency.

$$H_0: \quad \mu_i = \sum_{j=1}^{n_i} \left(\sum_{k=1}^{n_{ij}} \frac{w_{ijk}\mu_{ijk}}{w_{ij\cdot}} \right) \frac{w_{ij\cdot}}{w_{i\cdot\cdot}}$$

$$H_0: \quad \mu_i \neq \sum_{j=1}^{n_i} \left(\sum_{k=1}^{n_{ij}} \frac{w_{ijk}\mu_{ijk}}{w_{ij\cdot}} \right) \frac{w_{ij\cdot}}{w_{i\cdot\cdot}}$$

In order to make a geographical evaluation of where the model does not fit especially well, we investigate which municipalities has an observed count larger than predicted by comparing to a Poisson distribution, $X_i \sim \text{Po}(w_{i\cdot\cdot}\mu_i)$. The municipality claim frequency is estimated by a weighted average of the predicted values in our model. We plot the municipalities that deviate more

Table 4: Tail probabilities for the observed counts for the model sample.

	Tail probabilities			
	5% <	2.5% <	1% <	0.5% <
Number of municipalities	73	54	37	32
Percent of total	25%	19%	13%	11%

than expected onto a map. The resulting figure does not exhibit any strong distinguishable patterns. Applying our model onto the control sample we predict that there would occur 2211 cases of water-damage compared to the actual count of 2218. When comparing the number of deviating municipal-

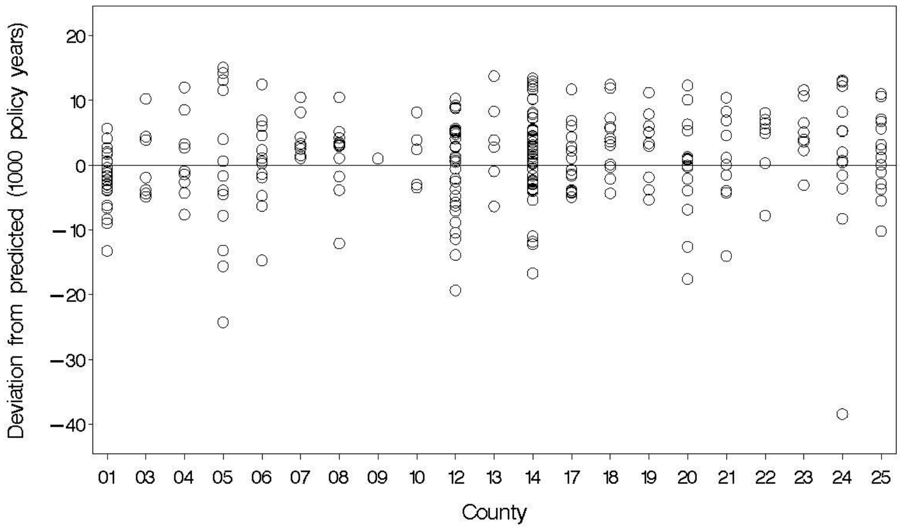

Figure 8: Deviation of observed number of counts from predicted by municipality in the control sample.

ities in the model and control sample we should remember that the model sample has six times as much exposure. If the confidence limits grows at a slower rate than the deviation from predicted then we will obtain more deviant municipalities as a function of more exposure. Matching the deviant

Table 5: Tail probabilities for the observed counts for the control sample.

	Tail probabilities				
	$< 5\%$	$< 2.5\%$	$< 1\%$	$< 0.5\%$	5% /290
Number of municipalities	36	17	11	4	0
Percent of total	12%	6%	4%	1%	0%

municipalities from the model and control sample at all levels, we obtain 14 matches of which four are from Stockholm county and three from Malmö. We should keep in mind that this is not a good way of conducting multiple hypotheses tests simultaneously. If we carry out n independent tests, each at level p we get a high false discovery rate. Introducing a binary variable for each hypothesis,

$$X_i = \begin{cases} 1, & \text{if test } i \text{ significant with probability } p \\ 0, & \text{if test } i \text{ non-significant with probability } 1 - p \end{cases}$$

and $Y = \sum_i^n X_i$. We obtain the probability of observing at leas one significant resultas

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1 - p)^n = \left\{ \begin{array}{c} p = 0.05 \\ n = 10 \end{array} \right\} = 40.1\%.$$

Since we are conducting 290 simultaneous tests the probability is approaching 1 and the probability of finding at least 36 significant 0, i.e. even if all hypotheses were true the probability of observing at least one significant would be 1. To control for the type-I error we can use Bonferroni correction and by adjusting the significance level of each test to 0.05/290 we obtain $P(\text{At least one significant}) = 0.0487$. None of the observed municipality counts are significant different from their predicted mean values on an overall level of 5%. The model has also a reasonably good fit when comparing to the estimated frequencies aggregated over rating factors and rescaling, see Figure 27 – 32.

3 Credibility theory

Credibility theory is a Bayesian approach to describing heterogeneous collectives, method for combining individual and collective experience used for experience rating and estimation of multi level factors. Credibility theory is an important tool in an actuary's toolbox, not only within pricing application but also in reserving. It can for example be used for smoothing over accident years, for using individual rather than aggregated data and share information among a number of related run-off triangles.

A factor with a high number of levels of which some only have such a small amount of recorded data that it would not be reasonable to include in a GLM-model is called a multi level factor (MLF) [10]. In our case the estimation of a geographical factor, where the levels might be given by individual geographical units, is a prime example of MLF-estimation. Using either postal code, parishes or municipalities of which there are more than 9000, about 1800 and 290 respectively, we would as a result have very sparse data in some tariff cells and hence very unreliable estimates. A question that then arises is whether it is possible and in that case how we could manage to estimate the effect of a single geographical unit with limited amount of data. An answer would be using credibility theory following the trail paved by Hans Bühlmann. The following section is primarily based on the book [2], but also [10] which among other things describe the recursive estimation procedure used.

3.1 Bayesian credibility

For a set of contracts and a particular kind of insurance, every individual/risk in this collective would be characterized by its own risk profile $\nu \in \Theta$, where Θ is the set of all risk profiles. The collective risk profiles could be described in a probabilistic way by a probability distribution $u(\nu)$, called structural function of the collective. Because some individuals have a higher propensity of acquiring claims, every individual should then as well have a premium reflecting his specific risk profile, but the quantity ν are to us hidden.

Definition 3.1. The correct individual premium of a risk with risk profile ν,

$$P^{Ind}(\nu) = E[X_{n+1}|\nu] = \mu(\nu) \tag{10}$$

where X_{n+1} is the claim amount year $n+1$. □

And the expected premium for the collective is then calculated using above and the structural function.

Definition 3.2. The collective premium is given by,

$$P^{Coll} = \int_{\nu\in\Theta} \mu(\nu)dU(\nu) =: \mu_0 \tag{11}$$

□

If we do not have access to any historical claims data we cannot distinguish between individual risk propensities, i.e. prior collection of data, our belief of different risk profiles and hence premiums would be equal. In terms of Bayesian statistics ν can be thought of as a realization of a random variable Θ. We are out to find the correct premium for each risk. In the search for such an estimator we define a optimality criterion that has to be fulfilled, minimizing the mean squared error of prediction [10] or in the terminology of [2] with respect to the quadratic loss function.

$$E\left[(\widehat{\mu(\Theta)} - \mu(\Theta))^2\right] \tag{12}$$

Applying this criteria and estimating premium such that the quadratic loss is minimized and hence the best premium given by the posterior expectation of $\mu(\Theta)$ named the Bayes premium, where the parameter Θ is written as an upper case letter to stress that it is a random variable.

$$P^{Bayes} = E[\mu(\Theta)|\mathbf{X}] = \widetilde{\mu(\Theta)} \tag{13}$$

The Bayes premium is *the best experience premium* and in estimating this quantity we need,

(1) the prior distribution

(2) the conditional distribution (sampling distribution)

- The random variables $X_i,\ i = 1, 2, ...$ are assumed to be conditionally i.i.d. given $\Theta = \nu$.

Given that our sampling distribution belongs to the exponential dispersion family $\mathcal{F}_{exp}^{b,c}$ with specified $b(\cdot)$ and $c(\cdot)$ functions we get some very handy properties if choosing it's conjugate prior distribution with hyper parameters x_0, τ_0 and the same $b(\cdot)$ function.

Theorem 3.1. *For the family of $\mathcal{F}_{exp}^{b,c}$ and it's conjugate prior family $\mathcal{U}_{exp}^{b}$ we have,*

(i) $P^{ind}(\nu) = E[X_j|\Theta = \nu] = \mu(\nu) = b'(\nu)$ *and* $Var(X_j|\Theta = \nu) = \frac{\phi b''(\nu)}{w_j}$

If the region Θ is such that $\exp\{x_0\nu - b(\nu)\}$ disappears on the boundary of Θ for each possible value x_0, then we have

(ii) $P^{Coll} = x_0$

(iii) $P^{Bayes} = \alpha\overline{X} + (1-\alpha)P^{Coll}$ *where* $\overline{X} = \sum_j \frac{w_j}{w.} X_j \quad \alpha = \frac{w.}{w. + \sigma^2/\tau^2}$

□

3.2 Credibility estimators

The Bayes premium minimizes the quadratic loss function but has the disadvantage of not necessarily being given by a closed form expression. This quantity might have to be calculated by computer intensive methods as MCMC, which can be a bit tedious. Moreover, calculation of a posterior distribution involves specification of a prior, which can lead to erroneous results and is also the main critic against any Bayesian approach. Instead, we could focus our attention on those estimators of $\mu(\Theta)$ that are linear in the observations, the so called credibility estimators (denoted by double hats).

$$P^{Cred} = \widehat{\widehat{\mu(\Theta)}} = \hat{a}_0 + \sum_{j=0}^{n} \hat{a}_j X_j \tag{14}$$

Assumptions 3.1.

(i) The variables $X_j|\Theta = \nu$ are independent with distribution function F_ν and moments $\mu(\nu) = E[X_j|\Theta = \nu]$ and $\sigma^2(\nu) = Var(X_j|\Theta = \nu)$

(ii) Θ is a random variable with distribution $U(\nu)$

□

Given the assumptions above we could derive the following results,

$$Cov(\overline{X}, \mu(\Theta)) = Var(\mu(\Theta)) =: \tau^2$$

$$Var(\overline{X}) = \frac{E[\sigma^2(\Theta)]}{n} + Var(\mu(\Theta)) =: \frac{\sigma^2}{n} + \tau^2.$$

Theorem 3.2. *Under assumptions 3.1 the credibility estimate is given by,*

$$\widehat{\widehat{\mu(\Theta)}} = \alpha\overline{X} + (1-\alpha)\mu_0 \quad \textit{where} \quad \mu_0 = E[\mu(\Theta)] \tag{15}$$

and

$$\alpha = \frac{n}{n + \sigma^2/\tau^2} \tag{16}$$

□

In the case when the distribution of $\{X_j\}$ conditional on Θ is an exponential family the credibility estimate would be the Bayes premium. The estimate is a weighted average of the collective premium and the more relevant measure of the individual experience $\overline{X}$. When a Bayesian premium is a credibility estimator, it is referred to as a exact credibility estimate. Furthermore, the factor α and hence the credibility estimate has the properties we would like it to possess.

(i) If we get more data, the number of policy years for each risk increase and more weight is put on the individual experience $\alpha \uparrow$.

(ii) If the within risk profile (group) variance increase $\sigma^2(\nu) \uparrow$, less weight will be put on the individual experience $\alpha \downarrow$ and they get less credible.

(iii) If the between risk profile variance increase $\tau^2 \uparrow$, more weight will be put on the individual experience $\alpha \uparrow$ and they get more credible.

3.3 The Bühlmann model

When we have a data set $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_I)$ where $\mathbf{X}_i = (X_{i1}, X_{i2}, ..., X_{in_i})$ is the experience of group $i = 1, ..., I$, and want to estimate each groups credibility premium $\mu(\Theta_i)$, another set of assumptions are needed.

Assumptions 3.2.

(i) The random variables $X_{ij}|\Theta_i = \nu,\ j = 1, ..., n_i$ are independent with distribution function F_ν and conditional moments,

$$\mu(\nu) = E[X_{ij}|\Theta_i = \nu] \quad \text{and} \quad \sigma^2(\nu) = Var(X_{ij}|\Theta_i = \nu)$$

(ii) The pair of observations $\{(\Theta_i, \mathbf{X}_i),\ i = 1, ..., I\}$ are independent and identically distributed.

□

We want an estimate of $\mu(\Theta_i)$ belonging to the class of estimates that are linear in all observations. Minimizing the quadratic loss function and after some derivations [2] we can conclude that the estimate coincides to the one i equation (15),

$$\widehat{\widehat{\mu(\Theta_i)}} = \alpha \overline{X}_{i\cdot} + (1 - \alpha)\mu_0. \tag{17}$$

An extension to this class of estimators is obtained if we impose the restriction of unbiasedness over collective and homogeneity. The estimator should be homogeneous in the sense that if all variables would be multiplied by a constant then it would be possible to factor out $f(c\mathbf{X}) = c^n f(\mathbf{X})$, which implies that in contrary to the inhomogeneous estimator it does not contain any separate constant terms. The best linear estimator fulfilling $E[\widehat{\mu(\Theta_i)}] = E[\sum_{i,j} b_{ij} X_{ij}] = E\left[\mu(\Theta_i)\right]$ is called a homogeneous credibility estimators. As previously defined the collective premium $E[\mu(\Theta_i)] = \mu_0$, using conditional expectation we have,

$$E[X_{kj}] = E[E[X_{kj}|\Theta_k = \nu]] = E[\mu(\Theta_i)] = \mu_0. \tag{18}$$

This means that changing the constant term in (17) into a linear expression in data yield an unbiased estimator. The most intuitive linear function that reflect the collective experience would be the overall average,

$$\overline{X}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{n_i} X_{ij} \quad \text{where} \quad N = \sum_{i=1}^{I} n_i \tag{19}$$

in conclusion,

$$\widehat{\widehat{\mu(\Theta_i)}}^{hom} = \alpha\overline{X}_{i\cdot} + (1-\alpha)\overline{X}_{\cdot\cdot} \tag{20}$$

3.4 The Bühlmann-Straub Model

Classifying data into tariff cells fulfilling the criteria (1) in section 1.4 of homogeneous within cell risk and considering a credibility estimator of some key ratio. Where in contrast to previous sections our response is a ratio of a sum of random variables and a measure of exposure (volume measure) $X_{ij} = S_{ij}/w_{ij}$, where w_{ij} is referred to as a weight. Since we want to model a weighted variable the variance assumption in Assumptions 3.2 is no longer valid. We should also take into account the amount of exposure, resulting in $Var(X_{ij}|\Theta_i) = \sigma^2(\Theta_i)/w_{ij}$. If X_{ij} conditional on Θ_i belongs to the class of exponential dispersion models the conditional variance would take the form $Var(X_{ij}|\Theta_i) = \phi v(\mu(\Theta_i))/w_{ij}$. Using the standard approach of modeling the number of claims as Poisson and the claim amounts as Gamma distributed, the variance function can explicitly be parameterized as a general Tweedie model $\mu^p(\Theta_i)$.

Assumptions 3.3.

(i) Conditionally given Θ_i, the $\{X_{ij} \mid j = 1, ..., n\}$ are independent with,

$$E[X_{ij}|\Theta_i] = \mu(\Theta_i) \quad \text{and} \quad Var(X_{ij}|\Theta_i) = \frac{\sigma^2(\Theta_i)}{w_{ij}}.$$

(ii) The pairs $\{(\mathbf{X}_i, \Theta_i), i = 1, 2, ...\}$ are independent, and $\Theta_1, \Theta_2, ...$ are independent and identically distributed.

In the description by [10] they explicitly assume that the credibility estimators are multiplicative in the risk categories/profiles, $\mu(\Theta_i) = \mu_0\Theta_i$ where the expected value of $\Theta_i, i = 1, 2, ...$ equals to 1. In contrast to the more general approach of not assuming any functional form of $\mu(\Theta_i)$, both yielding $E[\mu(\Theta_i)] = \mu_0$.

Definition 3.3. For the collective the following quantities are of interest,

$$E[\mu(\Theta_i)] = E[E[X_{ij}|\Theta_i]] = E[X_{ij}] = \mu_0 \tag{21}$$

$$Var(\mu(\Theta_i)) = \tau^2 \tag{22}$$

$$E[\sigma^2(\Theta_i)] = E[w_{ij}Var(X_{ij}|\Theta_i)] = \underbrace{\phi E[v(\mu(\Theta_i))]}_{\text{EDM}} = \underbrace{\phi E[\mu^p(\Theta_i)]}_{\text{Tweedie}} = \sigma^2 \tag{23}$$

□

By the last equality in equation (21) we see that without any knowledge of the individual risk profiles all individuals would have the same expected key ratio.

Theorem 3.3. *The credibility estimator under Assumptions 3.3 is given by,*

$$\widehat{\widehat{\mu(\Theta_i)}} = \alpha_i \overline{X}_{i\cdot} + (1-\alpha_i)\mu_0 \tag{24}$$

where

$$\overline{X}_{i\cdot} = \sum_{j=1}^{n} \frac{w_{ij}}{w_{i\cdot}} X_{ij}, \quad \textit{and} \quad \alpha_i = \frac{w_{i\cdot}}{w_{i\cdot} + \frac{\sigma^2}{\tau^2}}. \tag{25}$$

□

The individual estimators $\overline{X}_{i\cdot}$ are the best linear unbiased estimators (BLUE) with the smallest variance conditional on Θ_i. In the homogeneous case it would be reasonable to believe that is our best choice of estimator for μ_0 is the overall average, $\widehat{\mu_0} = \overline{X}_{\cdot\cdot} = \sum_i \frac{w_{i\cdot}}{w_{\cdot\cdot}} \overline{X}_{i\cdot}$ which is used in [10]. In fact the best homogeneous estimator of μ_0 is $\sum_i \frac{\alpha_i}{\alpha_\cdot} \overline{X}_{i\cdot}$, which is formulated in the following theorem.

Theorem 3.4. *The homogenous credibility estimator of $\mu(\Theta_i)$ in the Bühlmann-Straub model given Assumptions 3.3 is,*

$$\widehat{\widehat{\mu(\Theta_i)}}^{hom} = \alpha_i \overline{X}_{i\cdot} + (1-\alpha_i)\widehat{\widehat{\mu}}_0 \tag{26}$$

where

$$\widehat{\widehat{\mu}}_0 = \sum_i \frac{\alpha_i}{\alpha_\cdot} \overline{X}_{i\cdot} \quad \textit{where} \quad \alpha_i = \frac{w_{i\cdot}}{w_{i\cdot} + \frac{\sigma^2}{\tau}} \tag{27}$$

□

In order to obtain a credibility estimator we have to estimate the parameters μ_0, σ^2 and τ^2. By choosing to use the homogeneous estimator, the problem is reduced to estimate the two latter as the estimate of μ_0 is built in. By Definition 3.3 and equation (23) σ^2 is the expected within group variance, an unbiased estimator of the variance within group i is,

$$\hat{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} w_{ij}(X_{ij} - \overline{X}_{i\cdot})^2 \tag{28}$$

and

$$E[\sigma_i^2] = E[E[\hat{\sigma}_i^2|\Theta_i]] = E[\sigma^2(\Theta_i)] = \sigma^2.$$

The next step would be to weight the group variance estimates into an overall unbiased estimate. As suggested by [2], we could assume that the observations are normally distributed and therefore use an arithmetic mean over group variances. A more appealing approach that is more in line with the "credibility theory intuition" and do not involve a unjustified assumption is to weight the group variance estimates by degrees of freedom [10] which also yield a unbiased estimate.

$$\hat{\sigma}^2 = \frac{\sum_i (n_i - 1)\hat{\sigma}_i^2}{\sum_i (n_i - 1)} = \frac{\sum_i \sum_j w_{ij}(X_{ij} - \overline{X}_{i\cdot})^2}{\sum_i (n_i - 1)} \tag{29}$$

Turning to τ^2, an unbiased and consistent estimate is given by the following expression,

$$\hat{\tau}^2 = \frac{\sum_{i=1}^{I} w_{i\cdot}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2 - (I-1)\hat{\sigma}^2}{w_{\cdot\cdot} - \sum_{i=1}^{I} w_{i\cdot}^2 / w_{\cdot\cdot}}. \tag{30}$$

We are now able to estimate the risk of obtaining a water-damage in each

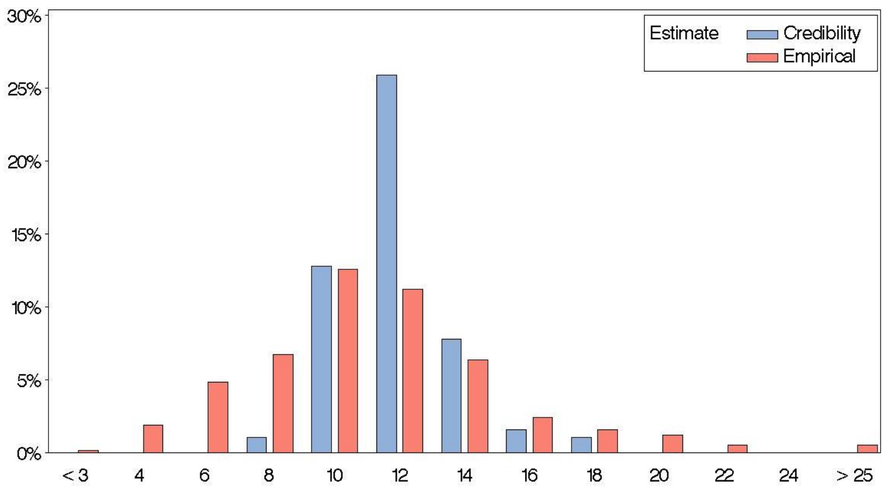

Figure 9: Histogram of the empirical and credibility claim frequency estimates per 1000 policy years.

municipality taking into account how credible the individual estimates are. Using equation (15) for claim frequency we obtain the results tabulated in Table 6, sorted descending in credibility factor.

Table 6: Pure credibility at municipality level with estimates in terms of claim frequency per 1000 policy years and an overall average of 12.42.

Municipality	Exposure	Claims	Credibility factor	Credibility estimate	Empirical frequency
Stockholm	54 094	926	0,963	16,9	17,1
Göteborg	37 610	511	0,947	13,5	13,6
Västerås	20 016	226	0,905	11,4	11,3
Huddinge	19 978	300	0,905	14,8	15,0
Nacka	19 090	292	0,901	15,0	15,3
⋮		⋮	⋮		⋮
Aneby	354	7	0,145	13,5	19,8
Boxholm	313	8	0,130	14,1	25,6
Ödeshög	270	4	0,114	12,7	14,8
Bjurholm	232	2	0,100	12,0	8,6
Ydre	220	3	0,095	12,5	13,6

3.5 Multiplicative model

As previously mentioned an assumption of a multiplicative structure of $\mu(\Theta_i)$ in the individual risk profiles Θ_i could be imposed, $\mu(\Theta_i) = \mu_0\Theta_i$. Since our prime interest is focused on the random vairables Θ_i, of which an estimate is obtained by dividing $\widehat{\mu(\Theta_i)}$ by μ_0.

$$\widehat{\mu(\Theta_i)} = \mu_0\widehat{\Theta_i} = \mu_0\left(\alpha_i\frac{\overline{X}_{i\cdot}}{\mu_0} + (1-\alpha_i)\right) \tag{31}$$

A reasonable approach in insurance practice would be to use a multiplicative model for a set of ordinary rating factors $\{\gamma_i^k\}, i = 1, ..., n, k = 1, ..., m$, where γ_i^k is relativity k for risk classification level i i.e. the risk classes are given by $\gamma_i = \gamma_i^1 \cdots \gamma_i^m$. Implying that conditional on the individual risk profiles our model takes the form

$$E[X_{ijt}|\Theta_j] = \gamma_i\mu(\Theta_j) \qquad E[X_{ijt}] = \gamma_i E[\mu(\Theta_j)] = \mu_0\gamma_i. \tag{32}$$

X_{ijt} is the key ratio of random variable S_{ijt} with volume measure w_{ijt} where i is risk classification level belonging to group j of which it is the t'th observation. Because we have introduced rating factors into the model specification the parameter μ_0 is the expected value of the key-ratio given that the rating factors adopt the base level value (=1). Once again assuming that the individual risk profiles are multiplicative in the ordinary rating factors

$$E[X_{ijt}|\Theta_j] = \mu_0\gamma_i\Theta_j \quad \text{implying} \quad E[\Theta_j] = 1. \tag{33}$$

As motivation for the variance structure [10] assume that the conditional key-ratios follows a Tweedie distribution and hence also exponential dispersion family, yielding

$$Var(X_{ijt}|\Theta_j) = \frac{\phi v(\gamma_i \mu(\Theta_j))}{w_{ijt}} = \frac{\phi \gamma_i^p \mu(\Theta_j)^p}{w_{ijt}} \tag{34}$$

and

$$E[Var(X_{ijt}|\Theta_j)] = \frac{\gamma_i^p \sigma^2}{w_{ijt}} \quad \Rightarrow \quad E[\phi\mu(\Theta_j)^p] = \sigma^2. \tag{35}$$

We continue by redefining our random variables X_{ijt} and weights w_{ijt} as follows.

Definition 3.4. If the random variables X_{ijt} conditional on Θ_j follows a Tweedie type of distribution where the mean value is multiplicative in the factors $\gamma_i = \gamma_i^1 \ldots \gamma_i^m$ the transformed weights and variables are,

$$\widetilde{X}_{ijt} := \frac{X_{ijt}}{\gamma_i} \quad \text{with weights} \quad \widetilde{w}_{ijt} = w_{ijt}\gamma_i^{2-p}.$$

□

By Definition 3.4 we get,

$$E[\widetilde{X}_{ijt}|\Theta_j] = \frac{E[X_{ijt}|\Theta_j]}{\gamma_i} = \mu(\Theta_i) \tag{36}$$

and

$$Var(\widetilde{X}_{ijt}|\Theta_j) = \frac{\phi\gamma_i^p \mu(\Theta_j)^p}{\gamma_i^2} = \frac{\phi\mu(\Theta_j)^p}{\widetilde{w}_{ijt}}. \tag{37}$$

Applying the newly defined variables together with conditional expectation and variance in equations (36) and (37) onto Assumption 3.3 the credibility estimator in Theorem 3.3 follows. Estimates of the variance parameters are obtained by applying the transformed variables instead of the original into equations (28), (29) and (30).

We calculate the credibility estimates in the multiplicative model described in section 2.1 recursively with the backfitting algorithm in Appendix B.1 [10]. After twenty iterations the algorithm has converged with a largest absolute difference of 0.000079. Sorted by credibility factor the estimates is tabulated in Table 7. Calculating the credibility estimates in this way implies that we believe in the hypothesis that all residual variation in the multiplicative model is a pure effect of territory. Creating territories by simply classifying estimates into classes by level of estimate and including in a GLM, we would obtain a very good fit to the data but this need not necessarily be a good territorial division. The territories should reflect a higher geographical level of risk and should be tested on a separate data set. A good idea would also be to investigate the territorial robustness by year.

Table 7: Credibility estimates of municipality risk in multiplicative model specified in section 2.1.

Municipality	Exposure	Claims	Credibility factor	Credibility estimate
Stockholm	54094	929	0.967	1.24
Göteborg	37610	511	0.948	1.10
Huddinge	19978	300	0.901	1.29
Nacka	19090	292	0.900	1.26
Västerås	20016	226	0.894	1.06
⋮		⋮		⋮
Grästorp	392	2	0.145	0.92
Boxholm	313	8	0.128	1.15
Ydre	220	3	0.113	0.99
Ödeshög	270	4	0.110	1.03
Bjurholm	232	2	0.096	0.98

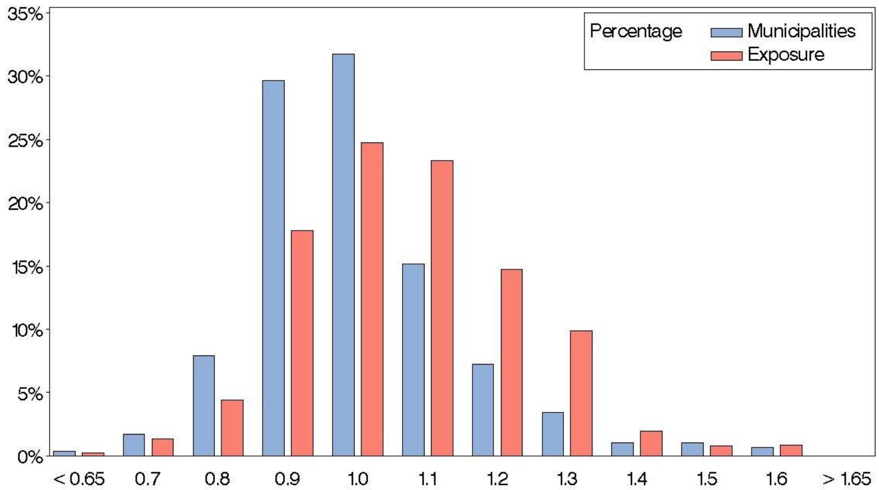

Figure 10: Histogram of the credibility estimates in the multiplicative model.

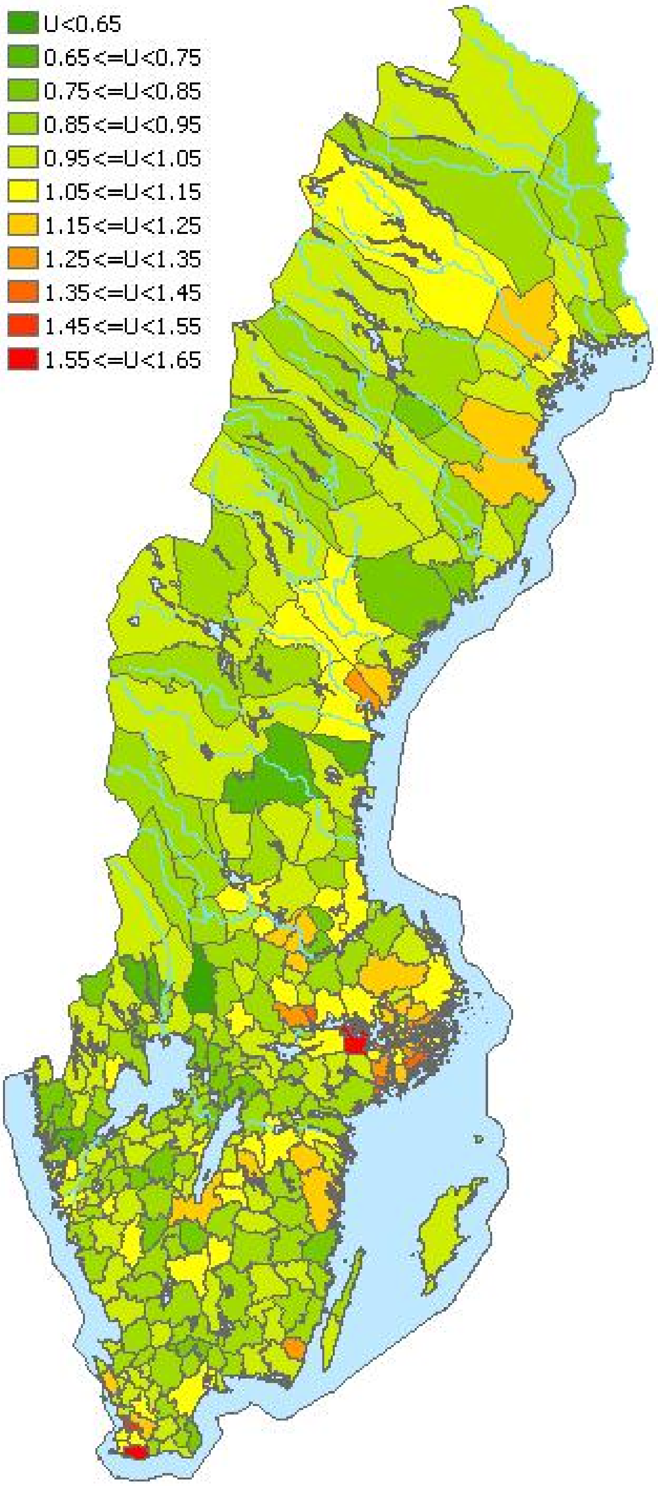

Figure 11: Credibility estimates in the multiplicative model

4 Smoothing techniques

The primary purpose of using any smoothing technique regardless if it is based on neighboring or the distance between the geographical units is to determine elevation in risk – "risk-terraces", which define the boundaries for geographical territories. By using data outside a individual unit we assume that the local conditions could be estimated by nearby units. It is reasonable to assume that adjacent areas, opposed to those distant, display similar geographical and meteorological characteristics and hence are more alike. In some cases even the surrounding social and physical conditions influence the local risk. When it comes to pure credibility estimation, this implies that we might consider an alternative weighting scheme for the estimator of μ_0 than proposed in Theorem 3.4 and equation (27).

4.1 Adjacency smoothing

4.1.1 First order

There is very sparse literature on the subject of adjacency-smoothing, therefore we propose a simple but yet effective method employing a bulls eye approach. Smoothing by rings of adjacent municipalities, acquiring a moving average of municipality risk – the smoothed estimates, Figure. 12.

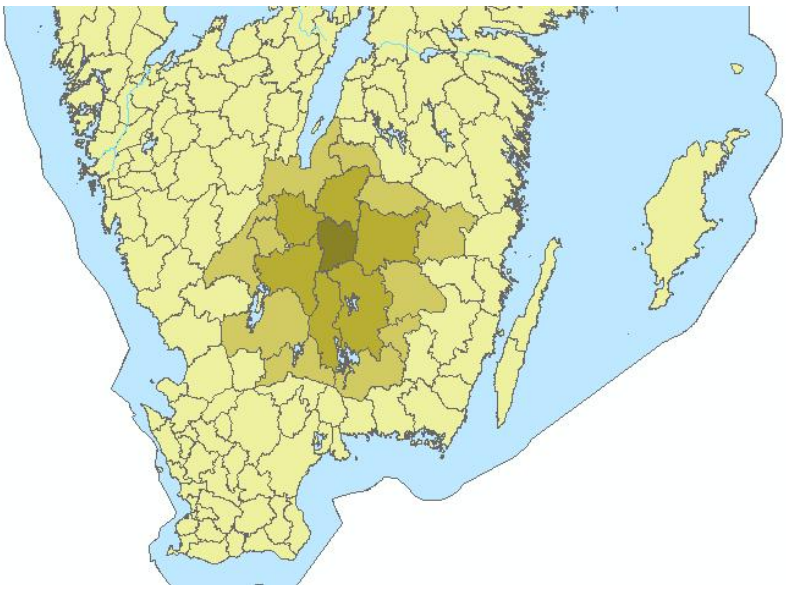

Figure 12: First and second order adjacent units for the municipality Sävsjö.

The smoothing weights are subjectively determined based on how similar we believe adjacent units are and how smooth the risk-terraces should be. If all weight is put onto the primary unit we obtain the least smoothed risk

Unit weight		
Primary	1st order adjacent	Number of adjacent units
α	$\beta^{(1)}$	$n_{adj}^{(1)}$

adjusted surface. The more weight applied onto adjacent units the greater smoothing over risk is achieved. Since the weights should sum to 1 and all first order adjacent units have the same weight we get the relationship,

$$\alpha + \sum_{i=1}^{n_{adj}^{(1)}} \beta^{(1)} = 1 \quad \Leftrightarrow \quad \alpha + n_{adj}^{(1)} \beta^{(1)} = 1. \tag{38}$$

Next, we introduce a parameter $\xi^{(1)}$ determining how much overall weight in relation to the primary unit to apply to all first order adjacent units.

$$\alpha + \xi^{(1)}\alpha = 1 \quad \Leftrightarrow \quad \alpha = \frac{1}{(1+\xi^{(1)})} \quad \text{and} \quad \beta^{(1)} = \frac{1}{n_{adj}^{(1)}} \xi^{(1)} \left(\frac{1}{1+\xi^{(1)}} \right). \tag{39}$$

For example if $\xi^{(1)} = 0$ then $\alpha = 1$, i.e. full weight on the primary unit. Or assuming that the estimates of all adjacent units are as important as the primary one $\xi^{(1)} = 1$, hence $\alpha = 1/2$ and $n_{adj}^{(1)} \beta^{(1)} = 1/2$.

4.1.2 First order conditions

In order for our adjacency weighting scheme to work desirably we need to impose conditions on the α and $\beta^{(j)}$ values based on the number of neighbors. In case of first order adjacency, if a geographical unit do not have any neighbors α need necessarily equal 1 e.g. Gotland municipality. If the unit on the other hand has one neighbor, it would not be reasonable having a $\xi^{(1)} > 1$, since the single neighbor should not be weighted more than the primary unit considered i.e. $\beta^{(1)} \not> \alpha$. For first order adjacency we obtain the following conditions.

$$\alpha = \begin{cases} 1, & \text{if} \quad n_{adj}^{(1)} = 0 \\ 1/(1+\xi^{(1)}), & \text{if} \quad \xi^{(1)} \leq n_{adj}^{(1)} \\ 1/(1+n_{adj}^{(1)}), & \text{if} \quad \xi^{(1)} > n_{adj}^{(1)} \end{cases} \tag{40}$$

$$\beta^{(1)} = \begin{cases} 0, & \text{if} \quad n_{adj}^{(1)} = 0 \\ (1/n_{adj}^{(1)})\xi^{(1)}\alpha, & \text{if} \quad \xi^{(1)} \leq n_{adj}^{(1)} \\ \alpha, & \text{if} \quad \xi^{(1)} > n_{adj}^{(1)} \end{cases} \tag{41}$$

4.1.3 Nth order

We can easily generalize the method to include higher order adjacent units by imposing the weighting scheme below.

Unit weight						
	Adjacent units			Number of adjacent units		
Primary	1st order	...	Nth order	1st order	...	Nth order
α	$\beta^{(1)}$	...	$\beta^{(N)}$	$n_{adj}^{(1)}$	...	$n_{adj}^{(N)}$

$$\alpha + \sum_{i=1}^{n_{adj}^{(1)}} \beta^{(1)} + ... + \sum_{j=1}^{n_{adj}^{(N)}} \beta^{(N)} = \alpha + n_{adj}^{(1)}\beta^{(1)} + ... + n_{adj}^{(N)}\beta^{(N)} = 1$$

Each adjacent unit order has it's own relation parameter to the primary unit, $\gamma^1, ..., \gamma^{(N)}$. The unit weights are therefore given by,

$$\begin{cases} \alpha = \left(\frac{1}{1+\xi^{(1)}+...+\xi^{(N)}}\right) \\ \\ \beta^{(j)} = \frac{1}{n_{adj}^{(1)}}\xi^{(j)}\left(\frac{1}{1+\xi^{(1)}+...+\xi^{(N)}}\right), \quad \text{for} \quad j = 1, ..., N. \end{cases}$$

Furthermore, in consistency with first order adjacency, conditions upon the influence parameters are needed and should fulfill the relationship,

$$\alpha \geq \beta^{(1)} \geq ... \geq \beta^{(N)}.$$

Higher than first order adjacency smoothing is usually not best practice when considering smoothing over geographical risk for units of irregular shape. Standard regional divisions as municipalities, parishes or post codes have units shaped such that some higher order adjacent units can be closer to the primary unit than other of lower order. As a solution, a sensible approach is using the actual distances instead. There are several distances to contemplate, distances between borders, geometric centers, centers of gravity i terms of policy distribution etc. More advanced methods might even take into account that units share common geographical characteristics, such as shoreline to the same lake. In these methods, we can even consider the amount of exposure. One can also construct a smoothing routine with respect to both adjacency as well as distance e.g. first order adjacency smoothing and a distance weighting scheme for the units farther away.

Applying the method on our municipality credibility estimates with $\xi^{(1)}$

equal to 0.5, 1, 2 and 3. The resulting distribution of the estimates is compressed towards its base level value in a larger extent as the values of $\xi^{(1)}$ increase and estimates are smoothed. In Figure 25 we depict the individual credibility in the multiplicative model with overall weight parameter $\xi^{(1)} = 3$, clear risk-terraces can be distinguished and hence potential territories. A distinct feature that can be observed is higher claim frequency level for high populated areas with the highest values in the center of the largest cities Malmö and Stockholm with rings of lower level values for nearby municipalities. Defining a territory as a group of geographically intertwined municipalities at the same classification level. Using this definition we obtain in total 48 territories where the largest territory consist of 47 and the smallest of one municipality. For $\xi^{(1)} > 3$ no great changes in the smoothed values occur in general. The parameterization means that primary units with $\leq$ 3 neighbors weights each first order adjacent units estimates as much as the primary unit.

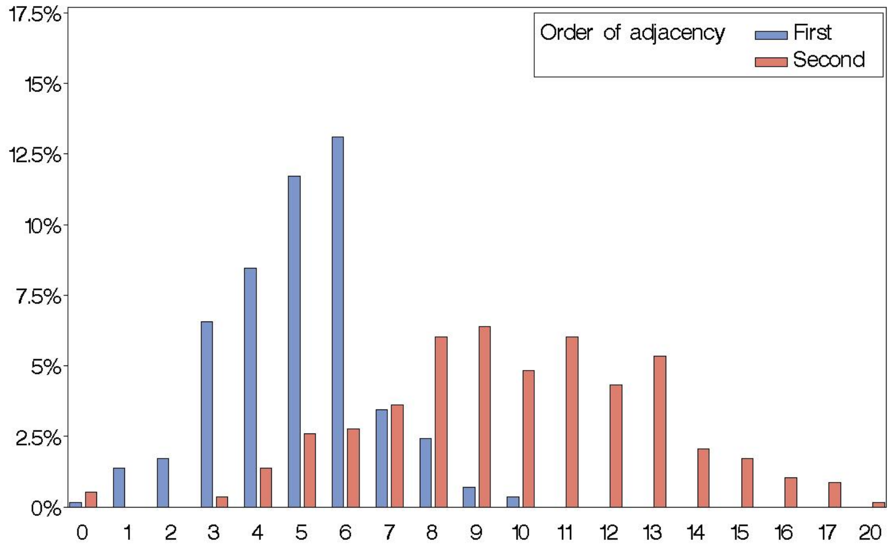

Figure 13: Histogram of the number of adjacent units where the values on the vertical axis is the percentage of all first or second order neighbors.

4.2 Distance smoothing

A similar approach would be to smooth with respect to distance. Appropriately using the Euclidean distance between municipality centroids. If the coordinates for the center of unit i is given by (x_i, y_i) the distance is calculated as $d_{ik} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2}$. Next, we need to establish a weighting scheme. The least complicated and most intuitive would be to use a strictly linear relationship [3],

$$\beta_{ij} = \begin{cases} 1, & \text{if} \quad d_{ij} = 0 \\ \left(d^{max} - d_{ij}\right)/d^{max}, & \text{if} \quad 0 < d_{ij} < d^{max} \\ 0, & \text{if} \quad d_{ij} \geq d^{max} \end{cases}$$

and hence,

$$\widehat{\Theta}_i^{\text{smooth}} = \frac{1}{\beta_{i\cdot}}\widehat{\Theta}_i + \sum_{j \neq i} \frac{\beta_{ij}}{\beta_{i\cdot}}\widehat{\Theta}_j.$$

We set the maximal distance d^{max} to 60km and smooth over the credibility estimates obtained in the multiplicative model, Figure. 15. The result is quite similar in the southern half of the country as in the first order adjacency approach implemented previously. As the municipalities become larger up-county, the number of territories increase which is in line with the homogeneity assumption i.e. that each territory should display homogeneous risks and as the municipalities become larger the less likely are two units to be similar.

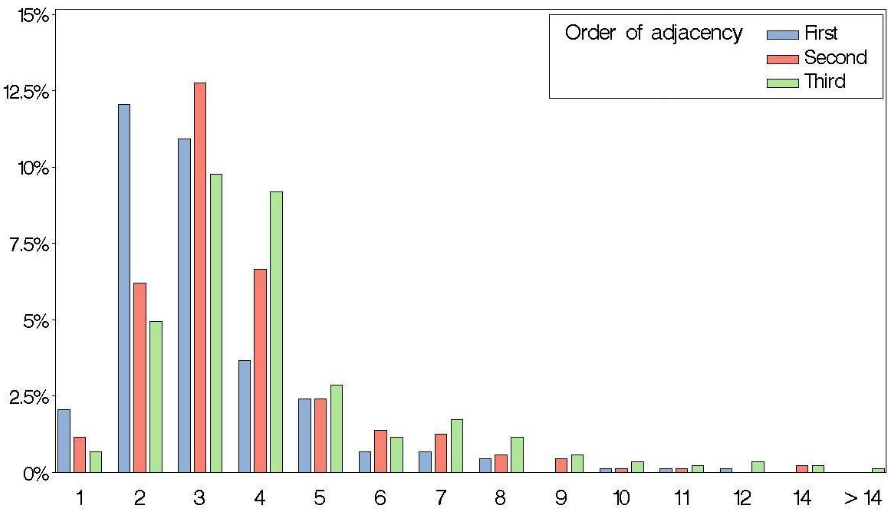

Figure 14: Histogram of the number of first, second and third order adjacent units by distance to centroid.

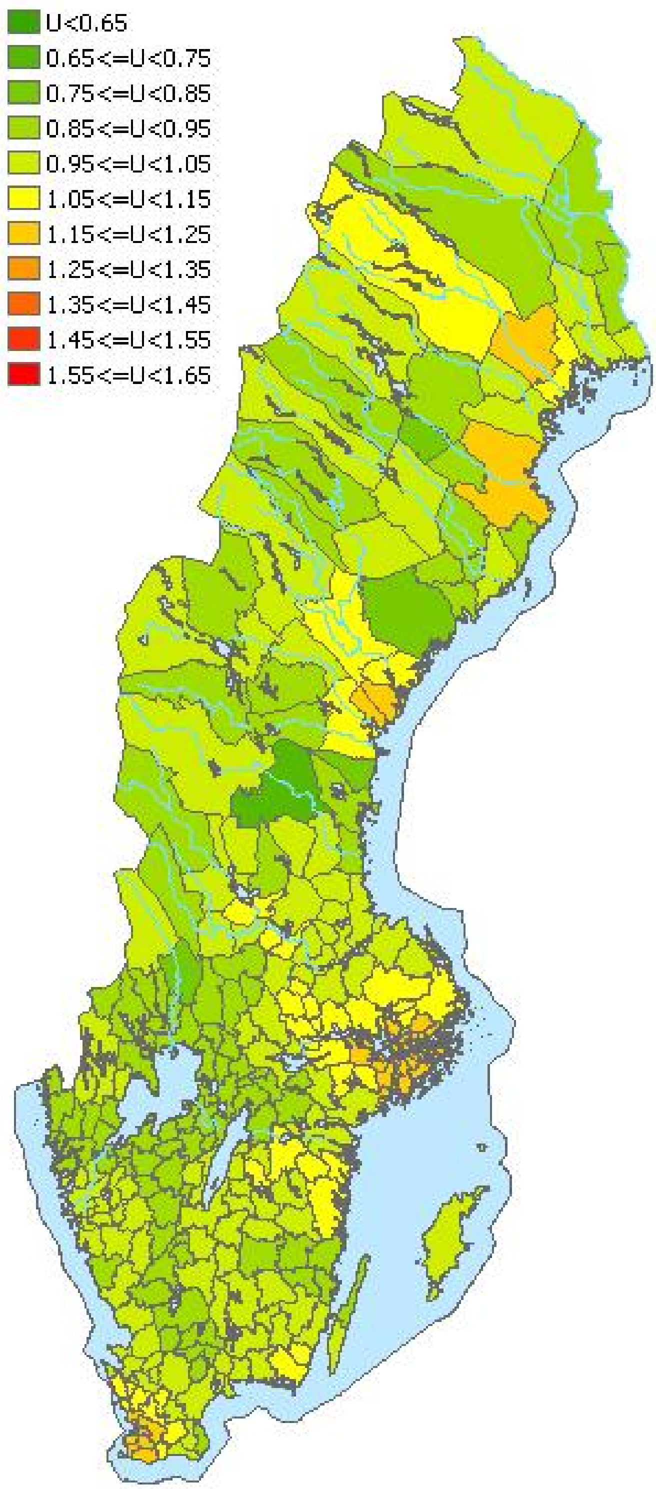

Figure 15: Linear distance smoothed credibility estimates with d^{max} = 60km.

5 Risk classification

In order to differentiate the pricing between individuals with different characteristics into risk classes the author of [5] specifies some guidelines for good practice risk classification.

1. Equity: Premiums should accurately reflect expected losses, benefits and expenses.
2. Homogeneity: Within any class, there should be no subgroups identifiable at reasonable cost that have different expected costs.
3. Intuition: The rating variables should be intuitively related to the loss hazards.
4. Practicality: Classification variables should be objectively measurable and not subject to manipulation by the insured.
5. Incentives: The classification system ideally should provide incentives to reduce risk.
6. Legal: Rating dimensions must be legal and socially acceptable.

A more precise way to put it, is that we should strive for a low level of variability within the risk classes i.e accurate and reliable, and also reflect the cause of loss. Obtaining accurate estimates relies upon number of exposure units and hence the law of large numbers.

5.1 Similarity of risks

Geographical factors that influence the claims experience might be quite different in adjacent areas, e.g. if there is a lake that yearly is overflowing, only affecting residents in a close proximity. We might imagine that the risk within this area is somewhat higher than others a few hundred meters away. Even though it might be difficult to define the risk in terms of distance to the lake at hand, using a small enough geographical unit would permit us to assume that all residents within an area are exposed to the same level of risk. If we actually are aware of the existence of additional factors influencing the claims experience, we should take them into account in the same way as used for risk classification [10]. In a multiplicative model we should estimate the ordinary rating factors γ_i together with the auxiliaries γ_i^{aux}, calculate the unit credibility estimates, consider the auxilary and unit credibility estimates together and classifying them into risk categories by level of risk.

$$\Omega_i = \gamma_i^{aux} \Theta_i. \tag{42}$$

Note that the unit estimate is the residual effect in terms of an additional multiplicative factor when auxiliary is included. Smoothing over the $\hat{\Omega}_i$'s is still advisable.

5.1.1 Auxiliaries

General data on municipality and parish level have been collected from Statistics Sweden and climate data from the Swedish Meteorological and Hydrological Institute gathered form 51 weather stations located around the country. These data could not be incorporated in a tariff since they do not tell us anything about the individual policies. On the other hand, they can be used to describe local conditions and hence the influence on the level of geographical risk.

Precipitation and temperature data containing monthly and yearly records between 1961 – 2009. Precipitation is described in terms of amount in millimeters and temperature in terms of average degrees Celsius. The measurements from each single weather station is only valid for its exact position. To estimate the precipitation and temperature in a municipality for a given month, we calculate the distances between municipality centers and station positions that have a measurement value, choose the value for which the distance to the stations is the smallest, fulfilling a maximal distance requirement and average over all values. In order for us to obtain an estimate for all municipalities we need to set the maximal distance requirement to 110km since there are only 50 stations and all stations does not have a measurement each year. Since we are considering water-damages the amount of precipitation in a region is expected to be highly correlated with the frequency of claims. We proceed by looking at claims and precipitation at a aggregated level by month and year, Figure 16. As expected there are a sim-

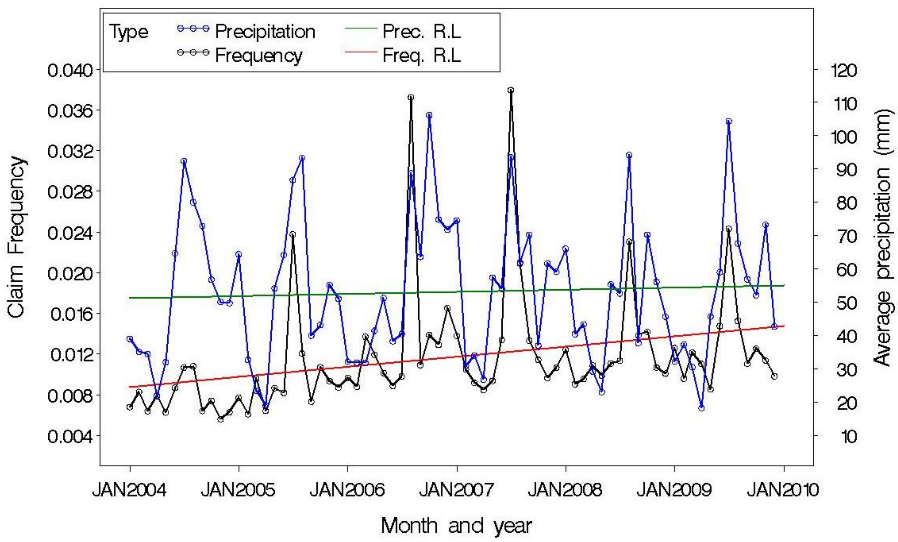

Figure 16: Time-series of claim frequency and amount of precipitation, each together with together with a fitted regression line.

ilar pattern in both series, which are increasing and peaks in July/August,

Figure 17. The trend can be considered as one explanation of the increasing

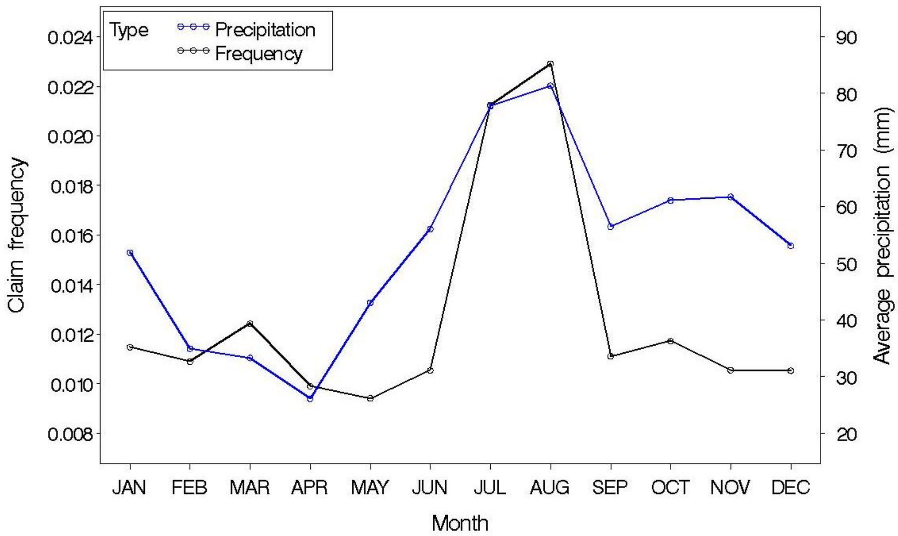

Figure 17: Seasonal variation in precipitation.

calendar year effect seen earlier. We believe that it is the heavy rains in a limited periods of time that drives these kind of claims. With these evidence we construct a normal July/August precipitation variable with values averaged over a period of 15 years for each municipality, 1995 – 2009. At first, we use a maximum distance between municipality center and weather station center corresponding to 110km. This distance is used in order to get at least one measurement value for each municipality and thereafter a condition of 50km which result in a large missing category corresponding to 123 municipalities and 302'000 years of exposure. The estimates vary between 53 and 109mm and are grouped into 12 categories with an interval of 5mm, included in the model the results are not as expected. Rather than an increasing relativity we obtain a slight increase in the first two categories up to $[60, 65)$ and slight decreasing estimates for the values above with some variation. The same pattern is obtained if using average yearly precipitation as well. This deviation from what we expect might be a consequence of not taking into account other important factors in our model or that the approximation is just too rough.

In high populated areas a large amount of concrete is used for construction of buildings and thought to limit the ability of rain-water to flow in its natural directions, which in turn is leading to floods. The systems for handling excess water by municipalities might in some cases be under dimensioned if these initially were built for a smaller number of inhabitants than the actual number of residents as a result of high population growth. We create a population density variable at municipality level with data ob-

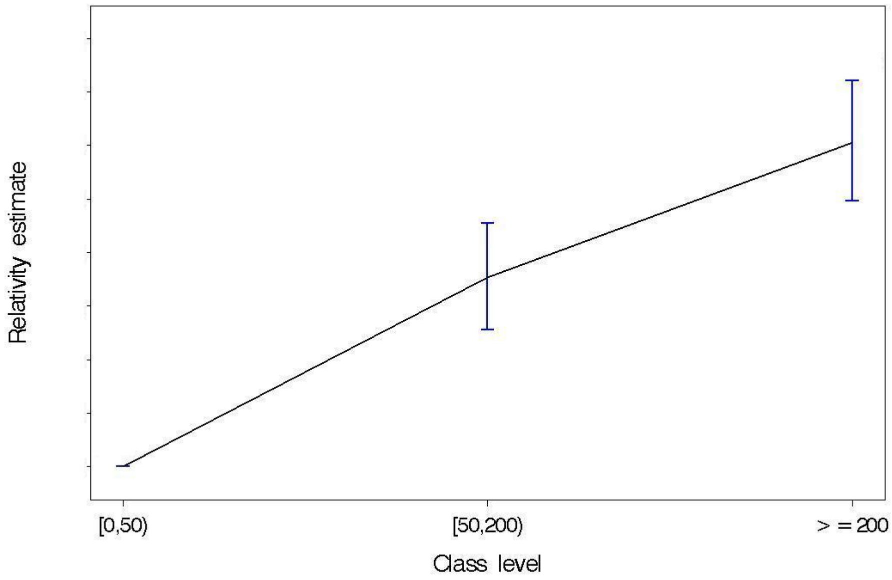

Figure 18: Population density factor at municipality level.

tained from SCB of landarea and population in 2010. The most sparsely populated municipality is Arjeplog in the north with 0.2 persons per km^2 and most densely populated is Stockholm municipality with 4504.3 per km^2. By including the factor in our model we get an increasing relativity estimates with density and a threshold where the population exceeds 200 persons per km^2. A likelihood ratio test type-III for the factor is significant at a 0.0001 level, as a result the small effect in the number of buildings disappear.

We could also consider income, net wealth or average property values, which are all available at municipality level from Statistics Sweden, but these tend to be collinear and we should choose at most one. Buildings with high assessed property values are situated at the most attractive locations, which often are close to water, but do also reflect demand for properties and hence high population density. In general higher value of home or income is hypothesized to be associated with a higher propensity to hire contractors and hence report even the smallest damages. The average taxed income 2010 by municipality varies between 198 and 463 thousand kronor and 90% of the municipalities has an income between 207 and 297. The estimates has a threshold at 300 which correspond to a reasonable income to afford maintaining a house. The trend in income is increasing suggesting that location and population density outweighs disposable income. In fact net wealth and average property values exhibit the same pattern as income when included in the model. If we include both population density and average taxed income the effect of population density is reduced a lot. Also, large corresponding confidence intervals which is obtained, indicating the suspected collinearity.

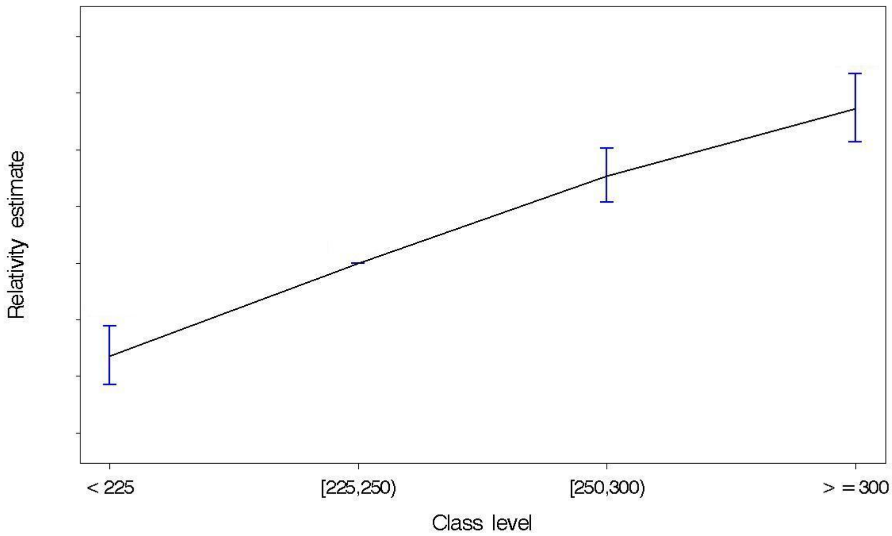

Figure 19: Taxed income factor at municipality level.

Since income to a large extent reflect population density the loss of information is small. We choose only to include average taxed income as a single auxiliary variable for risk classification of municipality credibility estimates.

The credibility estimates is once again estimated with 20 iterations in the Backfitting algorithm. The variation in the credibility estimates is heavily reduced since we have included additional information into the model and hence one explanation of the hypothesized geographical variation, Figure. 20. Smoothing by distance and grouping adjacent units with the same class level value result in 57 closed territories which display differences from the ones obtained by distance smoothing without auxiliaries.

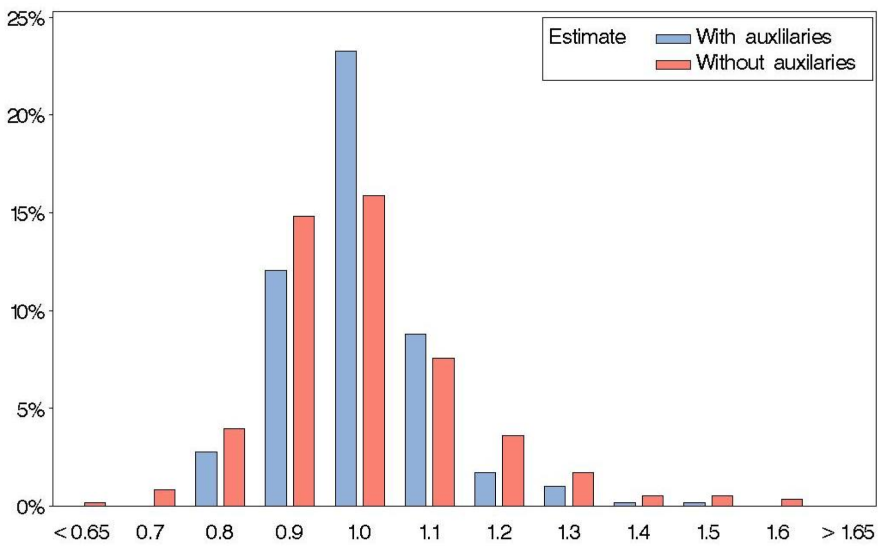

Figure 20: Comparison of the credibility estimates with and without inclusion of taxed income as auxiliary variable, where the vertical axis represent percentage of total exposure.

6 Cluster analysis

Clustering is the process of grouping objects that display a distinguishing pattern into classes, determined by a measure of similarity. A cluster is obtained by grouping observations by their relative distances together with some minimization criterion and perhaps a lower limit of the number of exposure units in each cluster. The two primary reasons for considering clustering in our application are pattern evaluation and reduction of the number of geographical levels. As a result we want to obtain $g \leq N$ homogeneous mutually exclusive classes where individuals within clusters are as similar as possible and individuals between clusters are as different as possible [7].

A similarity measure is a function of the variable values used for clustering possessed by two individuals,

$$s_{ij} = f(\mathbf{x}_i, \mathbf{x}_j) \quad \mathbf{x}_i = \{x_{i1}, ..., x_{im}\} \tag{43}$$

where x_{ik} is the value of variable k possessed by individual i. The measure is symmetric $s_{ij} = s_{ji}$, $s_{ik} = 1$ for all $i = k$ and usually bounded by zero and one. In contrast we could define the complement of similarity $d_{ij} = 1 - s_{ij}$, a dissimilarity measure. If several variables of different kinds or measured in different units are used in construction of the similarity measure we should account for the difference in variability that they might display, because high volatile variables will outweigh the less. If we are dealing with interval variables [4] suggests standardization with the standard deviation of all observations $z_{ik} = x_{ik}/s_k$ but notes that dilution of group differences might occur. We want to find clusters where both the distance between territorial risk as well as location is minimized. To accomplish this we use the Euclidean distance,

$$d_{ij} = \sqrt{\sum_{k=1}^{m}(x_{ik} - x_{jk})^2}. \tag{44}$$

However the geographical risk and location are of different types and we need to make an adjustment to the dissimilarity measure. Assuming that the variance of geographical risk estimates are the same in all geographical units σ_θ^2, the standardized squared Euclidean distance for credibility is,

$$f(\theta_i; \theta_j; \sigma_\theta) = (z_{i1} - z_{j1})^2 = \frac{1}{\sigma_\theta^2}(\theta_i - \theta_j)^2. \tag{45}$$

On the other hand, standardizing the x- and y-coordinates would not make any sense, as previously mentioned. It would seriously dilute the clustering, since the y-variable has a larger range. The Euclidean distance for location,

$$g((x_{i2}, y_{i3}); (x_{j2}, y_{i3})) = (x_{i2} - x_{j2})^2 + (y_{i3} - y_{j3})^2.$$

The author of [14] suggest that these should be combined by applying a preliminary weight w on f, which can be thought of as including the factor $1/\sigma^2$.

$$g(\cdot) + wf(\cdot) \quad \Rightarrow \quad d_{ij} = \sqrt{(x_{i2} - x_{j2})^2 + (y_{i3} - y_{j3})^2 + w(\theta_i - \theta_j)^2}. \quad (46)$$

We should note that such a variable weighting scheme has its flaws. The weight are chosen on pure judgment and might simply reflect the prior belief of how the territories should be constructed [4], as opposed to searching for previously unknown patterns in data.

There are numerous methods we can use for clustering. The hierarchical techniques are divided into two main classes, agglomerative and divisive. Agglomerative routines starts with N clusters and successively merge them into larger groups according to a optimization criterion. The sequence of events can thereafter be displayed graphically in a dendogram three. The divisive class of techniques runs in the opposite direction, starting with one single cluster and gradually divides it into smaller sub-clusters.

6.1 Agglomerative routines

The agglomerative clustering algorithm goes through the following general steps.

1. Define a measure of dissimilarity and calculate a triangular proximity matrix.
$$D_1 = \begin{bmatrix} d_{11} & & & \\ d_{21} & d_{22} & & \\ \vdots & \vdots & \ddots & \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}$$

2. Search the proximity matrix for the most similar pairs, $\min(D_k)$ and group them into one cluster.

3. Update the proximity matrix with the new clustering.

4. Repeat steps $2 - 3$ until all objects belong to the same cluster.

Single linkage or nearest neighbor clustering is the simplest in the class of agglomerative techniques. The distance between clusters is the Euclidean and proximity is defined as the smallest distance between the closest units in two clusters, i.e. the nearest neighbors. If we denote two separate clusters by C_A and C_B the single linkage distance between them is formally written as,

$$D_{AB} = \min_{\substack{i \in C_A \\ j \in C_B}} d_{ij}$$

where d_{ij} is defined in equation (46). The primary advantage of the method is that it does not put any restrictions on the shape or size of a cluster. Applying the routine onto our data we encounter a great drawback – chaining. Meaning that even though groups of observations clearly belong to separate clusters, intermediate observations in between the clusters tend to connect them together. The problem occurs primarily in the southern part of the country, as almost all municipalities are chained together with exception for some single municipalities who make up their own clusters. Since the municipality centers are closer in the southern part than up-country, we partition the country into north and south, and look for clusters in each partition. Unfortunately this does not resolve the issue of chaining.

A opposite approach to proximity is using complete-linkage. The distance between two clusters is measured by the distance between the units furthest away in two separate clusters, i.e. furthest neighbors. The routine create compact clusters of different shapes, but they tend build clusters with an equal diameter. Other similar methods are average-linkage, centroid- and median clustering, but these will not be considered in this paper.

6.2 Ward's minimum variance method

The algorithm goes through the same general steps as any other agglomerative clustering routine but differs in the way similarity between clusters is defined. This particular method employs an analysis of variance approach. If we denote observation j with attribute i and in cluster c by X_{cij}, the well known decomposition of the total sum of squares (TSS) into between group sum of squares (BSS) and error sum of squares (ESS) become,

$$\begin{aligned}\underbrace{\sum_{c}\sum_{k=1}^{m}\sum_{j=1}^{n_c}\left(X_{ckj}-\bar{X}_{\cdots}\right)^2}_{TSS} =& \underbrace{\sum_{c}\sum_{k=1}^{m}\sum_{j=1}^{n_c}\left(\bar{X}_{ck\cdot}-\bar{X}_{\cdots}\right)^2}_{BSS}\\ &+\underbrace{\sum_{c}\sum_{k=1}^{m}\sum_{j=1}^{n_c}\left(X_{ckj}-\bar{X}_{ck\cdot}\right)^2}_{ESS}.\end{aligned}$$

We focus our attention on the ESS, which is composed of the sum of each clusters ESS. $\text{ESS}_c = \sum_{k=1}^{m}\sum_{j=1}^{n_c}\left(X_{ckj}-\bar{X}_{ck\cdot}\right)^2$, i.e. sum of the squared differences within each cluster. Initially, we have n objects/clusters, for each recursion of the algorithm two clusters are merged into one and the ESS will therefore be reduced by one term. At recursion stage t the whole ESS is denoted by $\text{ESS}^t = \sum_{c=1}^{n-t}\text{ESS}_c$. The loss increase as a result of merging two clusters at recursion t is defined as,

$$z_t = \text{ESS}^t - \text{ESS}^{t-1}, \quad t = 1, ..., n.$$

The z_t is minimized at each iteration of the routine, which represent the minimum variance criterion. But minimizaton of z_t does not ensure that the solution is the overall minimum solution [9] i.e. for all possible permutations of the observations into $n - t$ clusters. Ward's method is particularly compelling since it attempts to maximize our homogeneity criterion. The variation within each cluster should be as small as possible (ESS) and the variation between clusters as large as possible.

Before applying the method onto our data we need to address the issue of determining the credibility weighting constant w in equation (46). A reasonable approach is for a given number of clusters, determine at which value of w the clusters start to break up in the (x,y)-plane. Equivalently at what level municipality credibility is as important as location. In order for the clusters to start breaking apart, a relatively large weighting constant is needed when specifying a small number of clusters. With coordinate data in terms 10km, we obtain the relation in Table 8. Secondly we want to

Table 8: Credibility weighting constant.

Number of clusters	$\sqrt{w}$
10	52
20	36
40	21
60	21

determine the optimal number of clusters. To assess homogeneity we use the $R^2 = \frac{BSS}{TSS}$ i.e. how much of the variation in the data that is explained by a given number of clusters, Figure 21. When the number of clusters are greater than seven, the marginal utility in terms of explained variation is leveling off and we can see that it is reasonable to have at least ten clusters. For higher values in the constant, the marginal gain for including an additional cluster is slightly greater than when the number of clusters are small. A method not mentioned in the literature is to refit and evaluate Akaike's information criterion (AIC) in our model from section 2.1 using the clusters as zone-variables. For a constant equal to 36 AIC reaches its minimum at about 70 clusters and 80 for a constant of 21. Such a large number of territories is a bit too much considering that we start off with 290 geographical units to begin with. Studying the cluster distribution in the $(x, \widehat{\Theta})$- and $(y, \widehat{\Theta})$-plane, they are well separated by location level (x or y), but can adopt a wide range of $w\widehat{\Theta}$ values. In order to separate the clusters into reasonably sized groups on the credibility axis, not create too many and not to deviate from the subjectively determined credibility weighting constant we need to have about 40 clusters. Applying $w = 21$ and specifying the number of clusters to 40, the resulting territorial division is depicted in

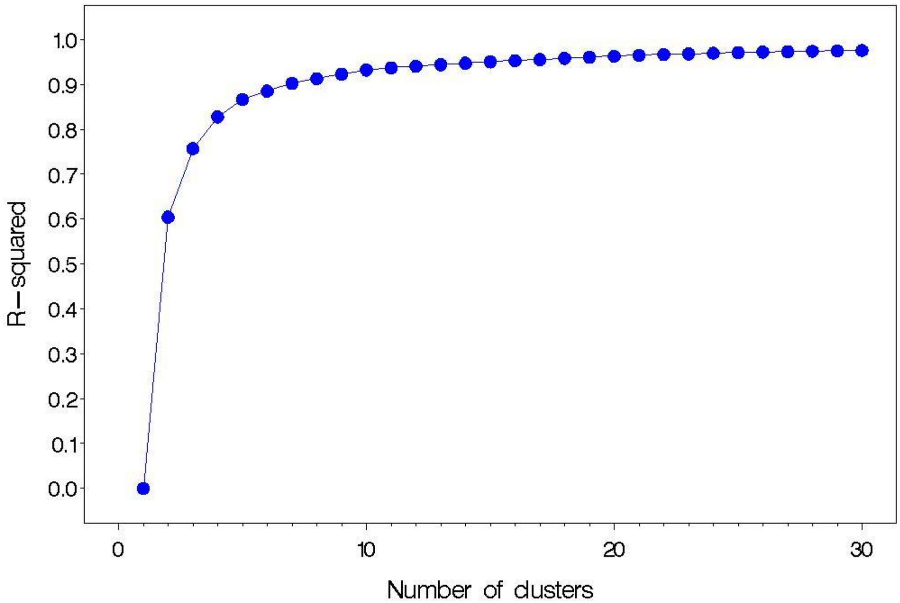

Figure 21: R-squared against the number of clusters for a credibility weighting constant equal to 26.

Figure 26.

6.3 Minimum variance adjacency clustering

There are several disadvantages inherent in Ward's method such as determination of weighting constant and finding irregular shaped territories. Because our prime aspiration is to construct territories by level of risk, we create a Ward type of clustering routine based on the level of risk and adjacency. The ESS is only based on municipality credibility,

$$\text{ESS} = \sum_{c} \sum_{j=1}^{n_c} \left(X_{ci} - \bar{X}_{c\cdot} \right)^2 \tag{47}$$

we minimize z_t at each iteration under the condition that the clusters are adjacent. Since we only use one clustering variable we do not need to standardize or using any weighting constant and the homogeneity criteria based on cluster risk is optimized (at each iteration). Plotting the proportion of variance within clusters against the number of clusters ESS/TSS, we are able to observe two level effects, Figure 22. The first obtains at 13 clusters and the second at 29, after which the additional reduction of variance within clusters for each additional cluster is obtained at a lower rate. These levels can be considered as optimal number of clusters. Whilst considering AIC, a minimum is achieved at 65 clusters.

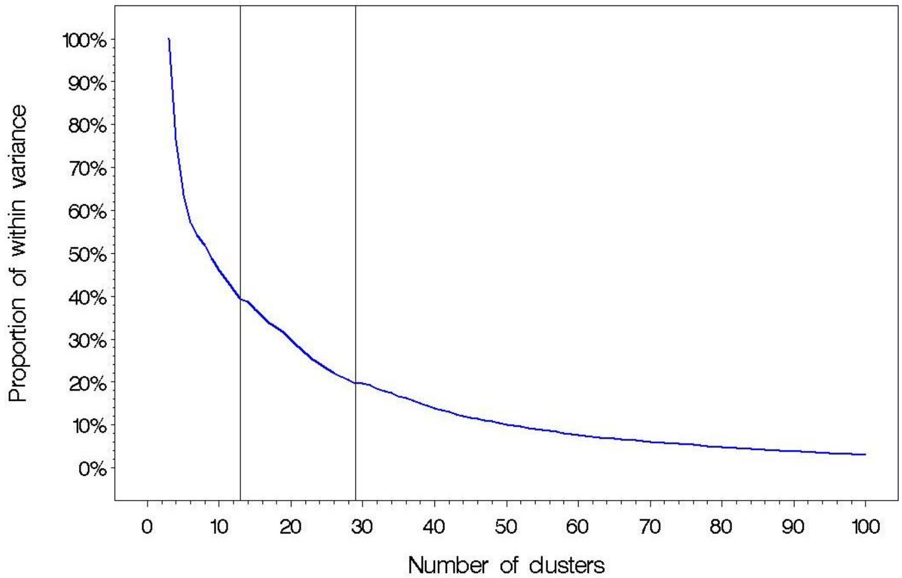

Figure 22: Proportion of variance within clusters to total variance against the number of clusters.

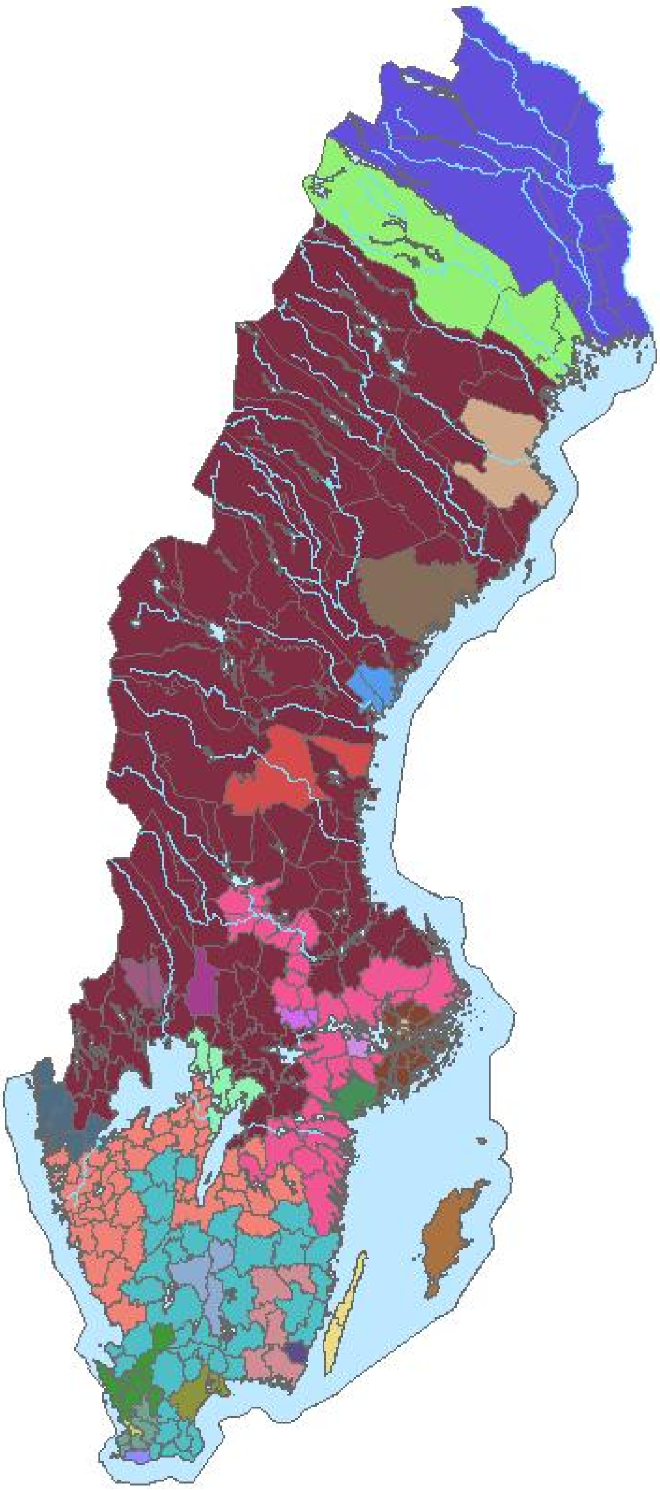

Figure 23: Minimum variance adjacency clustering territories.

7 Comparison and conclusions

In comparing the methods, we investigate the proportion of variance within clusters $1 - R^2$ and the fit of the model when including zone variable with Akaike's information criterion, Table 9. The likelihood ratio tests of type-III turns out highly significant for all territory divisions considered. We can conclude that the minimum variance adjacency clustering (MVAC) method performed best overall. It has by far the smallest variation within territories and AIC in the model sample. The fit is slightly worse than adjacency smoothed territories including auxiliary variable in the control sample. Considering that the MVAC-method obtains its minimum AIC at 65 clusters and has nine territories less, it is preferable. There are unfortunately too little data in the control sample at municipality level in order to obtain credible estimates of the municipality risk and hence the proportion of variance within territories by method. We try estimating the residual municipality

Table 9: Performance of territorial division

Model	Number of territories	Proportion of within variance	AIC model sample	AIC control sample
Without zones	0		133054	23669
Adjacency	40	43,0%	132689	23663
Adjacency with aux.	38	28,6%	132681	23637
Distance	48	40,5%	132690	23670
Distance with aux.	57	33,9%	132706	23642
MVAC	29	19,7%	132522	23640
Ward	40	39,0%	132675	23652

effect with credibility when MVAC zone variable is included, but to obtain an estimate of the variance between groups $\hat{\tau}^2 = -9.01 * 10^{-7} < 0$. This can happen if we cannot detect any differences between the groups [2]. Because the unbiased estimator of τ^2 in equation (30) contains a negative term that outweighs the between group sum of squares, we might consider selecting a strictly positive estimator. Choosing estimator of τ^2 based on $\sum_{i=1}^{I} \widetilde{w}_{i\cdot}(\overline{\overline{X}}_{i\cdot} - \overline{\overline{X}}_{\cdot\cdot})^2$ that is biased, we obtain credibility estimates highly concentrated around 1, further supporting the evidence of no group differences.

In order to answer the question whether the "non-random pattern" in model residuals is reduced when including zone variable we compare the observed and predicted claim frequencies at municipality level. When predicting a random variable In the model sample, we get a sum of squared error weighted by exposure of 8.752×10^6 without zone and 2.135×10^6

with MVAC zone variable. Corresponding for the control sample the sum of squared errors are 2.474×10^5 without and 2.173×10^5 with zone variable i.e. either the variance of the random variable is reduced, the bias ("non random pattern") or both. And since the zone variable is a highly significant we have evidence supporting that the systematic error of prediction is reduced.

An evaluation of the spatial distribution of claims is preferably conducted by plotting coordinates of each claim on a map and investigating the resulting patterns. This should also be complemented by claim frequencies in small geographical units. In our case, this is not easily done, because the current coordinate definitions are not complete and based on different projections of the earth surface. At municipality level we can conclude that there is a higher level of claims in high populated areas, especially around Stockholm and Malmö. But also south of lake Vättern and spreading to the coast, along Dalälven and by the west coast in municipalities such as Timrå, Härnösand, Skellefteå and Boden.

8 Discussion

There are three simple but important realizations one has to come to before embarking territorial construction and rating. First, we are looking for geographical risk terraces and the boundaries should be defined as the point where the level of risk drops. Second, the choice of method should be based on the feature we would like the territories to reflect. Third, there is no perfect method for obtaining territories; all have their own flaws and advantages. The methods are heuristic and will lead to a more or less good solution that should be complemented by manual revision to reduce the complexity, number of parameters and to obtain sought territory boundaries. The clustering routines excluding MVAC suffer from problems of determining the weighting constant, the number of clusters, creation of clusters of irregular shape and chaining. Although Ward's method is conditioned by such difficulties it has the superior strength of optimizing the homogeneity within clusters at each iteration. Ward's method tends also to fuse groups that have a small number of observations and create clusters of equal size with a spherical shape [9]. Whilst smoothing techniques are more intuitive and does not restrict the shape and size of the resulting territories, we need to define an appropriate smoothing constant or weighting function, which can be rather arbitrary. Both adjacency and distance based smoothing should perform equally well if the units were about the same size and shape, but as we have been considering municipalities, using a distance function is preferable. A good idea would be to evaluate resulting territories acquired from both smoothing and clustering as the methods can be seen as complementary. Boundaries based on smoothed estimates will reflect a general elevation in risk and a territory will largely be dependent on the levels of classification. Clustering routines group units by a criterion such as location and risk or continuity and risk. Smoothing is a good way to get an idea of risk terrace boundaries but MVAC is probably best for aggregation into territories since we are primarily interested in optimizing homogeneity of risks.

If accurate coordinate data is available a higher precision in determining the underlying territory boundaries is possible to achieve, given that we have enough exposure and claims. By partitioning the country is into a square grid we are allowing the actuary to define the size of each unit such that all requirements in section 1.4 are fulfilled and the actual territory boundaries more apparent. Even though the applied unit definition is not perfect, it is good enough for acquiring a satisfactory zone variable.

For the territories to be valid estimates of the geographical risk the underlying model have to be as good as possible. Since the credibility estimates are obtained under the hypothesis that all residual variation in terms of an additional multiplicative factor is a pure effect of territory. Inclusion of auxiliaries will reduce the residual variation and hence increase the predictive power of the model. The geographic risk in municipalities with a very small amount of exposure will be adjusted from an estimate close to the base level towards the auxiliary variable estimate, which should be closer to the truth. Combining credibility and auxiliaries, smoothing and classifying estimates in order to observe boundaries, then creating territories with MVAC and at last manually revise territories compared to smoothed value boundaries, is to the best of my knowledge to be considered as best practice.

By creating territories based on the belief that the risk in a group of units is based on location we assume that there is a hierarchy determining the risk of damage.

$$\text{Territory} \quad \Rightarrow \quad \text{Unit} \quad \Rightarrow \quad \text{Policy}$$

For us to obtain a measure of risk based on location- and unit experience, we could estimate the hierarchy by another,

$$\text{County} \quad \Rightarrow \quad \text{Municipality} \quad \Rightarrow \quad \text{Policy}.$$

This would lead us to consider hierarchical credibility models. Taking into account a set of ordinary rating factors γ_i we assume that the expected value of the key ratio is given by,

$$E\left[X_{ijkt} \mid \Theta_j, \Theta_{jk}\right] = \mu_0 \gamma_i \Theta_j \Theta_{jk}.$$

An evaluation of these models for territorial construction is left as work for the future.

References

[1] The Actuarial Profession. *Vehicle postcode zoning in personal lines rating*, 1999 General Insurance Convention, 1999. `http://www.actuaries.org.uk/sites/all/files/documents/pdf/0173-0205.pdf`.

[2] H. Bühlmann and A. Gisler. *A Course in Credibility Theory and its Applications*. Springer, 2005.

[3] S. Christopherson and D.L. Werland. Using a geographic information system to identify territory boundaries. Casualty actuarial society forum - arlington, Casualty Actuarial Society, 1996.

[4] B. S. Everitt. *Cluster Analysis*. Edward Arnold, 3 edition, 1993.

[5] S. Feldblum. Risk classification, pricing aspects. In J.L Teugels and B. Sundt, editors, *Encyclopedia of Actuarial Science*, volume 3, pages 1477–1483. Wiley, 2004.

[6] P.J. Jennings. Using cluster analysis to define geographical rating terretories. Discussion paper program, Casualty Actuarial Society, 2008.

[7] U. Jensen. Multivariate statistics. In J.L Teugels and B. Sundt, editors, *Encyclopedia of Actuarial Science*, volume 2, pages 1147–1152. Wiley, 2004.

[8] P. McCullagh and J.A. Nelder. *Generalized Linear Models*. Chapman & Hall, 1989.

[9] R. Mojena. Ward's clustering algorithm. In S. Kotz, C.B. Read, and D.L. Banks, editors, *Encyclopedia of statistical sciences*, volume 9, pages 529–532. Wiley, 1988.

[10] E. Ohlsson and B. Johansson. *Non-Life Insurance Pricing with Generalized Linear Models*. Springer, 2010.

[11] R. Sundberg. Statistical modelling by exponential families. Lecture notes, Department of Mathematics at Stockholm University, March 2010.

[12] R. Verrall. Claims reserving using credibility methods. In J.L Teugels and B. Sundt, editors, *Encyclopedia of Actuarial Science*, volume 1, pages 277–282. Wiley, 2004.

[13] G. Werner and C. Modlin. *Basic Ratemaking*. Casualty Actuarial Society, forth edition, 2010.

[14] J. Yao. Clustering in ratemaking: with application in territories clustering. Discussion paper program, Casualty Actuarial Society, 2008.

[15] A. C Yeo, K. A. Smith, R. J. Willis, and M. Brooks. Clustering technique for risk classificatino and prediction of claim costs in the automobile insurance industry. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 10(1):39–50, 2001.

A Figures

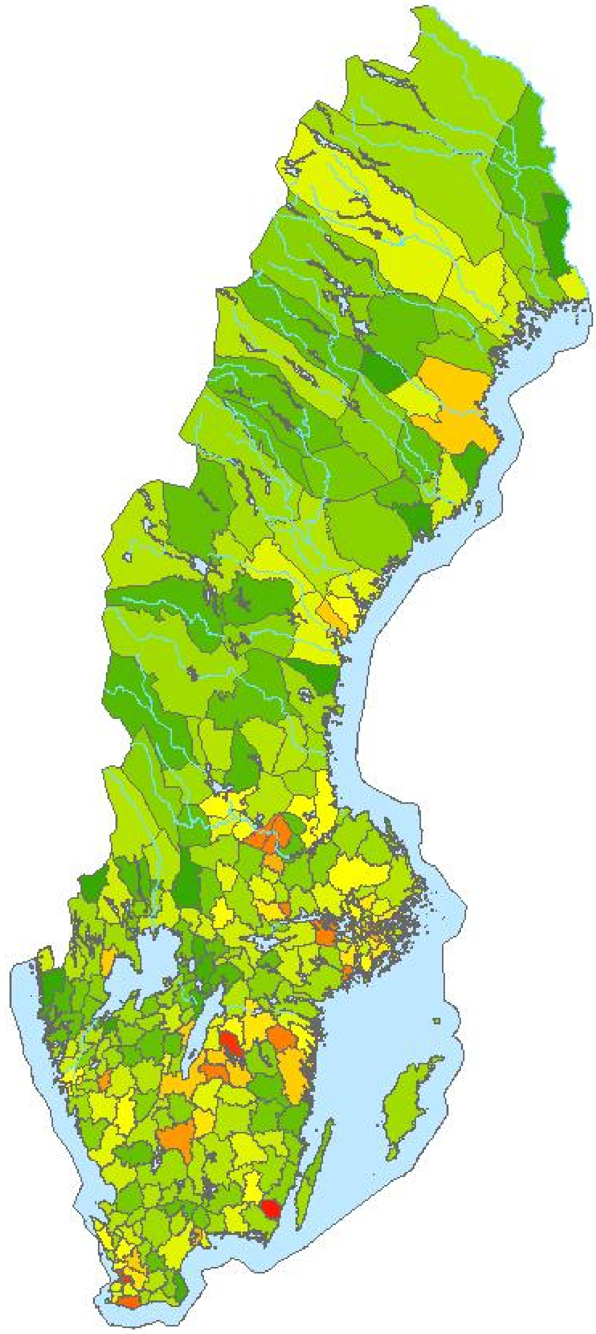

Figure 24: Empirical claim frequencies classified by the categories in Figure. 1. The figure is referenced on page 9.

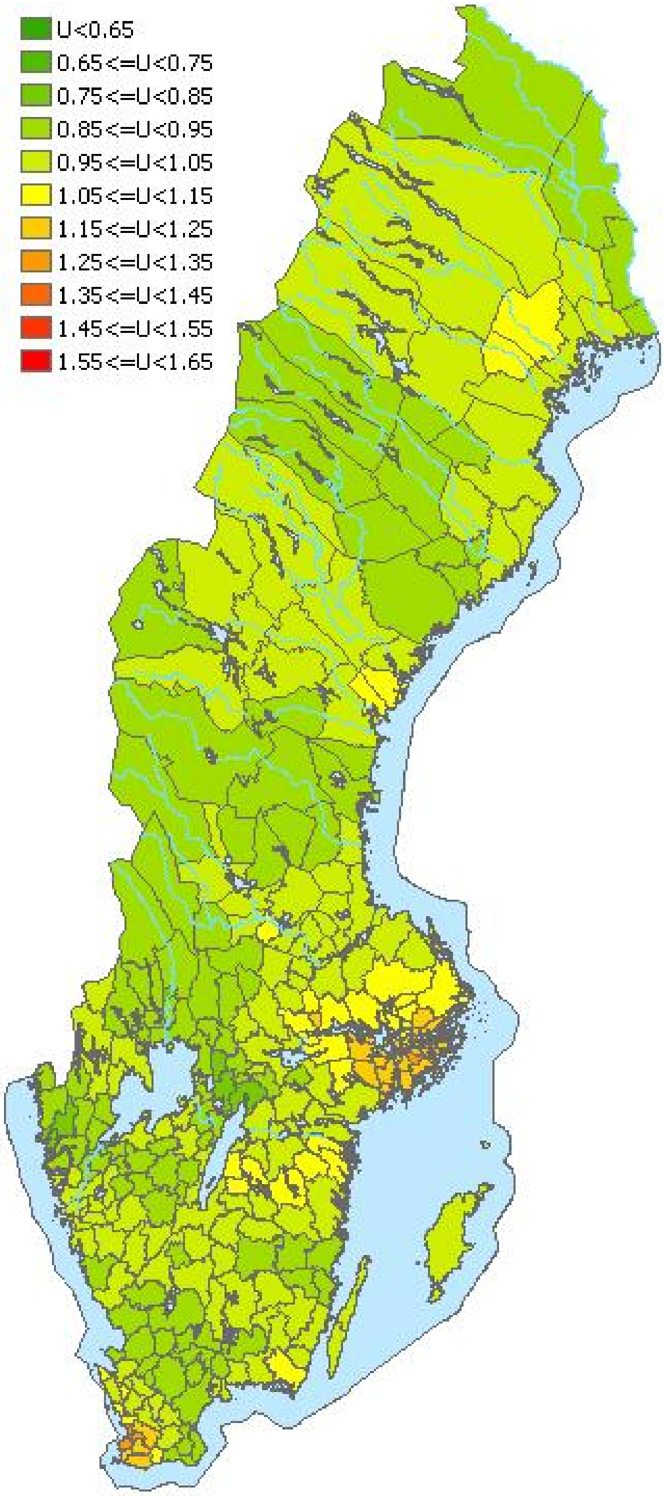

Figure 25: First order adjacency smoothed credibility with $\gamma^{(1)} = 3$. The figure is referenced on page 34.

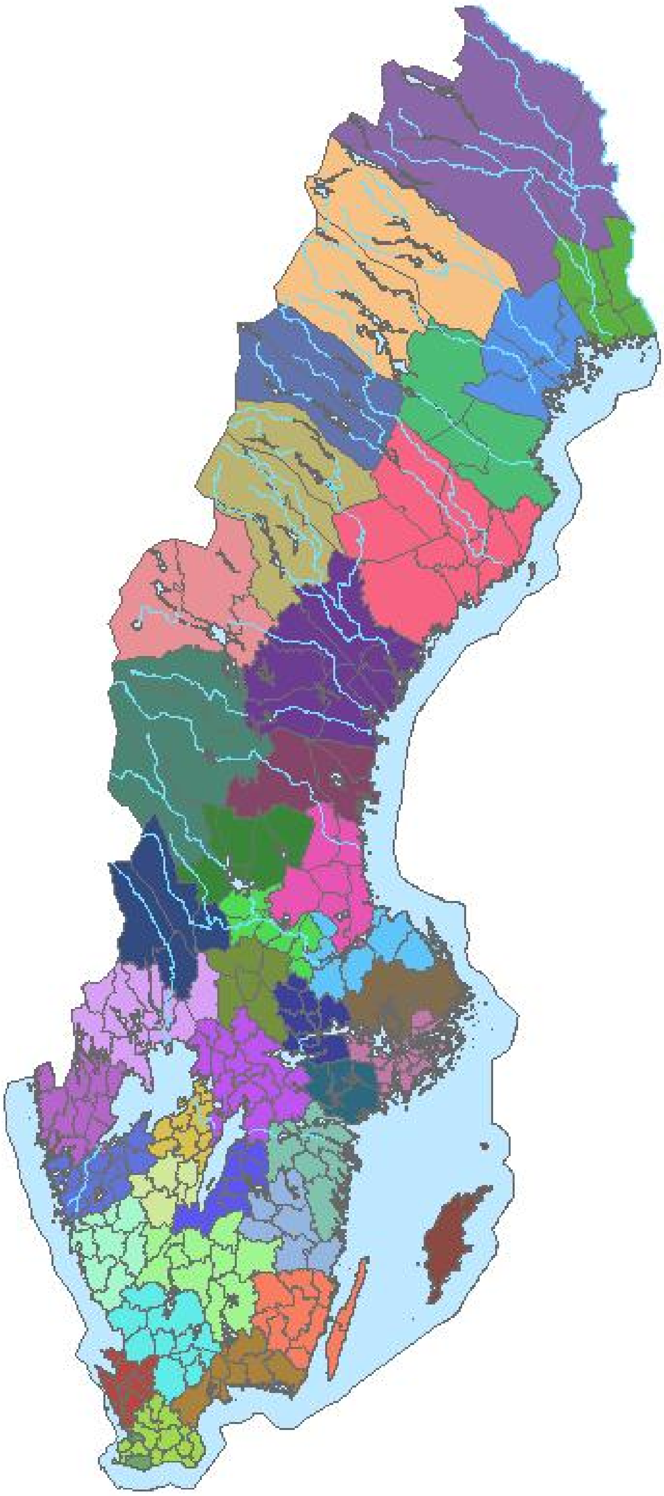

Figure 26: Ward territories.

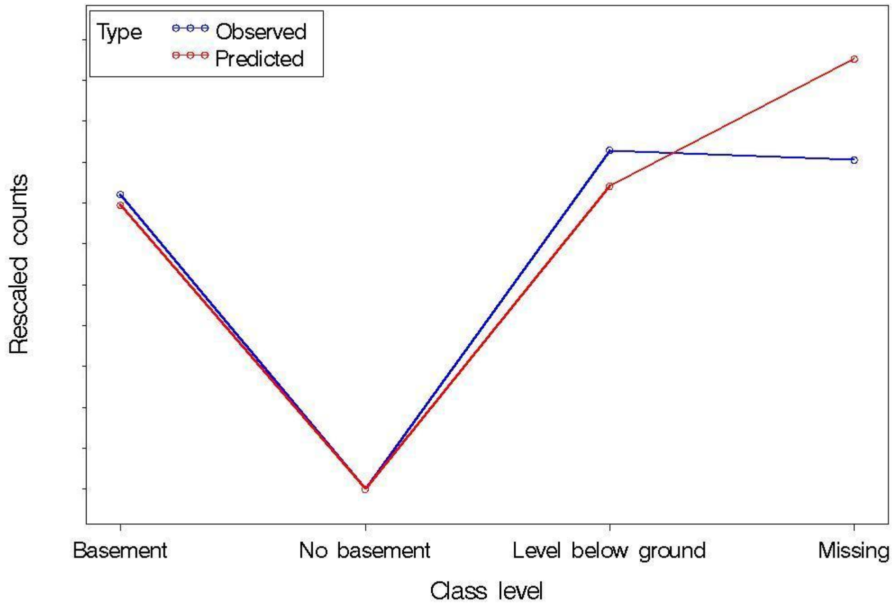

Figure 27: Basement variable parameter estimates compared with rescaled observed counts. The figure is referenced on page 19.

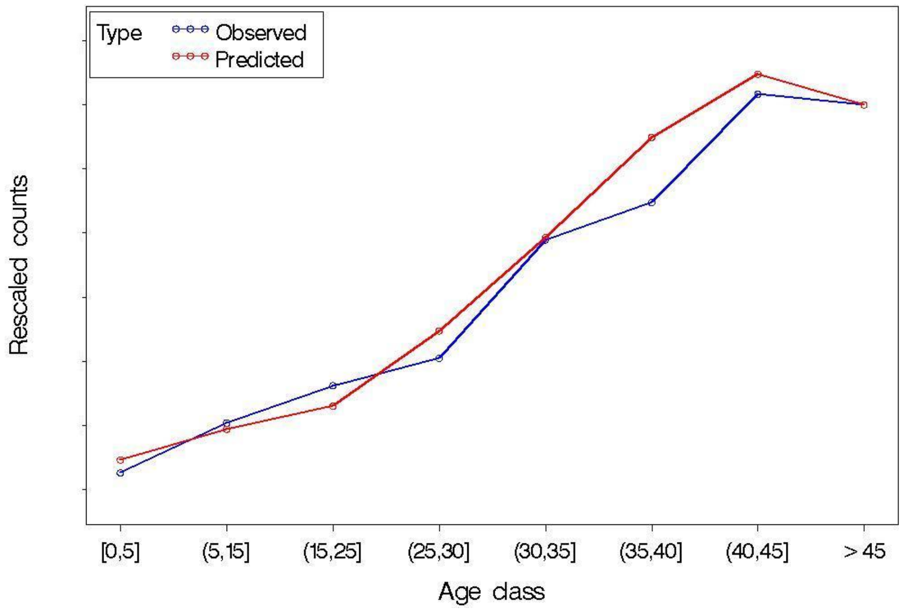

Figure 28: Building year parameter estimates compared with rescaled observed counts. The figure is referenced on page 19.

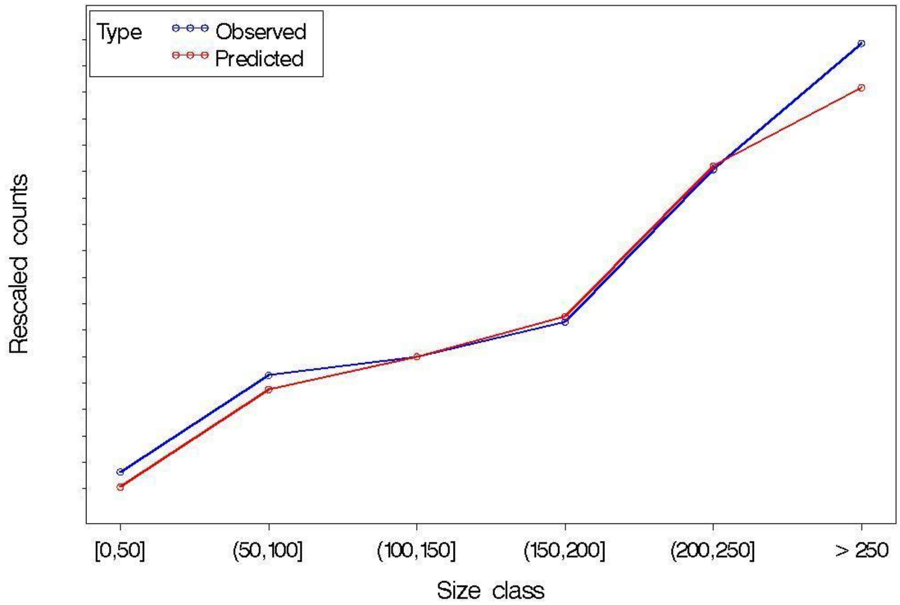

Figure 29: Living area size parameter estimates compared with rescaled observed counts. The figure is referenced on page 19.

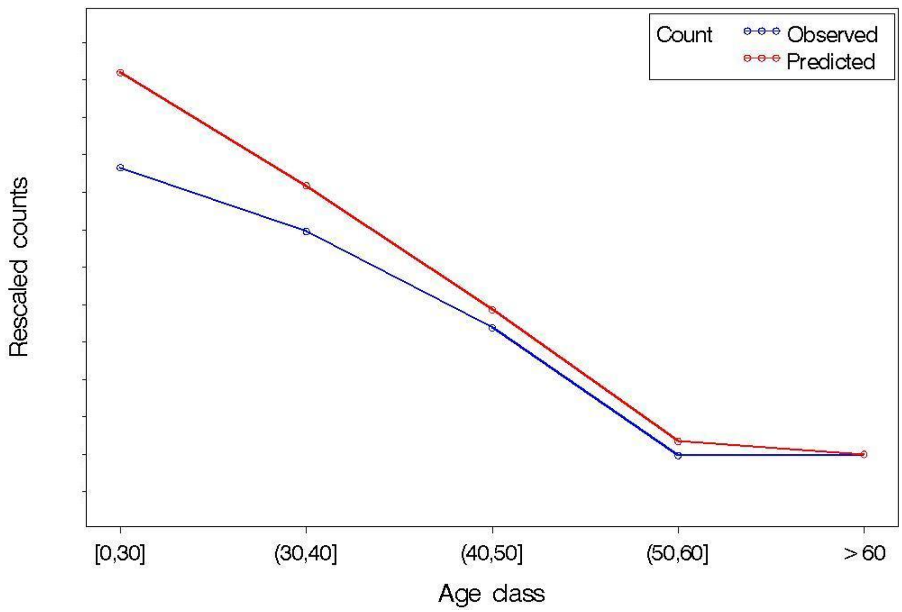

Figure 30: Policy holder age parameter estimates compared with rescaled observed counts. The figure is referenced on page 19.

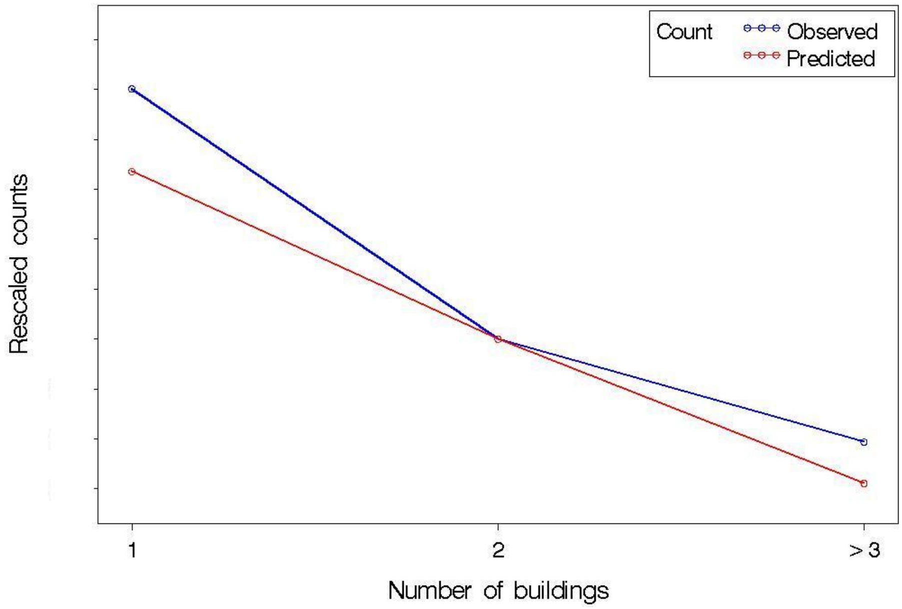

Figure 31: Number of buildings parameter estimates compared with rescaled observed counts. The figure is referenced on page 19.

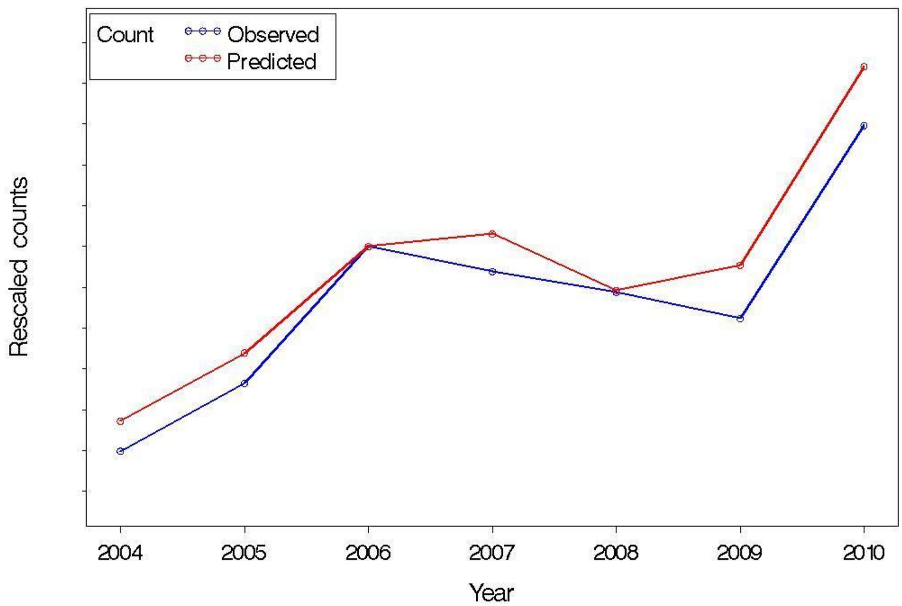

Figure 32: Calendar year parameter estimates compared with rescaled observed counts. The figure is referenced on page 19.

B Appendix

B.1 Backfitting algorithm

The algorithm for simultaneous estimation of rating and credibility factors in a GLM proposed by [10].

Step 0: Initially, let $\widehat{\Theta}_j = 1$ for all j.

Step 1: Estimate the parameters for the ordinary rating factors by a Tweedie GLM (typically Poisson or Gamma) with log-link, using $\log(\widehat{\Theta}_j)$ as an offset-variable. This yields $\hat{\mu}$ and $\hat{\gamma}_1^i, ..., \hat{\gamma}_R^i$.

Step 2: Compute $\hat{\sigma}^2$ and $\hat{\tau}^2$, using the outcome of Step 1.

Step 3: Use equation (31) to compute $\hat{\Theta}_j$, using the estimates from Step 1 and 2.

Step 4: Return to Step 1 with the new $\hat{\Theta}_j$ from Step 3.

Repeat Step 1-4 until convergence.