Pricing of General Insurance and the Impact of Asymmetric Information

Martin Englund

May 28, 2010
Preface

This thesis has been prepared in fulfillment of the requirements for both the Ph. D. degree in finance at the Finance Research Group, Department of Business Studies at Aarhus School of Business, Aarhus University, and the industrial Ph.D. degree under the Danish Ministry of Science, Technology and Innovation. The work has been carried out in the period from April 2007 to May 2010 under the supervision of Professor Carsten Tanggard and Professor Anders Grosen at Aarhus University and Professor Jens Perch Nielsen at Cass Business School, City University, London. The project has been funded by RSA Scandinavia (Codan Forsikring) and the Danish Agency for Science, Technology and Innovation under the Ministry of Science, Technology and Innovation.

In this thesis each chapter is written as an academic paper. The chapters are self-contained and can be read independently. This structure results in some notational discrepancies among the different chapters. There is also some overlap in the contents. The papers are changed in that the references have been merged and are placed at the end of the thesis. Hence, bibliographical repetitions are avoided. The papers are placed in logical order rather than in the order they were written or published.

As an industrial Ph. D. student I have experienced two worlds, both challenging but in very different ways. Doing research and development in the Danish business sector has been an excellent opportunity for me, but nevertheless it has been somewhat harder than I originally expected. Merging the academic and the business worlds has been one of the major challenges during my Ph. D. studies; to find out, focus on and balance between what can improve the business, be implemented due to systems and data available and yet is enough theoretically advanced and correct to contribute to the academic society. Another challenge for me has been communicating my findings at all levels, in these two worlds. These are skills I value highly and they will be most useful also in the future. A somewhat unforeseen challenge for me, was to bridge
the research areas of insurance mathematics and economics. Though, it led to many interesting discussions, which have given me a deeper knowledge within the two fields. Summarizing this Ph.D. programme I am also proud to say that besides the academic findings of the papers it also resulted in a new experience-rated pricing scheme of a personal line product in Sweden, and a new commercial line product with an eligible risk-based bonus in Denmark.

Acknowledgements

I take the opportunity to thank my supervisors Professor Carsten Tanggaard for tirelessly reading through my manuscripts, and for his important comments on and suggestions for this thesis, and Professor Anders Grosen for his support and encouragement throughout this Ph.D. programme. I would also like to thank Carsten and his wife for their generous hospitality when I was invited to their home. A special thank you goes to my industrial supervisor Professor Jens Perch Nielsen for encouragement, support and countless interesting discussions within the areas of both actuarial science, economics, innovation and management. I have learnt so much from him while writing my Master Thesis and throughout the last three years. I would also like to thank the co-authors of the papers included in this thesis.

On a personal level I would like to thank my (present and former) colleagues at Codan Forsikring; Jørgen, Robert, Jim, Fredrik, Tine, Kaspar and Anders for providing an enjoyable work atmosphere and for countless interesting discussions, both professional and personal, and for being a great inspiration to work with. Furthermore, I want to thank my family and friends for general support and encouragement through the past three years. Finally, but most of all, I would like to thank my wife Emma for her endless love, understanding and support.

Martin Englund
Löddeköpinge, May 2010
Summary

To lose one parent, Mr Worthing, may be regarded as a misfortune; to lose both looks like carelessness.

Oscar Wilde, The Importance of Being Earnest, 1895, Act I

To set the insurance premium correctly is of utmost importance on a competitive insurance market. Hence the overall objective of this thesis is to improve the pricing, first by using individual claims information, and second by using information about the individuals’ choice of coverage. Regarding how to use claims information, the quotation above raises a lot of questions: What information lies in an occurred event? When is an accident to be regarded as misfortune and when is it to be regarded as due to carelessness? How much can we rely on information given?

In the first two papers we present models for multiline experience rating of claim frequency, with prior known differences of the individuals. Here multiline refers to using claims information in additional lines of business, which means that to price one insurance coverage we also consider claim information in other correlated coverages or products. In the first paper of the thesis, Englund, Guillén, Gustafsson, Nielsen, and Nielsen (2008), the model is applied to two lines of business, but can easily be extended to arbitrary numbers of dimensions. The paper focus on describing the new models and investigating how these can affect the pricing of individual policies. In the second paper, Englund, Gustafsson, Nielsen, and Thuring (2009), we introduce a multidimensional credibility model, similar to the one in the first paper, but with closed-form expressions of the estimators, and we investigate another time dependent model. These models are investigated on a portfolio level, to see if they actually improve the pricing. The conclusion of our empirical study is that experience rating is extremely useful for pricing. While this hardly is a surprising conclusion, it might be surprising that we are able to present a situation in which the inclusion of experience rating gives an extra
improvement of the same order of magnitude as the improvement obtained from leaving the trivial flat rate and entering sophisticated covariate-based rating principles without experience rating. We also conclude that our multivariate credibility approach indeed is capable of improving the quality of estimation compared to classical one-dimensional credibility theory. However, the increased pricing precision of using claims informations from additional insurance products is not that large compared to the level of sophistication of the method and thereby also the costs for implementing it. We also introduced a time-dependent model. Adding the time effect does not generally improve prediction in our case. The reason seems to be that we have too few years of observations available in our data set. The introduced multidimensional models can also be used in cross-selling or profitability models, but unfortunately we still can not improve the pricing for entirely new customers. On the other hand, so can no experience rating scheme, due to the lack of claim information. Thus, in the second part of the thesis we seek an improvement of the pricing, by utilizing asymmetric information. This is the subject of the remaining two papers. The third paper, Englund (2010), investigates if asymmetric information is present on the Danish automobile insurance market (the terms asymmetric information, adverse selection and moral hazard are explained in the introduction, chapter 1). Evidence of asymmetric information is found for policyholders with a fixed price (not experience-rated) product. The findings are coherent with adverse selection models. We also find a positive claim occurrence dependency, which imply the correctness in using experience-rating as in the first two papers.

Traditionally asymmetric information is considered a problem since only policyholders with a higher risk than expected can benefit from their information; in order to purchase an insurance they do not have to reveal themselves as being higher risks than expected and can buy their insurance at a too low price. No one will believe the policyholders with a lower than expected risk when they say that they actually should have a cheaper insurance policy, since anyone can say that. Instead of considering asymmetric information as a problem, we are now interested in how to use it. In the fourth paper, Englund, Nielsen, and Tanggaard (2009), we assume that the policyholder knows something, about herself, which cannot be captured in the statistical pricing method. We present a risk-based bonus system to deal with the effects of asymmetric information, either as a price differentiation or a self-selection mechanism, or both. We find that the insurer easily can improve the pricing by implementing a bonus system where every policyholder has an opportunity of a bonus with an individual bonus rate, in contrast to the present fixed bonus rates for a limited subset of policyholders. An
even larger improvement in pricing performance is obtained by letting the policyholders choose, to buy the bonus product or not, by themselves. A third and last improvement is obtained by letting the policyholders choose the size of entrance fee, based on their private information on risk, preferences and wealth.

To summarize, experience rating mitigates the effects of asymmetric information. Experience rating can be introduced either as a pricing scheme or as an additional bonus product.

Note that insurance is to be understood as only general insurance and not life insurance. Even though they are quite similar in many aspects the life insurance market is not considered in this thesis.
At miste en forælder, hr. Worthing, kan betragtes som en ulykke, at miste både ligner skødesløshed.

Oscar Wilde, The Importance of Being Earnest, 1895, 1. Akt

At prisfastsætte forsikringspræmien korrekt er af yderste vigtighed på et konkurrencepræget forsikringsmarked. Derfor er det overordnede formål for denne afhandling at forbedre prisfastsættelsesmetoden, ved at først anvende individuel skadesinformation, og derefter ved at bruge information vedrørende individernes valg af dækning. Med hensyn til hvordan man bruger skadesinformation, rejser citatet ovenfor en række spørgsmål: Hvilken information ligger der i et indtruffet begivenhed? Hvornår er en ulykke at betragte som uheld og hvornår skal den betragtes som følge af skødesløshed? Hvor meget kan vi stole på oplysningerne?

I de to første artikler præsenterer vi modeller for multidimensionel erfaringsstariffering for skadesfrekvens med individuel information. Her henviser multidimensionel til brugen af oplysninger om yderligere dækninger, hvilket betyder, at for at prisfastsætte en forsikringsdækning bruger vi også skadesinformation fra andre korrelerede dækninger eller produkter. I det første papir i afhandlingen, Englund, Guillén, Gustafsson, Nielsen, and Nielsen (2008), anvendes modellen på to forretningsområder, men kan let udvides til vilkårlige numre af forretningsområder. Papiret fokuserer på at beskrive de nye modeller og undersøge, hvordan disse kan påvirke prisfastsættelsen af de enkelte policer. I det næste papir, Englund, Gustafsson, Nielsen, and Thuring (2009), introduceres en multidimensionel kredibilitetsmodel, svarende til den i det første papir, men med estimatorerne bestemt på lukket form, og vi undersøger en anden tidsafhængig model. Disse modeller er valideret på porteføljeniveau, for at se om de rent faktisk forbedrer prissættningen. Konklusionen af vores empiriske undersøgelse er, at erfaringsstariffering er meget nyttigt til prisfastsættelse. Selv om dette næppe er en overraskende konklusion, kan

I den anden del af afhandlingen forsøger vi at opnå vi en forbedring af prisfastsættelsen, ved at bruge asymmetricisk information. Dette er emnet for de resterende to papirer. Det tredje papir, Englund (2010), undersøger om asymmetricisk information er til stede på det danske bilforsikringsmarked (asymmetric information, adverse selection og moral hazard er erklæret i indledningen, kapitel 1). Beviser for asymmetricisk information er fundet for forsikringstagere med en fast pris. Resultaterne er i overensstemmelse med adverse selection modeller. Vi finder også en positiv skadesafhængighed over tiden, som styrker korrekheden i at bruge erfaringstariffering som i de første to artikler. Traditionelt er asymmetricisk information betragtet som et problem, da kun forsikringstagere med en højere risiko end forventet kan drage fordel af deres informationer; de behøver ikke at afsløre sig selv og kan købe deres forsikring til en for lavt pris. Sagen er den, at ingen vil tro forsikringstagere med en lavere end forventet risiko, når de siger, at de rent faktisk burde have en billigere forsikring, da enhver kan sige det. I stedet for at betragte asymmetricisk information som et problem, er vi nu interesserede i, hvordan man konstruktivt kan bruge det.

I det fjerde papir, Englund, Nielsen, and Tanggaard (2009), antager vi, at forsikringstagere ved noget, om sig selv, som ikke kan fanges i de statistiske prisfastsættelsesmetoder. Vi introducerer et risikobaseret bonussystem til at håndtere følgerne af asymmetricisk information, enten som en prisdifferentiering eller en selvestændig ud-
vælgelse mekanisme, eller begge dele.
Vi konkluderer, at et forsikringsselskab kan forbedre prissætningen ved at indføre et bonussystem, hvor alle forsikringstagere får bonus baseret på en individuel bonussats, i modsætning til det nuværende system hvor kun en begrænset andel af forsikringstagerne får bonus, med faste bonussatser. En endnu større forbedring af prissætningen opnåes ved at lade forsikringstagerne frivilligt kunne vælge dette bonusprodukt. En tredje og sidste forbedring er opnået ved at lade forsikringstagerne vælge hvormeget de vil betale for denne bonusordning (hvilket påvirker størrelsen på bonusen) baseret på deres private oplysninger om risiko, præferencer og formue. For at opsommere, så reducerer erfaringstariffering virkningerne af asymmetrisk information. Erfaringstariffering kan indføres som enten en prissætningsordning eller som et ekstra bonusprodukt.

Det skal bemærkes, at når denne afhandling omtaler forsikring omfatter dette kun skadesforsikring. Selvom der formentlig er mange lighedspunkter, vil denne afhandling ikke beskæftige sig med livsforsikring.
Resumé
List of papers

This thesis is based on the following four papers:


## Contents

Preface i  
Summary iii  
Resumé vii  
List of papers xi  

### 1 Introduction 1

1.1 Insurance Pricing 1  
1.2 Experience rating and Credibility theory 3  
1.3 Asymmetric information, Adverse selection and Moral hazard 4  

### 2 Paper A 5

2.1 Introduction 6  
2.2 Multidimensional Bühlmann-Straub credibility model 8  
2.3 Multidimensional Bühlmann-Straub credibility model with time dependence 10  
2.4 Data study 11  
2.5 Conclusions 15
3 Paper B

4 Paper C
  4.1 Introduction ......................................................... 32
  4.2 Economical Model ................................................... 35
  4.3 Empirical issues ..................................................... 40
  4.4 Results ............................................................... 45
  4.5 Discussion .......................................................... 50
  4.6 Conclusions ......................................................... 53

5 Paper D
  5.1 Introduction ........................................................ 56
  5.2 Motivation ........................................................... 58
  5.3 The Model ........................................................... 61
  5.4 Pricing contracts: simulated evidence ............................. 64
  5.5 Individual Choice and Self-selection .............................. 69
  5.6 Discussion .......................................................... 76
  5.7 Conclusions ........................................................ 79

A Appendix
  A.1 Data Process ....................................................... 81
  A.2 Credibility Model and Estimators ................................. 82

Bibliography 85
Chapter 1

Introduction

Since this thesis should be able to be read not only by actuaries but also by people whom may not necessarily have a more detailed knowledge about the insurance market, apart from the experience they have made through their own personal insurances, we start with a short survey of the most fundamental terms used in this thesis.

1.1 Insurance Pricing

*Predictions are hard, especially about the future.*

Lawrence “Yogi” Berra

The basic idea of insurance is that individuals transfer their risk to an insurance company, and for this they pay premiums. It would be ideal if the company could offer an individual premium, to each and every customer, which precisely reflects the risk of this customer, called the (actuarially) fair premium. In the applied world the complete insurance premium consists of different parts such as the actuarially fair premium, capital reserve, profit etc, but in this thesis we focus on the fair premium. In the actuarial literature the fair premium is sometimes called the (pure) risk premium. This might be a bit confusing for economists, since the risk premium, in the research area of economy, is the minimum additional return, over the risk-free rate, a person requires to be willing to take an uncertain bet. The confusion over what is meant by the risk premium is mutual, so we will only use the terms (actuarially) fair premium or actuarial
premium in the following.
By paying the individually fair premium each customer would level out her risk over the time, meaning that each customer would pay for her expected claim costs, while the company covers her experienced claim costs, which is the fundamental idea of insurance, but without the customers paying for each other’s risks. The insurer tries to estimate the fair premium as good as possible, usually by using statistical models and methods. The fair premium is most often estimated on the basis of a number of covariates, which are characteristics of the individual or the object to be insured. Covariates have to fulfil a number of criteria such as being objectively quantifiable, measurable and legal. For instance the gender of an individual can be a significant covariate but cannot be used in some countries, e.g. Sweden, since it is considered to be discrimination and is therefore prohibited by law. Covariates that can be taken into consideration for the pricing of an automobile insurance are e.g. the power of the engine, the age of the car, the brand and the geographic area where it is driven. In the reality we also have to consider the costs compared to the increased precision of our pricing, when including more covariates. When covariates are used to set premiums one assume that each individual with the same characteristics have the same risk. Even though the fair premium is not known by the company or the customer, the customers are basically indifferent with his or her fair premium; the customers only want the best possible coverage to the lowest possible price. Therefore it is of decisive importance for the company how good the statistical models are, since they affect the pricing, which is illustrated in the following example: Assume that company A has a simple rating scheme for car insurance, where all customer pay 1000 DKK in yearly premium. A new company B enters the Danish market and according to their rating women should pay 900 DKK and men should pay 1100 DKK. Assume that the fair premium actually is 900 DKK and 1100 DKK for women and men respectively. As long as company A has as many women as men in their portfolio they will not make any loss (at least over time). But as soon as the women in company A finds out that they can get the same insurance coverage for a smaller amount of money, they can save 100 DKK, they will change to company A as soon as possible, and on the contrary for the men in company B. They will change to company A to save money. Soon this transition results in that company A only has male customers and company B only has female customers. Having set the correct fair premium company B is indifferent to men and women in their portfolio. Company A on the other hand loses 100 DKK on every male customer, since their tariff (pricing scheme) was too blunt, and not detailed enough. This example is of course a very
simplified version of the real world, but it hopefully illustrates why it is important to understand the risks and to price as good as possible.

In the simple example above of how insurance is priced and how the market works, all customers are assumed to have the same risk, based on the covariates used. The problem in the real world is that two people with the same measurable properties most likely will behave differently, or have different abilities. We interpret this as there is some kind of latent individual risk profile, which can be used in experience rating and estimated with e.g. credibility theory.

1.2 Experience rating and Credibility theory

By experience rating means that the premium is adjusted based on what happened in the previous insurance period. If a claim is reported then the premium for next insurance period increases, and if no claims are reported it usually decreases, ceteris paribus. For most rating schemes there are floors as well as ceilings for how low and how high the premium may become. Experience rating is applied under the assumption that there might be something overlooked by the pricing scheme, or that a future reward will affect the customers behaviour to lower the propensity of claims. Credibility theory is a way of weighting individual and collective information, in this case claims information, based on how credible the information is. If we create a weighted mean the weight will be affected by basically three parameters. The individual variance, the collective variance and the individual duration. This means that if a person’s claims pattern is similar each year, say one claim each year, then we will use more of this individual claims information than if the person had ten claims in one year but none in the following. If the claim pattern for the group of policyholders, the collective, is similar, we will use more collective information, than if the group is heterogeneous. Finally we will find the individual claims information more credible if we have a lot of individual information, a long duration. When we introduce time dependence, as in the first two papers of this thesis, this means that we allow the individuals to change over time, by relying more on recent data then old one.

In the first two papers of this thesis we try to find ways of pricing insurance more correct based on firstly individual claim information and secondly individual choice of coverage.
1.3 Asymmetric information, Adverse selection and Moral hazard

When one party has access to information that is denied to the other, there is presence of asymmetric information. There are basically two special cases of asymmetric information: adverse selection and moral hazard. Adverse selection refers to a market process in which the "bad" products or customers are more likely to be selected, due to asymmetric information. The asymmetry can be caused by private information of one of the parties or by regulations or social norms which prevent e.g. the insurer from using certain characteristic to set prices. Two ways to model adverse selection are with signaling games and screening games. In insurance moral hazard means that the insured may have some degree of control of the probability of the occurrence, or size, of an insured event, due to care taken. Care can be interpreted as money; but also diligence, mental concentration, or intensity of effort. We can distinguish between expenditure undertaken to reduce the probability of an accident (self-protection) and expenditure to reduce the size of the contingent loss (self-insurance or loss reduction). Expenditures on fire sprinklers reduce the size of a loss, but not the probability of an accident. Expenditures on a burglar alarm reduce the probability of a theft whereas the decision not to leave the silverware in an unlocked container reduces the loss if there is a household theft. Driving an automobile more slowly and carefully reduces both the probability of an accident and the likely costs of an accident if it should occur.
Chapter 2

Paper A

**Multivariate Latent Risk:**
A Credibility Approach

**Abstract:** We investigate a concept of multivariate pricing, which includes claim history for more than one line of business and is a generalization of the Bühlmann-Straub model. The multivariate credibility model is extended to allow for the age of claims to influence the estimation of future claims. The model is applied to data from a portfolio of commercial lines of business.
2.1 Introduction

We start by considering the simplest possible sampling scheme of the multivariate credibility approach of Bühlmann and Gisler (2005): Let $X_1, X_2$ be stochastic variables that are conditionally independent given an one-dimensional unobservable variable $\Theta$. This sampling scheme is excellent for catching heterogeneity properties in correlated variables with the same underlying risk such as small claims versus big claims or claim frequencies versus claim severities. In this paper we also consider a multivariate credibility approach. We are interested in the situation where the two considered stochastic variables $X_1, X_2$ are correlated, not through the same unobservable underlying risk, but with an underlying unobservable risk parameter tied to each of the stochastic variables.

Let $X_1, X_2$ be conditionally independent stochastic variables given $\Theta = (\Theta_1, \Theta_2)$, where $\Theta_1$ is the latent risk variable describing the severity of $X_1$ and $\Theta_2$ is the latent risk variable describing the severity of $X_2$. This sampling scheme is useful when analyzing more than one line of business per customer. If $\Theta_1$ and $\Theta_2$ are correlated, information from one line of business will impact on the knowledge of the other line of business. The sampling scheme of Bühlmann and Gisler (2005), used in this particular context, would correspond to an absolute correlation of 1 between the two latent variables $(\Theta_1, \Theta_2)$. In empirical studies we see the whole spectrum of positive correlations ranging between zero and one. In this paper, we present in an empirical illustration study with correlation 0.7.

There is not a large literature on methods incorporating data from different business lines on the same policyholder. Exceptions includes the studies by Desjardins et al. (2001), in which Bonus-Malus systems for fleets of vehicles are derived from the claims or safety offences history, and Pinquet (1997, 1998), in which examples are given of experience rating that incorporates additional claim information, from the number and the cost of claims, and claims at fault and not at fault within the same line of business. Instead of following the expected value principle, as in the latter papers, we follow the tradition of linear credibility.

Traditional approaches of credibility theory, based on the ideas of Lundberg (1966), Bichsel (1967), Bühlmann (1967) and more recently Lemaire (1995), consider one unobserved risk parameter, for each customer and treat one policy or coverage, which we will call a line of business, at a time in order to calculate a premium based on both
collective and individual information. In this paper we consider two recent extensions
to the traditional approach; firstly the multidimensional generalization of the credibility
approach to more than one line of business, which was first introduced by Jewell (1973)
and Jewell (1974), and secondly to allow the multidimensional individual latent risk parameter to change over time.

Pioneering papers on one dimensional evolutionary credibility models are Gerber and
Jones (1975b), Sundt (1981) and Pinquet et al. (2001). The latter shows that the date
of the claim does matter because the effects of a claim on the risk evaluation diminish
over time. A concise review of credibility theory can be found in Norberg (2004), and
as a textbook on multiple time series analysis we recommend Lütkepohl (2005).

While our approach has an obvious potential for improved pricing, it is also promis-
ing for other purposes where a joint understanding of more than one line of business at
a time is important. One such example is cross-selling where policyholders may have
policies underwritten in several lines within the same company. Here the insurance
company would like to use the information of a policyholders latent risk of one line of
business to be able to cross-sell him a policy of another line of business at a competitive
price. This requires a multidimensional model where the correlation between the latent
risk parameters of the two lines of business are understood. So, for a customer with
one policy only, the experience rating mechanism should allow to incorporate the col-
lected historical information when pricing some other policy of interest of the customer.
Therefore, the insurance company can turn its data base into a competitive advantage
even in situations where they do not have historical information of the customer in the
line of business which they wish to cross-sell. This, however, needs a new derivation
of a modified Bühlmann-Straub model, since the ordinary multidimensional model is
reduces to model of lesser order when the customer lack one of the investigated lines of
business.

The Bühlmann-Straub credibility model assumes all customers to be a priori equal,
which is not the case for most insurance companies. So, to obtain some kind of contin-
uity between the present actuarial practice and our approach, we aim to develop our
models with prior known differences.

We will use a real data set from commercial insurance in Denmark to see the differ-
ences between the one- and two-dimensional history rating schemes.
We show through our practical example that modern extensions of credibility theory, incorporating multidimensional credibility and time effects, can lead to improved pricing of the customer risk after some knowledge of the claims history has been acquired.

This paper is organized as follows, section 2.2 begins with a review of the classical Bühlmann-Straub credibility model and the extension to the multi-dimensional case. Section 2.3 introduces time-dependent random effects into the model. An application to the data is presented in section 2.4. Conclusions are given in the final section.

2.2 Multidimensional Bühlmann-Straub credibility model

In this section, we provide a method to obtain a multidimensional credibility estimator for claim frequencies, with prior known differences of the individuals. An excellent and comprehensive treatment of credibility theory is found in Bühlmann and Gisler (2005). We will use similar notation and refer to the results therein.

Let us in the univariate case consider the number of claims $N_{ij}$ of customer $i$ in insurance period $j$, usually years, on one particular coverage, with $N_i = (N_{ij})_{j=1,...,J}$. We assume that every customer $i$ has its individual risk profile $\theta_i$, a realization of the random variable $\Theta_i$. We also assume that we have a priori knowledge on the expected number of claims from customer $i$ in period $j$, $\lambda_{ij}$, which depends on regression components represented by the vector $x_{ij}$ and the duration $\omega_{ij}$: $\lambda_{ij} = \omega_{ij} \exp(x_{ij}a)$; $a \in \mathbb{R}^k$, where $a$ is a vector of parameters and where $k$ is the number of regression components. This is particularly useful in practice because Bonus-Malus schemes are frequently super-imposed on tariffs given by regression components, or rating factors, that may change over time, e.g. age.

Formally, we assume that $N_{ij} | \Theta_i = \theta_i$ is Poisson distributed with mean $\theta_i \lambda_{ij}$, and moreover we assume that $(\Theta_1, N_1), (\Theta_2, N_2), ...$ are independent random vectors, where $\Theta_1, \Theta_2, ...$ are iid with $E(\Theta_i) = 1$ and $V(\Theta_i) = \tau^2$. This model corresponds to the case of known a priori differences, which we allow to change over time, studied in Bühlmann and Gisler (2005) where $Y_{ij} = N_{ij} / \lambda_{ij}$. Since $Y_{ij}$ satisfy the conditions of the Bühlmann-Straub model, and due to the linearity property of the projections, the
best linear predictor of $N_{i,J+1}$ given $N_{i,1}, ..., N_{i,J}$ and $\lambda_{i,J+1}$ is $\hat{\theta}_i \lambda_{i,J+1}$, where

$$
\hat{\theta}_i = \mathbb{E}(\Theta_i) + \text{COV}(\Theta_i, \bar{Y}_i) \mathbb{V}(\bar{Y}_i)^{-1}(\bar{Y}_i - \mathbb{E}(\bar{Y}_i)) = 1 + \alpha_i (\bar{Y}_i - 1) = 1 + \alpha_i \left( \frac{N_i}{\lambda_i} - 1 \right), \tag{2.1}
$$

and $N_i = \sum_j N_{ij}$, $\lambda_i = \sum_j \lambda_{ij}$, $\bar{Y}_i = \frac{1}{\lambda_i} \sum_j \lambda_{ij} Y_{ij}$ and $\alpha_i = \lambda_i / (\lambda_i + \tau^{-2})$.

The first expression follows easily from standard linear regression theory and the equivalence to orthogonal projection to minimizing the sum of squares, see Norberg (2004).

Since we assume a conditional Poisson distribution of the individual claims $N_{ij}$, given that the a priori differences are known and given the individual risk profile, $\mathbb{V}(N_{ij} | \theta_i) = \mathbb{E}(N_{ij} | \theta_i) = \lambda_{ij} \theta_i$. This implies that we assume equidispersion at the individual level, but it does not imply that there is a restriction on the dispersion in the portfolio. Moreover, we believe that Poisson is a model framework that is easy to apply in practice, since modelling the number of claims with the Poisson regression model leads to consistent estimates of $\lambda_{ij}$ under very general specification conditions (Gourieroux et al. (1984)). In practice, however, we will need to estimate $\tau^2$ before we obtain the credibility estimator $\hat{\theta}_i$.

One obvious generalization of the presented model is to increase the dimensions of $N_{ij}$, e.g. we may consider different types of claims related to the same policy or several types of claims arising from several policies held by the same customer. This is studied by Böhlmann et al. (2003). In the multivariate model presented below, we depart from their setting in that we assume that there are a priori differences, which are not necessarily the same, in each dimension, and, more importantly, we assume one specific risk parameter for each line of business.

In the multivariate case, we need to consider $P$ products (lines of business). $\theta_i = (\theta_{i1}, ..., \theta_{iP})$ is then a vector of random variables, where each component produces a particular individual risk profile for each type of insurance held by the same customer $i$. We assume $\mathbb{E}(\theta_i) = 1$ and $\text{COV}(\theta_i) = T$, with element $\tau^2_{kl}$ on row $k$ and column $l$. The covariance matrix of the risk profile random variables is constant for all individuals in the portfolio. No other restriction is initially imposed on the covariance between the risk profile of different components. The a priori differences now are $\lambda_{ijp}$ where
subscript $i$ refers to the customer, $j$ refers to the year and $p$ refers to the type of claim, and correspondingly, the risk profiles are $\theta_{ip}$ for every individual and product.

Again, using the best linear predictor principle in the multivariate setting, the credibility estimator follow immediately as:

$$\hat{\theta}_i = E(\theta_i) + \text{COV}(\theta_i, \mathbf{Y}_i) \mathbf{V}(\mathbf{Y}_i)^{-1} (\mathbf{Y}_i - E(\mathbf{Y}_i)),$$

(2.2)

where bold symbols are used to denote vectors. $\mathbf{Y}_i$ is $P$-dimensional and each component equals $\frac{1}{\lambda_{ip}} \sum_j \lambda_{ijp} Y_{ijp}$, with $Y_{ijp} = N_{ijp}/\lambda_{ijp}$ and $\lambda_{ip} = \sum_j \lambda_{ijp}$ and $\hat{\theta}_i = (\hat{\theta}_{i1}, ..., \hat{\theta}_{ip})$.

### 2.3 Multidimensional Bühlmann-Straub credibility model with time dependence

Time dependence can be introduced in the model by letting the individual risk profiles vary over time. Let us assume that $\Theta_{ij}$ is the (multivariate) random vector of individual risk profiles of customer $i$ in period $j$. We assume that $\Theta_{ij}$ are independent in $i$ with $E(\Theta_{ij}) = 1$. We assume a vector stationary process with covariance structure both between the $P$ types of policies and between the $J$ periods. To focus on the covariation between lines of business we follow and generalize Pinquet et al. (2001). We denote the covariances and autocorrelation coefficients as $\text{COV}(\theta_{ijp}, \theta_{i*j*p}) = \rho_{pp^*}^{j-j^*i} \tau_{pp^*}$ if $i = i^*$ and $\text{COV}(\theta_{ijp}, \theta_{i*j*p}) = 0$ otherwise, with $|\rho_{pp^*}^{j-j^*i}| \leq 1$. This holds for all individuals in the portfolio, $j, j^* = 1, ..., J$ and $p, p^* = 1, ..., P$. Theoretical, or intuitive, there are no restrictions for $\rho_{pp^*}$ and $\rho_{p^*p}$ to differ, but in our application below we put $\rho_{pp^*} = \rho_{p^*p}$, for simplicity.

In our application we have used a Weighted Least Squares (WLS) method, with the duration as weight:

$$\min_{\rho, \tau^2} \sum_{ijp} (N_{ijp} - \lambda_{ijp} \hat{\theta}_{ijp})^2 \omega_{ijp},$$

(2.3)

where $\hat{\theta}_{ijp}$ is the credibility estimator expression, with or without time dependence for individual $i$, year $j$ and product type $p$, which depends on the structural parameters.
2.4 Data study

In this section, we apply the models introduced in this paper to a portfolio of a Danish insurance company. By way of introduction, we need to discuss a few theoretical and implementational challenges, which arise when going from one-dimensional to multidimensional experience rating. Firstly, the number of underwritten lines of business may not be the same for all customers, but we would still like to apply the same framework. Secondly, we would like to be able to cope with a policyholder who has different lines of business issued at different points in time, so that the length of the claims history varies between lines. And thirdly, we have to take into account that durations of the contracts do not necessarily equal one year, because the policy may not be in force during the whole period. Our approach addresses all three issues and results in a practical scheme that can readily be applied in practice.

Actually, the first and third problem are the same in the ordinary multidimensional Bühlmann-Straub context, and is solved through using the estimated expected claim frequency as weight. The second problem is only a concern in time-dependent models and the problem is solved by treating the unobserved period as a missing value and replacing the missing information with full collective information.

The data used in the study come from the commercial portfolio of an active Danish insurance company and contain information on claim frequencies for various types of coverage: Fire, water damage, fungus and insect damage, glass damage and theft. For illustrative reasons, we have chosen to focus on only two types of coverage (lines of business): water damage and theft. The original data set contained information from 1995 to 2003 on 19,270 policies having both lines of business. For sake of simplicity, we excluded policies with less than 4 years of claims history in one of the business lines. This restriction reduced the number of policies to 10,212. Finally, we saved a random subsample of 25% of the policies for testing the estimated models, leaving 7,656 policies for analysis. In this final version, we had 288 claims on water damage and 739 on theft.

The data also contained an estimate of the annually a priori expected number of claims for each policy on each coverage $\lambda_{ijp}$, which were originally calculated using a Poisson regression model on the entire data set, see Dionne and Vanasse (1992) or McCullagh and Nelder (1989) for generalized linear models.
We estimated the unknown structural parameters in each credibility model by minimizing expression (2.3), and the results are shown in Table 2.1. Numerical minimization was achieved by an iterative method, with the one-dimensional estimates as starting values for the two-dimensional model parameters.

**TABLE 2.1**

Estimated parameters for each model using weighted least squares

<table>
<thead>
<tr>
<th>Model</th>
<th>$\tau^2_{11}$</th>
<th>$\tau^2_{22}$</th>
<th>$\tau^2_{12}$</th>
<th>$\rho_{11}$</th>
<th>$\rho_{22}$</th>
<th>$\rho_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple one-dimensional, Theft</td>
<td>0.377</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>One-dimensional with time, Theft</td>
<td>0.412</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.721</td>
<td>-</td>
</tr>
<tr>
<td>Simple one-dimensional, Water</td>
<td>-</td>
<td>1.686</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>One-dimensional with time, Water</td>
<td>-</td>
<td>1.712</td>
<td>-</td>
<td>-</td>
<td>0.811</td>
<td>-</td>
</tr>
<tr>
<td>Simple two-dimensional</td>
<td>0.447</td>
<td>1.702</td>
<td>0.619</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Two-dimensional with time</td>
<td>0.461</td>
<td>1.922</td>
<td>0.863</td>
<td>0.865</td>
<td>0.922</td>
<td>0.351</td>
</tr>
</tbody>
</table>

The estimates of Table 2.1 reveals a positive covariance between the two lines of business within the same policy ($\tau^2_{12} = 0.619$). This coincides with our expectations, since it corresponds to unmeasured risk factors which cannot be captured by the a priori differences alone. Further more, expected results of the time-correlation parameters were obtained. Risk profiles for consecutive time periods show positive correlation, as in Pinquet et al. (2001) and even for different types of coverage a positive correlation between consecutive periods was found ($\rho_{12} = 0.351$).

A comparison between the different models was made by applying the estimates of Table 2.1 to the testing sample containing 2,556 policies. The prediction power of the different models was assessed by calculating the weighted sum of squared residuals for the time-period $J + 1$. All models performed better than the a priori differences $\lambda_{ijp}$ alone, and the two-dimensional models showed slightly better results than the one-dimensional models.
2.4 Data study

We end this section with an investigation of the consequences of implementing the introduced models. This is done by choosing five arbitrary policies and realise how the different models respond to their claims history. Characteristics on the five policies can be seen in Table 2.2. Table 2.3 displays estimates of $\hat{\theta}_{ip}$ for the five policies using the different models. The column 'Crude estimate' shows the unsmoothed estimate $\frac{N_{ip}}{\lambda_{ip}}$.

<table>
<thead>
<tr>
<th>Policy</th>
<th>$N_{ij1}$</th>
<th>$N_{ij2}$</th>
<th>$\lambda_{ij1}$</th>
<th>$\lambda_{ij2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{0, 0, 0}</td>
<td>{0, 0, 1}</td>
<td>{0.008, 0.012, 0.011}</td>
<td>{0.248, 0.247, 0.247}</td>
</tr>
<tr>
<td>2</td>
<td>{0, 0, 1}</td>
<td>{0, 0, 0}</td>
<td>{0.102, 0.099, 0.097}</td>
<td>{0.102, 0.102, 0.084}</td>
</tr>
<tr>
<td>3</td>
<td>{1, 1, 1}</td>
<td>{1, 0, 0}</td>
<td>{0.438, 0.430, 0.422}</td>
<td>{0.105, 0.108, 0.107}</td>
</tr>
<tr>
<td>4</td>
<td>{0, 0, 0}</td>
<td>{0, 0, 0}</td>
<td>{0.111, 0.109, 0.108}</td>
<td>{0.014, 0.014, 0.014}</td>
</tr>
<tr>
<td>5</td>
<td>{0, 0, 0}</td>
<td>{1, 0, 0}</td>
<td>{0.024, 0.023, 0.023}</td>
<td>{0.169, 0.169, 0.169}</td>
</tr>
</tbody>
</table>
TABLE 2.3

Credibility estimates using the proposed models for the five policies

<table>
<thead>
<tr>
<th>Policy</th>
<th>Line of Business</th>
<th>Crude estimate</th>
<th>One-dimensional Simple</th>
<th>One-dimensional With time</th>
<th>Two-dimensional Simple</th>
<th>Two-dimensional With time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Theft</td>
<td>Water</td>
<td>0.988</td>
<td>0.993</td>
<td>1.060</td>
<td>1.151</td>
<td></td>
</tr>
<tr>
<td>2 Theft</td>
<td>Water</td>
<td>1.350</td>
<td>1.194</td>
<td>1.366</td>
<td>1.186</td>
<td>1.317</td>
</tr>
<tr>
<td>3 Theft</td>
<td>Water</td>
<td>0.673</td>
<td>0.773</td>
<td>0.946</td>
<td>0.857</td>
<td></td>
</tr>
<tr>
<td>4 Theft</td>
<td>Water</td>
<td>0.932</td>
<td>0.954</td>
<td>0.770</td>
<td>0.900</td>
<td></td>
</tr>
<tr>
<td>5 Theft</td>
<td>Water</td>
<td>1.973</td>
<td>1.448</td>
<td>1.127</td>
<td>1.424</td>
<td>1.311</td>
</tr>
</tbody>
</table>

Focusing on the first policy, we see that for the theft coverage the one-dimensional models give a rebate on the premium, while the two-dimensional models suggest a raise in the premium. This is because the individual history from the second coverage, water damage, is taken into account. It can also be seen that the two-dimensional model with time-dependence suggests a bigger raise in the theft line, since the claim in the water line appeared in the previous year (the two-dimensional model without time-dependence weights all claims equally). If we consider the water damage line for the same policy, both time-dependence models suggest an increase of more than 30% in the premium, while models without time-dependence suggest an increase of around 19%. An opposite scenario in the context of time-dependence structure arises in the third
policy, in the water line of business. Here, the customer experienced a claim during his first year. This is shown by the time-dependence models, leading to lower values than the other two models without a time-dependent structure. Policies 2, 4 and 5 should be interpreted in a similar manner.

2.5 Conclusions

The claims history for a customer, or policy, is commonly used to adjust the future premium, and this is often implemented using a Bonus-Malus System. The idea behind multiline experience rating is to extend this concept so that the rating also depends on performance in other lines of business. The motivation behind this is that if a customer performs badly in one line, we will expect her to perform badly in all lines of business, to some extent, and vice versa for customers with none or few claims. The size of the impact on a particular line of business of claims made in other lines should be controlled by the correlation between the two lines, and the amount of observed exposure.

We have presented a model for multiline experience rating in this study. The model is applied for two lines of business, but can easily be extended to arbitrary numbers of dimensions and the time dependence can have a more general structure.

The model has been applied to data from a commercial insurance company in Denmark, and the results show that the models perform slightly better than the one-dimensional Bühlmann-Straub credibility model in terms of predicting error in a testing sample.
Chapter 3

Paper B

MULTIDIMENSIONAL CREDIBILITY WITH TIME EFFECTS: An Application to Commercial Business Lines

Abstract: This paper considers Danish insurance business lines, for which the pricing methodology recently has been dramatically upgraded. A costly affair, but nevertheless the benefits greatly exceed the costs; without a proper pricing mechanism, you are simply not competitive. We show that experience rating improves this sophisticated pricing method as much as it originally improved pricing compared to a trivial flat rate. Hence, it is very important to take advantage of available customer experience. We verify that recent developments in multivariate credibility theory improve the prediction significantly and we contribute to this theory with new robust estimation methods, for time (in-)dependency.
Introduction

In this paper credibility theory and experience rating mean more or less the same thing; however, strictly speaking credibility theory describes a theoretical model with a latent risk variable, while experience rating is the act of including observed experience in the rating process. This latter act is sometimes carried out in non-life insurance companies without a consistent theoretical model behind it. But since all experience rating in this paper is based on a theoretical model, we can more or less use the two expressions interchangeably. Credibility theory has a long tradition in actuarial science; we show in this paper that there is indeed a good reason for this. In our concrete application to Danish commercial business lines, we show that the use of experience rating is as important as the use of pricing as such. In other words, we double the quality of the price rating by the inclusion of credibility theory in the rating process. We also consider the recently developed method of multivariate experience rating, where the latent risk parameter is allowed to be multidimensional such that each dimension represents one cover from the business line, see Englund et al. (2008). We also introduce a method to estimate a time effect of this model. We show that this more general version of credibility theory gives better results than the results from classical one dimensional credibility theory. We follow the standard approach of actuarial practitioners and we only use frequency information in our credibility approach. However, the severity of experience claims should contain some valuable information as well, indicating that there might be even more to gain from credibility theory, if a robust and stable credibility method is developed incorporating severity information in the experience rating.

An early beginning of credibility theory appeared in Mowbray (1914) and Whitney (1918). After the elegant approach presented by Bühlmann (1967), and Bühlmann and Straub (1970), a large number of extensions have been derived. References can be made to Jewell (1974), Hachemeister (1975), Sundt (1979, 1981) and Zehnwirth (1985). See also Halliwell (1999), Greig (1999) and Bühlmann and Gisler (2005) for more comprehensive surveys.

Evolutionary models are not new in credibility theory. The idea is that recent claim information is more valuable than old claim information. This approach was introduced in the 1970’s for one-dimensional credibility models; see Gerber and Jones (1975b,a) and De Vylder (1976). Much of the work on the time-dependent models focused on credibility formulas of the updating type. These recursive estimators were introduced by
Mehra (1975) for credibility applications; and further developed by De Vylder (1977), Sundt (1981) and Kremer (1982). For the time dependence in this paper we use a multivariate generalization of the recursive credibility estimator of Sundt (1981), where the risk parameter itself is modelled as an auto-regressive process.

The paper is organized as follows. In ”Multidimensional Credibility Theory” we state the credibility model and the estimators in our multidimensional setup. The model is generalized in ”Evolutionary Effect” to incorporate an evolutionary effect, and a recursive credibility estimator is stated. The paper is finished with the results of an empirical data study, in ”An Application to Danish Commercial Business Lines”, from which we, in ”Discussion and Conclusions”, draw conclusions concerning the predictability of the credibility estimators.

**Multidimensional Credibility Theory**

In this section we repeat the multivariate credibility model of Englund et al. (2008) and define a new more robust variance estimator inspired by the elegant approach of Bühlmann and Gisler (2005). We consider only frequencies of claims.

**The Multidimensional Credibility Model**

We observe insurance claims $N_{ijk}$ with expected number of claims $\lambda_{ijk}$ for individual $i \in (1, \ldots, I)$, calendar year $j \in (1, \ldots, J)$ and coverage $k \in (1, \ldots, K)$. A dot, ‘·’, indicates summation over that index. Define the standardized number of claims as a vector of $k$-dimensions, $F_{ij} = \frac{N_{ij}}{\lambda_{ij}}$.

**Model Assumptions:**

(i) Given the vector of individual risk parameters $\Theta_i$, the insurance claims $N_{ij}$, are independent and Poisson-distributed with expected value $E[N_{ij} \mid \Theta_i] = \lambda_{ij}\Theta_i$. 
The expected value and (co-) variance of $F_{ij}$, conditional on $\Theta_i$, are:

$$E[F_{ij} \mid \Theta_i] = \Theta_i$$

$$\text{Cov}[F_{ij}, F_{ij}^T \mid \Theta_i] = S_i,$$

with diagonal elements

$$\text{Var}[F_{ijk} \mid \Theta_i] = \sigma_k^2(\Theta_i) \lambda_{ijk}$$

$$\text{Cov}[\Theta_i, \Theta_i^T] = T$$

(ii) The pairs $(\Theta_1, N_{1j}), (\Theta_2, N_{2j}), \ldots, (\Theta_I, N_{Ij})$ are independent, and $\Theta_1, \Theta_2, \ldots, \Theta_I$ are independent and identically distributed with

$$E[\Theta_i] = \begin{bmatrix} E[\Theta_{i1}] & E[\Theta_{i2}] & \cdots & E[\Theta_{iK}] \end{bmatrix}^T = \Theta_0.$$

The Multidimensional Credibility Estimator

The credibility estimators can be seen as projections in the Hilbert space of all square-integrable random variables, see e.g. Zehnwirth (1985). Hence, the best linear unbiased estimator of the multidimensional latent risk parameter $\Theta_i$, given the observed experience, is:

$$\hat{\Theta}_i = E[\Theta_i] + \text{Cov}[\Theta_i, F_i] \text{Cov}[F_i, F_i^T]^{-1} (F_i - E[F_i])$$ (3.1)

where

$$\text{Cov}[\Theta_i, F_i] = E[\text{Cov}[\Theta_i, F_i] \mid \Theta_i] + \text{Cov}[E[\Theta_i \mid \Theta_i], E[F_i \mid \Theta_i]]$$

$$= 0 + \text{Cov}[\Theta_i, \Theta_i^T] = T$$

and

$$\text{Cov}[F_i, F_i^T] = E[\text{Cov}[F_i, F_i^T] \mid \Theta_i] + \text{Cov}[E[F_i \mid \Theta_i], E[F_i \mid \Theta_i]^T]$$

$$= E[S_i] + \text{Cov}[\Theta_i, \Theta_i^T] = SW_i^{-1} + T$$

The vector of standardized number of claims is $F_i = \begin{bmatrix} F_{i1}, & F_{i2}, & \cdots, & F_{iK} \end{bmatrix}^T$, where $F_{ik} = \sum_{j=1}^{J} \lambda_{ijk} F_{ijk}$. The weight matrix $W_i$ is a diagonal matrix with $\lambda_{ik} = \sum_{j=1}^{J} \lambda_{ijk}$ in the $k^{th}$ diagonal element. The credibility weight $\alpha_i$ is estimated by

$$\hat{\alpha}_i = T(SW_i^{-1} + T)^{-1} = TW_i(TW_i + S)^{-1}.$$ This estimator differs from Bühlmann and Gisler (2005), but is able to use information from additional coverages to calculate the
individual risk parameter, even if we lack information in a specific (inactive) coverage. The resulting credibility estimator can be seen as both a weighted sum of individual and collective claim information, and as a linear regression:

\[
\hat{\Theta}_i = \hat{\alpha}_i F_i + (I - \hat{\alpha}_i) \Theta_0
\]

\[
= \Theta_0 + \hat{\alpha}_i (F_i - \Theta_0)
\]

(3.2)

\(I\) is the identity matrix. Note that the one-dimensional credibility estimator is a trivial special case of (3.2).

Estimation of the Parameters

The estimation of the parameters \(\Theta_0, T\) and \(S\) is inspired by the elegant estimators presented by Bühmann and Gisler (2005) for a different but related multivariate credibility problem. We estimate the elements of \(\Theta_0\) by

\[
\hat{\theta}_0 k = \left(\sum_{i=1}^I \hat{\alpha}_{ik}\right)^{-1} \sum_{i=1}^I \hat{\alpha}_{ik} F_{i.k}.
\]

The estimator of \(S\) is a diagonal matrix with \(\hat{\sigma}^2_k\) as \(k^{th}\) diagonal:

\[
\hat{\sigma}^2_k = \frac{1}{J_k - I_k} \sum_{i=1}^I \sum_{j=1}^J \lambda_{ijk} (F_{ij.k} - F_{i.k})^2.
\]

\(J_k\) is the sum of the number of yearly observations for all \(I_k\), individuals with an active coverage \(k\). Note that we assume the occurrence of claims to be Poisson-distributed and that this, theoretically, gives us \(\sigma^2_k = \theta_{0k}\). The estimation of \(\theta_{0k}\) precedes the estimation of \(\sigma^2_k\), due to the estimation of the credibility weights, wherefore we use two different estimators for this. The preliminary estimator of \(T, \tilde{T}\), has diagonal

\[
\tilde{\tau}^2_{kk} = c_k \left(\frac{1}{I_k - 1} \sum_{i=1}^I \frac{\lambda_{i.k}}{\lambda_{..k}} (F_{i.k} - F_{..k})^2 - \frac{\sigma^2_k}{\lambda_{..k}}\right), \quad F_{..k} = \left(\sum_{i=1}^I \lambda_{i.k}\right) \sum_{i=1}^I \lambda_{i.k} F_{i.k}
\]

and for \(k \neq k'\):

\[
\tilde{\tau}^2_{kk'} = \frac{c_k}{I_{kk'}} \sum_{i=1}^I \frac{\lambda_{i.k}}{\lambda_{..k}} (F_{i.k} - F_{..k}) (F_{i.k'} - F_{..k'})
\]

Here \(I_{kk'}\) is all individuals with both coverage \(k\) and \(k'\) active and \(\lambda_{..k} = \sum_{i=1}^I \lambda_{i.k}\) and \(\lambda_{i,k} = \sum_{j=1}^J \lambda_{ikj}\). The parameter \(c_k\) in the formulas above takes the expression
\[ c_k = (I - 1) \left( \sum_{i=1}^{I} \lambda_{i;k} \lambda_{i;k}^{-1} \right)^{-1}. \]

Since \( \tilde{\tau}_{kk'}^2 \) may be less than zero the final diagonal estimator is defined as: \( \tau_{kk'}^2 = \max \{ \tilde{\tau}_{kk'}^2, 0 \} \). If \( \tilde{\tau}_{kk'}^2 \leq \tilde{\tau}_{kk'}^2 \tilde{\tau}_{kk'}^2 \) we again follow the suggestions in Bühlmann and Gisler (2005) and replace \( \tilde{\tau}_{kk'}^2 \) by:

\[
\tau_{kk'}^2 = \text{signum} \left( \frac{\tilde{\tau}_{kk'}^2 + \tilde{\tau}_{kk'}^2}{2} \right) \min \left( \frac{|\tilde{\tau}_{kk'}^2 + \tilde{\tau}_{kk'}^2|}{2}, \sqrt{\tilde{\tau}_{kk'}^2 \tilde{\tau}_{kk'}^2} \right)
\]

Even the corrected estimator $\overline{T}$ resulting from (3.3) is not necessarily positive semidefinite, for $K > 2$. Therefore, to make the credibility estimation meaningful and achieve the positive semidefiniteness, we adjust the estimator one last time. We compute the eigenvalue decomposition of $\overline{T}$, put a floor of zero on the diagonal eigenvalue matrix, and compose the new eigenvalue matrix and we reconstruct the final estimator $\hat{T}$. This estimator is more robust to calculate than the original estimator of Englund et al. (2008). The only parameter left to estimate now is $\lambda_{ijk}$, which we assume to be estimated from some pricing model developed by the company. This can be done at various levels of sophistication.

**Evolutionary Effects**

A time-independent credibility estimator implies that the risk parameter for each coverage and policyholder is constant over time and the estimator will therefore treat old and new claim information equally. However, this might be quite insufficient in some cases, e.g., the abilities of a car driver are not constant. Hence, instead of assuming that the risk characteristics are given once and for all by the parameter $\Theta_i$, we now suppose that the risk characteristics of year $s$ are given by an unknown parameter $\Theta_{is}$, and that the dependence between $\Theta_{is}$ and $\Theta_{it}$ decreases as $|s - t|$ increases. This is done by modeling the risk parameter as a stationary process, more specifically, as an auto-regressive process. The interpretation of this approach is that new claim information will affect the claim prediction more than old claim information. We apply the same time dependency model as Englund et al. (2008), since it is intuitive and easy to interpret. When it comes to estimation, we follow the recursive estimation principle of Sundt (1981), which is more stable and easier to implement than the estimator of Englund et al. (2008).
The Time-Dependent Model and Estimator

As a model for the time-dependent credibility estimator we generalize the model stated in "The Multidimensional Credibility Model" by introducing a new index, and thereby incorporate a time-dependence and a correlation structure of the latent risk parameter.

We use the same model as in Englund et al. (2008), i.e. we assume the insurance claims $N_{ijk}$ to be independent and Poisson-distributed, given $\Theta_{ijk}$, with expected value $E[N_{ijk} | \Theta_{ijk}] = \lambda_{ijk} \Theta_{ijk}$ and the process $\{\Theta_{ijk}\}_{j=1,...,J}$ to be an auto-regressive process with lag 1, with the covariance structure: $\text{Cov} [\Theta_{ijk}, \Theta_{ij'k'}^T] = \tau_{kk'}^2 (\rho_{kk'})^{|j-j'|}$ if $i = i'$ and $\text{Cov} [\Theta_{ijk}, \Theta_{ij'k'}^T] = 0$ otherwise, with $\tau_{kk'}^2 = \tau_{k'k}^2$, $\rho_{kk'} = \rho_{k'k}$ and $|\rho| \leq 1$.

The recursive estimator of the latent risk parameter is the result of a straightforward generalization of a theorem which is found and proved in Sundt (1981) for the one-dimensional case, and in Bühmann and Gisler (2005) for the multidimensional case. The estimator is

$$\hat{\Theta}_{i(j+1)} = \mathbf{a} \left( \hat{\alpha}_{ij} \mathbf{F}_{ij} + (I - \hat{\alpha}_{ij}) \hat{\Theta}_{ij} \right) + (I - \mathbf{a}) \Theta_0 \quad (3.4)$$

where $\mathbf{F}_{ij} = \left[ F_{ij1}, F_{ij2}, \cdots, F_{ijK} \right]^T$, as previously noted. $\mathbf{a}$ is a diagonal matrix of elements in $[0,1)$ driving the stationary auto-regressive processes. The credibility weight is $\hat{\alpha}_{ij} = T_{ij} W_{ij} (T_{ij} W_{ij} + S)^{-1}$, which is the same as in (3.2), but with the annual index $j$. The matrix $T_{i(j+1)}$ is defined as

$$T_{i(j+1)} = \mathbf{a}^2 (I - \hat{\alpha}_{ij}) T_{ij} + (I - \mathbf{a}^2) (T_1 + S)$$

The starting values for the recursion are $\hat{\Theta}_{i1} = \hat{\Theta}_0$ and $T_1 = \hat{T}$, as in "The Multidimensional Credibility Estimator". The elements in $\mathbf{a}$ are estimated by performing a minimization of the residual sum of squares (see "The Results") based on the one-dimensional time-dependent credibility estimator for each coverage. Note that the estimator deals with individual missing values automatically.
An Application to Danish Commercial Business Lines

In this section we present an extensive study of eighteen different estimators of each line of business in the four-dimensional commercial data set. We calculate the simple flat rate, which no actuaries would recommend in practice. The comparison of the performance of respectively the flat rate and the sophisticated estimate \( \hat{\lambda}_{ijk} \) gives us the possibility to evaluate the extra information that can be extracted from the observed experience. The remaining 16 estimators are credibility estimators with and without time effect based on varying dimensions of the data set: one, two, three or four dimensions.

The Data Set

The data set includes four coverages: Fire, Glass, Other and Water, and consists of insurance information for 2842 policyholders. We have available the estimated (expected) claim frequency, the duration \( \omega_{ijk} \), and the number of reported claims \( n_{ijk} \), for each policyholder, coverage and year. The estimated number of claims, \( \hat{\lambda}_{ijk} \), is the product of the estimated claim frequency and the duration. The estimated claim frequency was originally determined via a Poisson regression, based on a large number of covariates. We have up to eight years of information. Only few policyholders have eight years of information for all four coverages. In Table 3.1 we present a comparison of the expected and the reported number of claims for the different coverages. The number of expected claims is lower than the reported in Glass and Other, while it is the opposite situation for Fire and Water. However, the total number of expected claims is rather close to the total number of reported ones.

The Results

Let us first consider the situation without a time effect. For every coverage, for example Fire, we have one credibility estimator based on all four dimensions, three credibility estimators based on three dimensions, three credibility estimators based on two dimensions, and one credibility estimator based on one dimension. That is eight estimators. We also consider the same eight credibility estimators with time effect, therefore we have a total of sixteen credibility estimators for each coverage. In the following we evaluate their
Table 3.1: The expected and the reported total number of claims

<table>
<thead>
<tr>
<th>Insurance coverage</th>
<th>Total number of expected claims</th>
<th>Total number of reported claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building - Fire</td>
<td>809</td>
<td>787</td>
</tr>
<tr>
<td>Building - Glass</td>
<td>7455</td>
<td>7797</td>
</tr>
<tr>
<td>Building - Other</td>
<td>1980</td>
<td>2004</td>
</tr>
<tr>
<td>Building - Water</td>
<td>2763</td>
<td>2731</td>
</tr>
</tbody>
</table>

The expected and the reported total number of claims in the different coverages.

performance and compare them to the flat rate and to the sophisticated rating that does not take advantage of experience rating.

1d CE: the one-dimensional credibility estimator, influenced only by claims occurring in that coverage,

2d CE: the two-dimensional credibility estimator, influenced by claims occurring in that and one other coverage,

3d CE: the three-dimensional credibility estimator, influenced by claims occurring in that and two other coverages,

4d CE: the four-dimensional credibility estimator, influenced by claims occurring in all four coverages.

A time-dependent credibility estimator is denoted with a 't', e.g. 3d tCE for the three-dimensional time-dependent credibility estimator.

We use the residual sum of squares (SS) as our performance measure:

$$SS = \sum_{i=1}^{I} \left( \hat{N}_{ijk} - n_{ijk} \right)^2, \quad k = 1, 2, 3, 4$$
where \( \hat{N}_{ijk} \) is the estimated number of claims for individual \( i \) in coverage \( k \), i.e. either \( \hat{N}_{ijk} = \tilde{\lambda}_k \), \( \hat{N}_{ijk} = \hat{\lambda}_{ijk} \) or \( \hat{N}_{ijk} = \hat{\lambda}_{ijk}\hat{\theta}_{ik} \). The estimate \( \tilde{\lambda}_k \) is calculated with the estimator \( \bar{\lambda}_k = \frac{1}{\sum_{i=1}^I \sum_{j=1}^J n_{ijk}(\sum_{i=1}^I \sum_{j=1}^J \omega_{ijk})^{-1}} \), which is the so-called mean value estimator (MVE), also called the flat rate. This estimator is introduced to get a better notion of the improvement in prediction received by using any of the credibility estimators. With the MVE every individual is considered having equal risk, which means that the estimator is one of the simplest possible. It is neither affected by covariates nor by individual claim record. The estimate of the latent risk parameter \( \hat{\theta}_{ik} \) is calculated with any of the credibility estimators, and \( n_{ijk} \) is the reported number of claims in the validation data set, which consists of all individual information in a specific year \( j = J + 1 \), for each policyholder. The results are presented in Figure 3.1. In this figure the SS values, normalized (divided) with the SS value for the present estimator \( \hat{\lambda}_{ijk} \), for 18 different estimators are plotted, which from the left are the mean value estimator \( \bar{\lambda}_k \), the present estimator \( \hat{\lambda}_{ijk} \) and the 16 different credibility estimators \( \hat{\lambda}_{ijk}\hat{\theta}_{ik} \). In Table 3.2 we present a sheet to make the interpretation of Figure 3.1 easier. Table 3.3 contains the resulting SS values represented in Figure 3.1. The numbers on the horizontal axis of the subfigures in Figure 3.1 are keyed to Table 3.2.

For the four-dimensional case we present the estimated covariance matrix \( \hat{T} \) to show how the claim experience is connected between the different coverages, which corresponds to the correlation matrix \( C \):

\[
\hat{T} = \begin{bmatrix}
2.708 & 0.1537 & 0.2315 & 0.1744 \\
0.1537 & 1.692 & 0.5838 & 0.2394 \\
0.2315 & 0.5838 & 1.444 & 0.1015 \\
0.1744 & 0.2394 & 0.1015 & 1.025
\end{bmatrix},
\quad C = \begin{bmatrix}
1 & 0.0718 & 0.1171 & 0.1047 \\
0.0718 & 1 & 0.3735 & 0.1818 \\
0.1171 & 0.3735 & 1 & 0.0835 \\
0.1047 & 0.1818 & 0.0835 & 1
\end{bmatrix}.
\]

According to Figure 3.1 we see that a credibility estimator is the best choice for claim prediction in every one of the four coverages. However, the optimal choice of estimator, for each specific coverage, varies, and a universal answer to which is the best estimator can therefore not be found according to Figure 3.1. The best estimators for the four different coverages are:
Table 3.2: Table to help interpret Figure 3.1. MVE stands for mean value estimator, Present for the present covariate-based estimator, (t)CE for credibility estimators with and without time dependence. The name of the coverages in the parentheses show the additional coverages used.

<table>
<thead>
<tr>
<th>x-Axis</th>
<th>Building - Fire</th>
<th>Building - Glass</th>
<th>Building - Other</th>
<th>Building - Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MVE</td>
<td>MVE</td>
<td>MVE</td>
<td>MVE</td>
</tr>
<tr>
<td>2</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>3</td>
<td>1d CE</td>
<td>1d CE</td>
<td>1d CE</td>
<td>1d CE</td>
</tr>
<tr>
<td>4</td>
<td>2d CE (Glass)</td>
<td>2d CE (Fire)</td>
<td>2d CE (Fire)</td>
<td>2d CE (Fire)</td>
</tr>
<tr>
<td>5</td>
<td>2d CE (Other)</td>
<td>2d CE (Other)</td>
<td>2d CE (Glass)</td>
<td>2d CE (Glass)</td>
</tr>
<tr>
<td>6</td>
<td>2d CE (Water)</td>
<td>2d CE (Water)</td>
<td>2d CE (Water)</td>
<td>2d CE (Other)</td>
</tr>
<tr>
<td>7</td>
<td>3d CE (Glass, Other)</td>
<td>3d CE (Fire, Other)</td>
<td>3d CE (Fire, Glass)</td>
<td>3d CE (Fire, Glass)</td>
</tr>
<tr>
<td>8</td>
<td>3d CE (Glass, Water)</td>
<td>3d CE (Fire, Water)</td>
<td>3d CE (Fire, Water)</td>
<td>3d CE (Fire, Other)</td>
</tr>
<tr>
<td>9</td>
<td>3d CE (Other, Water)</td>
<td>3d CE (Other, Water)</td>
<td>3d CE (Glass, Water)</td>
<td>3d CE (Glass, Other)</td>
</tr>
<tr>
<td>10</td>
<td>4d CE (all)</td>
<td>4d CE (all)</td>
<td>4d CE (all)</td>
<td>4d CE (all)</td>
</tr>
<tr>
<td>11</td>
<td>1d tCE</td>
<td>1d tCE</td>
<td>1d tCE</td>
<td>1d tCE</td>
</tr>
<tr>
<td>12</td>
<td>2d tCE (Glass)</td>
<td>2d tCE (Fire)</td>
<td>2d tCE (Fire)</td>
<td>2d tCE (Fire)</td>
</tr>
<tr>
<td>13</td>
<td>2d tCE (Other)</td>
<td>2d tCE (Other)</td>
<td>2d tCE (Glass)</td>
<td>2d tCE (Glass)</td>
</tr>
<tr>
<td>14</td>
<td>2d tCE (Water)</td>
<td>2d tCE (Water)</td>
<td>2d tCE (Water)</td>
<td>2d tCE (Other)</td>
</tr>
<tr>
<td>15</td>
<td>3d tCE (Glass, Other)</td>
<td>3d tCE (Fire, Other)</td>
<td>3d tCE (Fire, Glass)</td>
<td>3d tCE (Fire, Glass)</td>
</tr>
<tr>
<td>16</td>
<td>3d tCE (Glass, Water)</td>
<td>3d tCE (Fire, Water)</td>
<td>3d tCE (Fire, Water)</td>
<td>3d tCE (Fire, Other)</td>
</tr>
<tr>
<td>17</td>
<td>3d tCE (Other, Water)</td>
<td>3d tCE (Other, Water)</td>
<td>3d tCE (Glass, Water)</td>
<td>3d tCE (Glass, Other)</td>
</tr>
<tr>
<td>18</td>
<td>4d tCE (all)</td>
<td>4d tCE (all)</td>
<td>4d tCE (all)</td>
<td>4d tCE (all)</td>
</tr>
</tbody>
</table>
Table 3.3: Table showing the resulting SS values of the validation analysis. The values are represented in Figure 3.1.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building</td>
<td>138.97</td>
<td>127.38</td>
<td>124.96</td>
<td>125.02</td>
<td>124.91</td>
<td>124.45</td>
<td>124.93</td>
<td>124.37</td>
<td>124.51</td>
</tr>
<tr>
<td>- Fire</td>
<td>2637.1</td>
<td>2061.2</td>
<td>1640.2</td>
<td>1641.7</td>
<td>1629.6</td>
<td>1648.8</td>
<td>1630.9</td>
<td>1649.7</td>
<td>1639.1</td>
</tr>
<tr>
<td>- Glass</td>
<td>830.12</td>
<td>744.38</td>
<td>637.94</td>
<td>634.70</td>
<td>635.61</td>
<td>637.86</td>
<td>632.83</td>
<td>634.87</td>
<td>636.18</td>
</tr>
<tr>
<td>- Other</td>
<td>376.82</td>
<td>313.42</td>
<td>304.38</td>
<td>304.07</td>
<td>304.12</td>
<td>304.38</td>
<td>303.81</td>
<td>304.04</td>
<td>304.07</td>
</tr>
<tr>
<td>- Water</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>Building</td>
<td>124.39</td>
<td>125.02</td>
<td>125.08</td>
<td>124.96</td>
<td>124.65</td>
<td>124.99</td>
<td>124.56</td>
<td>124.71</td>
<td>124.59</td>
</tr>
<tr>
<td>- Fire</td>
<td>1640.0</td>
<td>2116.7</td>
<td>1715.6</td>
<td>1704.3</td>
<td>1718.2</td>
<td>1703.5</td>
<td>1716.7</td>
<td>1705.3</td>
<td>1704.7</td>
</tr>
<tr>
<td>- Glass</td>
<td>633.54</td>
<td>674.46</td>
<td>664.46</td>
<td>663.24</td>
<td>667.76</td>
<td>660.48</td>
<td>664.66</td>
<td>663.69</td>
<td>661.05</td>
</tr>
<tr>
<td>- Other</td>
<td>303.70</td>
<td>303.30</td>
<td>301.28</td>
<td>301.68</td>
<td>301.71</td>
<td>301.28</td>
<td>301.28</td>
<td>301.74</td>
<td>301.31</td>
</tr>
</tbody>
</table>
Figure 3.1: The normalized Sum of Squares (nSS) values for the different estimators.

**Fire**: The three-dimensional estimator with *Glass* and *Water* as additional coverages.

**Glass**: The two-dimensional estimator with *Other* as additional coverages.

**Other**: The three-dimensional estimator with *Fire* and *Glass* as additional coverages.

**Water**: The three-dimensional time-dependent estimator with *Fire* and *Other* as additional coverages.

**Discussion and Conclusions**

The conclusion of our empirical study is that experience rating is extremely useful for pricing. While this hardly is a surprising conclusion, it might be surprising that we are able to present a situation in which the inclusion of experience rating gives
an extra improvement of the same order of magnitude as the improvement obtained from leaving the trivial flat rate and entering sophisticated rating principles without experience rating. We also conclude that our multivariate credibility approach indeed is capable of improving the quality of estimation compared to classical one-dimensional credibility theory. However, the most important thing is to use experience rating; the multivariate approach is just an extra improvement. Note that adding the time effect does not generally improve prediction in our case. The reason seems to be that our average duration of information is too short to divide it into old and new observations. However, adding a time effect does not harm our prediction either. We therefore expect that adding a time effect would improve prediction in our case if, for example, an average of ten years of observed history were available.
Chapter 4

Paper C

ASYMMETRIC INFORMATION: Evidence from the Automobile Insurance Market

Abstract: This paper puts one of the predictions of adverse-selection models to the test, using data from the Danish automobile insurance market. We find evidence that is consistent with the presence of informational asymmetries; policyholders choosing more comprehensive insurance coverage are associated with more claims. A significant correlation between claims and coverage is found for policyholders purchasing fixed-priced products, but not for those with an experience-rated premium. We find no evidence of moral hazard.
4.1 Introduction

Asymmetric information is present everywhere! Insurance policyholders certainly know their preferences better than the insurer; they know if they prefer a blue car instead of a red one. They might also be more aware of their behaviour or reactions, e.g. if they get stressed or are relaxed while driving. Information about the preferences or the conduct of the policyholders can be very difficult, or even impossible, to obtain by the insurer, and this information might not be relevant at all. But, if it is correlated with the risk (the probability or the severity of claims) then the policyholders do have a better knowledge of their risk than the insurer, and asymmetric information is present. On the other hand the insurer might be better able to evaluate the given collective information, why it is unclear which way the asymmetry goes. In this paper we investigate if there are any useful asymmetries left for the policyholder to gain from, or if the insurer has managed to take all relevant risk information into account when setting the price. Hence, we are interested in if there is such a phenomenon as good or bad, better or worse, policyholders, and if we can distinguish these already at the time of purchase.

There are mainly two special cases of asymmetric information: adverse selection and moral hazard. Adverse selection refers to a market process in which poor products or customers are more likely to be selected, due to asymmetric information. The asymmetry can be caused by private information of one of the parties or by regulations or social norms which prevent e.g. the insurer from using certain characteristic to set the prices. Two ways to model adverse selection are with signaling games and screening games.\footnote{Spence (1973)} Moral hazard means that the policyholder may have some degree of control over the probability or size of the occurrence, of an insured event, due to the care taken. Care can be interpreted as money, but also diligence, mental concentration, or intensity of effort.\footnote{Ehrlich and Becker (1972)}

Under adverse selection, without moral hazard, the degree of coverage chosen by the policyholder is based on the ex-ante (before the event of an accident) assessment a policyholder makes of her likelihood of having an accidents. Each policyholder is fully described by a single number, e.g. her probability of accident, since neither the frequency nor the claim size are functions of the actions of the policyholders. It follows that the policyholders with a high propensity of experiencing accidents choose more
coverage than those with a low one. In the presence of moral hazard, the causality works in the opposite direction caused by the unobservability of efforts to prevent accidents, hence the choice of coverage affects the efforts to prevent accidents. The probability of accidents can now be seen as a function of the level of care taken, which is unobservable by the insurer. Generous coverage reduces the expected cost of an accident for the policyholder and therefore the incentives for safety. In this case, more coverage may lead to more accidents. In the end, both moral hazard and adverse selection predict a single-period positive correlation between risk and coverage within a risk class. Nevertheless risk-coverage correlation is a good test for asymmetric information, since policyholders that choose different levels of deductible must be different, either in risk, risk aversion or in both.

While evidence of a risk-coverage correlation is found in studies of some insurance markets, the findings have been rather diverse in the automobile insurance market. Three initial studies suggest the existence of coverage-risk correlation. But their findings were challenged by the research that followed. Chiappori and Salanie use more refine methods and cannot reject a zero correlation between higher insurance coverage and more accidents in their French data. Nor can Dionne et al. find evidence for asymmetric information using data from Quebec, and the same goes for Saito on the Japanese automobile insurance market. Cohen finds a positive risk-coverage correlation in the Israeli automobile insurance market. These scattered conclusions might, of course, be dependent on differences in the specific data sets, the definition of claim, the statistical tests chosen and the market structures. Kim et al., for instance, find evidence of asymmetric information in the automobile insurance market of Korea, using a multinomial measurement but not with the more traditional dichotomous measurement. But, there is also empirical evidence that risk-averse policyholders tend to buy more coverage and yet be more cautious, hence they are less risky. Sonnenholzner and Wambach find that patient policyholders use high effort and buy high insurance cover-

---

5 Chiappori and Salanié (2000) and Dionne et al. (2001)
6 Eventhough Chiappori et al. (2006) conclude that a positive coverage-risk correlation would be a natural and robust consequence of the assumption of a competitive market.
7 Dionne et al. (2001) and Saito (2006a)
8 Cohen (2005)
9 Kim et al. (2009)
10 de Meza and Webb (2001)
age, whereas the impatient ones do not. The effects of such preference-based selection can counteract, to some extent, the effects of “traditional” adverse selection, and the sign of the correlation between coverage and risk may therefore be indeterminate.

We see that most of the previous empirical analyses to find evidence of asymmetric information are focused on the coverage-risk correlation, described above. This paper contributes to the literature by reexamining whether such coverage-risk correlation exists. We find new evidence of asymmetric information for some types of policyholders, but not for all. These findings are consistent with the previous literature, and help to explain why some of the previous results may seem scattered. We do this on unique data from the Danish automobile insurance market, in which we have access to all data available for an insurer. This is important since we can use the same pricing scheme as the insurer, and avoid spurious effects due to under or over parameterization. We focus on the coverage of material damages, called collision coverage in some countries, since it is the only coverage of the automobile insurance product for which the policyholder chooses level of deductible. Based on Cohen’s findings we also investigate the differences between new and repeated policyholders, but with a focus on new policyholders who are more likely to have an advantage in terms of amount of information over the insurer. While Cohen looks separately at pools of policyholders with and without substantial driving experience prior to joining the insurer, this has been impossible in this paper since such information is not included in the data set. We find a significant coverage-risk correlation for both new and repeated policyholders with a fixed priced product, but not for those with an experience-rated product. Unlike Cohen, we find this correlation when the main part of the population, not the minority, have a low deductible. This finding is coherent with the Rothschild-Stiglitz model, saying that equilibrium exists if and only if there are enough policyholders with a high accident probability. Rothschild and Stiglitz assumes adverse selection without moral hazard.

The paper is arranged as follows: in section 4.2 we present the economical model of asymmetric information and the procedures for testing the assumptions. In section 4.3 we introduce some empirical issues such as the market, the products and the data which are under investigation. The results are presented in section 4.4. These are finally discussed and concluded in section 4.5 and 4.6.

---

11Sonnenholzner and Wambach (2006)
12Rothschild and Stiglitz (1976)
4.2 Economical Model

In this paper we deal with exclusive contracts. This means that we only consider markets in which the insurer can impose an exclusive relationship. A policyholder can e.g. not have more than one automobile insurance, due to possible moral hazard. This allows firms to implement non-linear and, in particular, convex pricing schemes, which are typically needed under asymmetric information.\(^{13}\) Furthermore, the policies offered by the insurer, as for all policies offered in the Danish automobile market, are one-period policies with no committment on the part of either the insurer or the policyholder. This means that each policyholder can switch to another insurance company in the next period if they wish to. In the Danish market, insurers are not required to share information about their policyholders with other insurers, and nor do they. Despite the inability to commit, both parties can sign a first period contract that should be followed by a second-period contract which is experience-rated. This is described more in detail later on. In this context the literature on adverse selection in a competitive setting builds on the influential paper of Rothschild and Stiglitz.\(^{14}\) Following Chiappori and Salanié, one can argue that three hypotheses seem fairly robust under adverse selection: (1) policyholders are likely to be offered menus of contracts, among which they are free to choose; (2) contracts with more comprehensive coverage are sold at a higher (unit) premium (so called convex pricing); (3) contracts with more comprehensive coverage are chosen by policyholders with higher expected probability of accident (or claim frequency).\(^{15}\)

The first prediction is essentially qualitative. It is also quite general, and holds true for different types of adverse selection; individuals may differ by their risk, but also by their wealth, preferences, risk aversion, etc. Hence, this hypothesis is not likely to provide very powerful tests.

The second prediction, in most circumstances, is a direct consequence of individual rationality and risk aversion in general: even in the absense of adverse selection, an individual will not choose a contract with higher deductible unless its unitary price is lower, at least if pricing is approximately fair.

The third prediction suggests however a fairly simple test; a positive correlation

\(^{13}\)Chiappori (2000)
\(^{14}\)Rothschild and Stiglitz (1976)
\(^{15}\)Chiappori and Salanié (2000)
Evidence of Asymmetric Information

between coverage and claim frequency should be observed on observationally identical policyholders. This will be the focus of the present paper.

However, the simplicity of testing prediction 3 has its cost; the risk-coverage correlation does not allow the identification of the type of informational asymmetry involved, if any. One can argue that the origin of claim does not really matter for the pricing process, as long as the insurer cannot observe the underlying parameter of characteristics. The insurer is liable for a claim regardless whether it is due to endogenous risk or lack of care. However, if one is interested in dealing with and reducing the informational asymmetry it is important to know the difference. Also theoretically this will affect our assumptions and conclusions since a Rothschild-Stiglitz equilibrium exists if and only if there are enough high risk agents in the economy, under the assumption of only adverse selection. When both problems (adverse selection and moral hazard) are present simultaneously, this condition is no longer true; an equilibrium may, or may fail to, exist whatever the proportion of agents of different types, depending on the parameters or the model.\(^{16}\)

The usual assumptions used in models on adverse selection are maintained in this paper: the individual information is costly to observe by the insurer, the provision of the insurance is costless and the insurer is risk-neutral while the policyholders are risk-averse, having identical twice-differentiable, increasing and strictly-concave utility functions \(U\). However, we relax the usual assumption that neither the frequency nor the claim size are functions of the actions of the policyholders, until we have investigated the possible existence of moral hazard further, since a correlation between coverage and accidents is consistent not only with the existence of adverse selection but also with the existence of moral hazard.\(^{17}\)

Hence our paper will firstly concentrate on testing prediction 3 of the Rothschild-Stiglitz model, without checking whether it holds because of adverse selection of moral hazard. Secondly we will test both for moral hazard and learning, respectively.

\(^{16}\)Chassagnon and Chiappori (1997)

\(^{17}\)For recent work investigating whether drivers' precautions are influenced by the accident costs they expect to bear, see Abbring et al. (2003) and Cohen and Einav (2003).
4.2 Economical Model

The Procedure

We want to test the conditional independence of the choice of coverage and the occurrence of accidents, in this case reported claims. Here conditional means conditional on all variables observed by the insurer. To do this we will test the hypothesis of zero correlation, $\rho = 0$, in a bivariate probit model. The motivation of the choice of a bivariate probit model is based on the findings of previous literature, see below.

Puelz and Snow consider an ordered logit formulation for the choice of deductible in which the observed number of accidents is introduced among the explanatory variables.\(^{18}\) Then they estimate two structural equations: the demand function for a deductible and a premium function that relates different pricing variables to the observed premium. By doing this they find a significant correlation between coverage and occurrence of accidents. However, it is not sufficient to use their linear modelling procedure, to test for the accident-coverage correlation we are interested in, due to the possible non-linear in risk and especially preferences.\(^{19}\) Instead, a pair of probits and the bivariate probit model are recommended by Chiappori and Salanie.\(^{20}\)

To find the exogeneous variables, the covariates, to be used when pricing the insurance coverage we perform the same pricing procedure as in the pricing department of the insurer:

We observe the set of exogeneous variables $X_{ij}$ for individual (policyholder) $i = 1, \ldots, I$ and calendar year $j \in (1, \ldots, J)$, see section 4.3 for more details about the data. We model the expected claim frequency as a generalized linear model (GLM); Poisson distribution and log link, why the expected number of claims $\nu_{ij}$ can be written as:

$$
\nu_{ij} = \omega_{ij} \exp(\mathbf{X}_{ij}\beta),
$$

where $\omega_{ij}$ is the exposure (the duration) and $\beta$ is a vector of parameters.

\(^{18}\)Puelz and Snow (1994)

\(^{19}\)Dionne et al. (2001) and Chiappori and Salanie (2000)

\(^{20}\)Chiappori and Salanie (2000)
A Bivariate Probit

We define two 0-1 endogeneous variables $Y_{ij}$ and $Z_{ij}$: $y_{ij} = 1$ if $i$ had at least one claim in which they were judged to be at fault in year $j$; otherwise (no claim or $i$ not at fault) $y_{ij} = 0$; $z_{ij} = 1$ if $i$ chose comprehensive coverage, i.e. with a low deductible, in year $j$; $z_{ij} = 0$ if $i$ chose less coverage, i.e. with high deductible. There are many ways to define comprehensive coverage. Chiappori and Salanie investigate the affects of buying minimum legal coverage (third party liability) and any additional coverage, or not. As they admit it is better to treat these coverages separately, since the choice of purchasing additional coverage is also dependent on the actual need of such coverage, but that this will greatly complicate their model. In this paper we address this challenge and will look at only one coverage (the material damages coverage) and focus on the level of deductible within that coverage. However, there are many levels of deductible available on the Danish insurance market, see figure 4.1, so we will test different levels of what should be called low and high deductible, respectively. Moreover we separate claims in which the insured is at fault from those in which she is not. The reason for this is that the information on her risk type may not be conveyed, if another driver is to blame. Furthermore we do not use the further information associated to drivers who had several claims during one year; these are very few; only 0.082% of all policyholders had more than one claim in 2008.

We now set up two probit models, one for the occurrence of an accident (or more precisely claim) and one for the choice of coverage. Let $\epsilon_{ij}$ and $\eta_{ij}$ be two independent centered normal errors with unit variance. Then

$$Y_{ij} = 1 \left( X_{ij}\beta + \epsilon_{ij} > 0 \right)$$

and

$$Z_{ij} = 1 \left( X_{ij}\gamma + \eta_{ij} > 0 \right)$$

We estimate these two probits independently, weighing each individual by the duration $\omega_{ij}$ under insurance, given in part of a year. Now we can easily compute the generalized residuals and apply the test in Chiappori and Salanie. But, as they also conclude, estimating the two probits independently is appropriate under conditional independence,
but it is inefficient under the alternative. Therefore we will instead estimate a bivariate probit in which $\epsilon_{ij}$ and $\eta_{ij}$ still are distributed $N(0, 1)$ but have a correlation coefficient $\rho$, which we also estimate. This will allow us to test $\rho = 0$ but also to get a confidence interval for $\rho$.

Moral Hazard or Adverse Selection

Many tests to distinguish whether asymmetric information depends on moral hazard or adverse selection have been proposed in the previous literature.\textsuperscript{23} These are often specific for the data or structures of the market under investigation, and are not applicable for the data set available in this paper. Instead we will use the hypothesis that a policyholder with an experience-rated insurance policy will have an incentive to be more cautious after experiencing a claim in one year and will therefore have less number of claims in the following year, under the assumption of moral hazard. Hence there will be a negative correlation between claims in one year and the following year, conditional on some explanatory variables, if moral hazard is present. We continue to make use of the bivariate probit model, but this time between claim or not in year 2007, $Y_{i2007}$, and claim or not in year 2008, $Y_{i2008}$, conditional on the same explanatory variables as before, for each year, respectively. If the correlation instead is positive this is called claims migration and is interesting also for fixed priced products, even though the hypothesis on moral hazard does not hold for these types of insurance products.

One can argue that completely new policyholders, in the sense of them lacking experience of driving and of the insurance market, and the insurer are equally uninformed regarding the policyholder’s risk. This is called symmetric incomplete information.\textsuperscript{24} In such case it would be inefficient to draw conclusions based on the ex ante selection of contracts from the offered menu of contracts. But, as time goes by both the insurer and the policyholder gain more experience, which is referred to as learning, though not necessarily equally which can lead to ex post adverse selection. Furthermore, if the policyholder then changes insurer we can definitely assume that the asymmetry truly exists. Cohen finds evidence of such asymmetric information for new policyholders

\textsuperscript{23}Abbring et al. (2003) and Chiappori and Salanić (2000) are specific to the French insurance system, since they make use of the bonus-malus coefficient of the French experience-rating scheme and the rules for how this can be transferred to the children of policyholders with maximum price reduction. See also Cohen and Einav (2003) for other possible tests.

\textsuperscript{24}De Garidel-Thoron (2000)
with three or more years of driving experience, which has been unobservable for the insurer.\textsuperscript{25} This is not possible to investigate in this paper since the number of years with a driving license is not included in the data set. But, if moral hazard is not present and a claim is followed by an increased deductible, ceteris paribus, this indicates that the policyholder has become more aware of her own risk, by learning. To test this we use the bivariate probit model on claims or not in year 2007, $Y_{i2007}$, and positive, zero or negative change of deductible between 2007 and 2008, $\Delta Z_{i}$ conditional on the same explanatory variables as before. Unfortunately, this test will be somewhat weak since a driver also may learn from unreported "near misses", i.e. when an accident almost happens but fortunately does not occur.

### 4.3 Empirical issues

#### The Data

The data set used in this paper is received from a Danish insurance company and covers all data available for the insurer on personal lines automobile insurance. The data set is specific for this insurer, but the information included is general for most insurers why the following results are considered general, at least for the Danish automobile insurance market. Due to this generality of information and the competitiveness of the market, where the pricing scheme is essential, some variables and estimates are unreported in table 4.2 even though they are included in the analysis. The data consists of two sets; the first contains information about 182,031 policyholders who held an active insurance policy during the years 2002-2006. This data set is used for finding the variables to specify the model. The second data set contains information about 125,437 policyholders who held an active insurance in 2008 and will be used to test the risk-coverage correlation.

Policyholders with registered payment default are dealt with special rules, why they are omitted in this study. To perform the tests of moral hazard and learning we create a third data set, which is a sub set of the second one, with the policyholders who held an active insurance policy in both 2007 and 2008.

\textsuperscript{25}Cohen (2005)
Table 4.1: The table shows the number of observations of new or repeated policyholders with a fixed priced (Elite) insurance or not (Step), in year 2008. For each time any information of a policyholder changes, a new observation is produced.

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>Step</th>
<th>Elite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeated</td>
<td>66316</td>
<td>132938</td>
</tr>
<tr>
<td>New</td>
<td>4328</td>
<td>14887</td>
</tr>
</tbody>
</table>

For each individual the insurer has documented around 80 covariates. These are for instance policyholder’s demographic characteristics such as age and gender for both the policyholder and the principal driver, residential area; characteristics of the insured car: brand, size of engine, fuel type, model year; period covered: the length (duration) of the period covered by the purchased policy and so on. Since we only investigate the material damages coverage, the claim frequency will be lower than if we looked at the entire automobile insurance.

**Deductible-premium menu**

The insurer, from which the data originates, offers basically two car products for personal lines (cars for personal usage at a maximum weight of 3.5 tonnes): one fixed price product called Elite and one based on experience rating, with a finite number of steps, therefore called Step. In Denmark it is mandatory for all car owners to have a third part liability insurance. The material damages coverage, which we will investigate, is voluntary and covers damages on the vehicle which originate from bodily injury, collateral damages, theft, robbery, vandalism, fire, explosion, lightning and thrown objects. The criteria for purchase of the Elite product are: the user has to be at least 25 years of age, and has not reported any claims on the third part liability coverage during the last five years and has not reported any claims during the last three years on the material damages coverage, though three glass damage claims are accepted. These criteria are specific for this insurer, but similar products are offered by most Danish insurers.

Almost anyone can purchase the experience-rated product. The level to start on depends on age and number of years without claims. For Step 13 (with lowest premium) the policyholder must fulfill the following criteria: at least 8 years of claim-free driving, of which the last three must be totally claim-free, though glass claims are accepted. There are no legal regulation of the rating scheme on the Danish insurance market, as
### Table 4.2: The table shows some of the parameter estimates of a GLM estimation on the expected claim frequency. The variables are the following: *Elite* stands for the fixed priced product, $d_{1990}$ is the level of deductible given thousands of DKK discounted to 1990, insurance years is the seniority of the policyholder and household density refers to how many households there are within an area of 100 square kilometers. Actually there are 13 variables included in the analysis, but some of these remain unreported, on demand by the insurer, due to competitive reasons. The estimates are based on data from the period of year 2002-2006. ***, **, * indicate significance at 1%, 5% and 10% level, respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>Chi-square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elite Yes</td>
<td>-1.83</td>
<td>0.242</td>
<td>57.23</td>
<td>&gt;.0001</td>
</tr>
<tr>
<td>Elite No</td>
<td>-2.24</td>
<td>0.239</td>
<td>87.8</td>
<td>&gt;.0001</td>
</tr>
<tr>
<td>$d_{1990}$</td>
<td>-0.0751</td>
<td>0.0061</td>
<td>152</td>
<td>&gt;.0001</td>
</tr>
<tr>
<td>age of driver</td>
<td>-0.0589</td>
<td>0.0080</td>
<td>54.4</td>
<td>&gt;.0001</td>
</tr>
<tr>
<td>insurance years</td>
<td>-0.0029</td>
<td>0.0008</td>
<td>13.4</td>
<td>0.0003</td>
</tr>
<tr>
<td>age of car</td>
<td>-0.0336</td>
<td>0.0026</td>
<td>167</td>
<td>&gt;.0001</td>
</tr>
<tr>
<td>household density</td>
<td>0.0350</td>
<td>0.0028</td>
<td>159</td>
<td>&gt;.0001</td>
</tr>
</tbody>
</table>

In order to specify the variables in our bivariate probit model, we perform a GLM estimation of the expected claim frequency and the results, for some of the variables, are shown in table 4.2. In addition to the variables in table 4.2, we also find the gender of the principal driver significant. If we set a dummy variable equal to 1 for female drivers and zero for male drivers, the parameter estimate of this variable is 0.0994 with a significance level of less than 0.0001. However, even though this variable is significant, it is not used in the pricing scheme of the material damages coverage due to interpretation difficulties, and was therefore excluded also in the forthcoming analysis.

Figure 4.1 shows the distribution of the chosen deductibles, by the policyholders in 2008, discounted to 1990 years value. In 1990 the levels of deductible were set at 500 DKK, 1 000 DKK, 1 500 DKK..., 10 000 DKK. Each year the deductibles have been regulated so that in 2008, the levels were 919DKK, ..., 18 970DKK, and chosen to an average of 4 176DKK. The average premium of a motor insurance is 4 901 DKK. Note

---

1. The regulations of the French automobile insurance market are described in Chiappori and Salanié (2000) and Richaud (1999), and the Canadian ones in Dionne et al. (2001).
2. DKK is the abbreviation for the Danish currency kroner.
Figure 4.1: Distribution of deductibles, in monetary units.
that there are many levels of deductibles in Denmark, compared to markets described in the previous literature.

In table 4.2 we see that the variable \( d_{i1990} \) has a corresponding parameter estimate of \(-0.0751\). \( d_{i1990} \) is the chosen deductible discounted to the corresponding deductible level of year 1990, given in thousands of Danish kroner (DKK). This means that policyholders with higher deductibles report fewer claims. Naturally, policyholders with lower deductibles are able to file claims for accidents with damage costs too small to claim under policies with a higher deductible. In insurance data we can only observe claims, not accidents. The decision to report a claim is, most often, made by the policyholder, and should be considered as a response to or outcome of a choice based on incentives. If the expected costs of reporting a claim exceed the expected compensation, due to, for instance, experience rating schemes resulting in a higher future premium or that the expected costs are below the deductible, then the policyholder is always free not to file the accident. Hence, if we count all claims reported by low-deductible policyholders we will expect more claims submitted by the low-deductible policyholders, even if the two groups do not at all differ in their risk type.

One way to deal with the bias, originating from different levels of deductibles, presented in Chiappori and Salanié, is to omit all accidents where only one vehicle is involved.\(^{28}\) An even more restrictive way is to consider only accidents involving bodily injuries, since it is mandatory to report these kind of accidents. However, this requires a lot more detailed level of documentation and processing of data and reduces the number of accidents radically. For the purpose of this paper we choose to follow Cohen in which all claims below a certain level are removed.\(^{29}\) This gives a focus on larger claims but at least all policyholders are fairly treated, and the spurious correlation can be dealt with, though it comes at the cost of lower claims frequency. Hence, throughout we only count claims that can be submitted by both groups of policyholders, i.e. only claims with a severity that exceed the highest level of deductible; in 2008 this was 18970 DKK. After this truncation the parameter estimate of \( d_{i1990} \) is 0.0357, but it is no longer significant at a reasonable level (the p-value is equal to 0.1249). Hence we have removed or at least reduced the bias. One of the more applied problems, and which is of most importance for the analysis and results, is where to draw the line for what is considered to be a low

\(^{28}\)Chiappori and Salanié (2000)  
\(^{29}\)Cohen (2005)
4.4 Results

<table>
<thead>
<tr>
<th>$T$</th>
<th>500</th>
<th>1500</th>
<th>2500</th>
<th>3500</th>
<th>4500</th>
<th>5500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta(Z_i)$</td>
<td>0.235***</td>
<td>0.182***</td>
<td>0.0401*</td>
<td>0.0215</td>
<td>0.0003</td>
<td>0.1663**</td>
</tr>
</tbody>
</table>

Table 4.3: The table shows the GLM estimate $\beta$ of the dummy variable for low and high deductible $Z_i$, when the level of deductible $d_{i1990}$ has been replaced by $Z_i$. Otherwise are all variables the same as in table 4.2. The estimates are based on data from the period of year 2002-2006. ***, **, * indicate significance at 1%, 5% and 10% level, respectively.

or high deductible. This is a delicate task due to the many possible levels of deductible and the distribution of the chosen ones in this set of data.

To get a first sign of the correlation we will perform a new parameter estimation on the GLM, where $d_{i1990}$ has been replaced by the new variable $Z_{ij} = 1(d_{i1990} \leq T)$. $T$ is a threshold, to define the 0-1 deductible variable $Z_{ij}$ for each individual, i.e. what is to be considered as a low or high deductible. We know that this is not sufficient to base conclusions on, but it will give us a hint of if it is worth proceeding. We also introduce a new variable $Z_{i}^{rel} = 1\left(\frac{d_{i1990}}{\text{premium}} \leq R\right)$, which stands for relatively low deductible, given the premium. $R$ is a new threshold.

4.4 Results

To ease the interpretation of the results and reduce the risk of spurious effects due to unforeseen expectations on future changes in premium, we start by focusing on Elite drivers. After reducing the number of claims, only counting claims that could be submitted by both types of policyholders, we start comparing the claims variable $Y_{ij}$ for policyholders with a low deductible with the claims submitted by those with a high deductible. The GLM parameter estimates for different value of $T$ are displayed in table 4.3. We see that it is possible to find a positive significant parameter, indicating that the claims are rarer for Elite policyholder choosing a high deductible than for Elite policyholder choosing a low one.
Evidence of Asymmetric Information

<table>
<thead>
<tr>
<th>Coverage-Risk Correlation (Elite)</th>
<th>T</th>
<th>500</th>
<th>1500</th>
<th>2500</th>
<th>3500</th>
<th>4500</th>
<th>5500</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td></td>
<td>0.00272</td>
<td>0.000343</td>
<td>-0.000117</td>
<td>-0.00488</td>
<td>-0.00563</td>
<td>0.00338</td>
</tr>
<tr>
<td>( \bar{Z}_i )</td>
<td></td>
<td>0.00241</td>
<td>0.0541</td>
<td>0.613</td>
<td>0.945</td>
<td>0.949</td>
<td>0.989</td>
</tr>
</tbody>
</table>

Table 4.4: The table shows the estimated correlation between the deductible variable \( Z_i \) and the claim variable \( Y_i \) using a bivariate probit model for policyholders with a fixed priced insurance, conditional on the variables in table 4.2 excluding \( d_{i1990} \). \( \bar{Z}_i \) is the mean of \( Z_i \) showing the fraction of policyholders choosing low deductible. No correlation estimates are significant. The estimation is based on data from the period of year 2008. The number of observations are 143919. ***, **, * indicate significance at 1%, 5% and 10% level, respectively.

<table>
<thead>
<tr>
<th>Coverage-Risk Correlation (new Elite)</th>
<th>T</th>
<th>500</th>
<th>1500</th>
<th>2500</th>
<th>3500</th>
<th>4500</th>
<th>5500</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td></td>
<td>0.0230</td>
<td>0.00361</td>
<td>0.0187</td>
<td>-0.0206</td>
<td>-0.0312**</td>
<td>-0.0515***</td>
</tr>
<tr>
<td>( \bar{Z}_i )</td>
<td></td>
<td>0.00334</td>
<td>0.189</td>
<td>0.625</td>
<td>0.973</td>
<td>0.980</td>
<td>0.996</td>
</tr>
</tbody>
</table>

Table 4.5: The table shows the estimated correlation between the deductible variable \( Z_i \) and the claims variable \( Y_i \), using a bivariate probit model, for new policyholders with a fixed-priced insurance, conditional on the same variables as in table 4.2, excluding \( d_{i1990} \). \( \bar{Z}_i \) is the mean of \( Z_i \) showing the fraction of policyholders choosing a low deductible. A significant correlation is found when \( T \) becomes large. The estimates are based on data from the period of year 2008. The number of observations are 14550. ***, **, * indicate significance at 1%, 5% and 10% level, respectively.

Deductible in monetary units

The second test we perform is the bivariate probit used in Chiappori and Salanie and Cohen. The bivariate probit estimates the correlation \( \rho \) between the error terms of two binary equations, see in section 4.2 above.

Table 4.4 displayes the estimated risk-coverage correlation for different value of the threshold \( T \). We see that the estimate of the correlation some times is negative, in contrast to what the standard theory of asymmetric information models predicts. However, none of the estimates are significant on a reasonable level and might be results of spurious effects.

The results for all new Elite policyholders are provided in table 4.5, where we find a significant negative correlation, indicating that policyholders with a low deductible (high coverage) also have fewer claims. Based on these results we cannot reject the

---

4.4 Results

<table>
<thead>
<tr>
<th>$R$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>-0.00260</td>
<td>0.00442</td>
<td>0.00637*</td>
<td>0.0108***</td>
<td>0.0115***</td>
</tr>
<tr>
<td>$\bar{Z}_i^{rel}$</td>
<td>0.190</td>
<td>0.837</td>
<td>0.954</td>
<td>0.972</td>
<td>0.977</td>
</tr>
</tbody>
</table>

Table 4.6: The table shows the estimated correlation between the deductible variable $Z_i^{rel}$ and the claim variable $Y_i$ using a bivariate probit model for policyholders with a fixed priced insurance, conditional on the variables in table 4.2 excluding $d_{i1990}$. $\bar{Z}_i^{rel}$ is the mean of $Z_i^{rel}$ showing the fraction of policyholders choosing a relatively low deductible. Significant correlation estimates are found for high values of $R$. The estimation is based on data from the period of year 2008. The number of observations are 143919. ***, **, * indicate significance at 1%, 5% and 10% level, respectively.

hypothesis of independent equations, or perhaps, as said above, reject the independence in benefit of the negative correlation. However we note that the significant negative correlation is found when the number of low risks (high deductible) is quite small. This might be explained by that some young drivers are forced to choose a high deductible in order to buy an insurance at all.

Instead of ending our analysis with just examining deductibles in monetary units, as in Chiappori and Salanie, we continue with investigating the distribution of deductibles divided by the paid premium, as in Cohen.\textsuperscript{31}

**Deductible in relation to premium**

Table 4.6 and 4.7 provide the estimates of the risk-coverage correlation for Elite and new Elite policyholders, respectively. We find positive correlations at a statistically significant level of 1%. Thus the hypothesis that the equations are independent can be rejected. If we look at the distribution of the deductible in relation to the paid premium (see figure 4.2) the distribution is smooth except for the case where the deductible is approximately four times the paid premium.

The new Elite policyholders with a high deductible in relation to their premium are younger (average of 43 years) than the ones with low deductible (average of 52 years). They have a lower yearly premium, on average 2 300 DKK compared to 5 000 DKK, and a higher deductible, 8 000 DKK compared to 5 000 DKK.

\textsuperscript{31}Chiappori and Salanié (2000) and Cohen (2005)
Figure 4.2: Distribution of deductibles in relation to paid premium

Table 4.7: The table shows the estimated correlation between the deductible variable $Z_r^{rel}$ and the claims variable $Y_i$, using a bivariate probit model, for new policyholders with a fixed-priced insurance, conditional on the same variables as in table 4.2, excluding $d_{i1990}$. $Z_r^{rel}$ is the mean of $Z_r^{rel}$ showing the fraction of policyholders choosing a relatively low deductible. A significant correlation is found when $R$ becomes large. The estimates are based on data from the period of year 2008. The number of observations are 14550. ***, **, * indicate significance at 1%, 5% and 10% level, respectively.
4.4 Results

Coverage-Risk Correlation (Step drivers)

<table>
<thead>
<tr>
<th></th>
<th>$T = 4$</th>
<th>$R = 4.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New</td>
<td>Repeated</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.0120</td>
<td>0.00289</td>
</tr>
</tbody>
</table>

Table 4.8: The table shows some examples of estimates of the correlation between the deductible variables $Z_i$ and $Z_{i}^{rel}$ and the claim variable $Y_i$, respectively, using a bivariate probit model for policyholders with a fixed priced insurance, conditional on the variables in table 4.2 excluding $d_{i1990}$. No correlation estimates are significant. The estimation is based on data from the period of year 2008. The number of observations are 3861 new Step drivers and 53200 repeated Step drivers. ***, **, * indicate significance at 1%, 5% and 10% level, respectively.

Claim Occurrence Dependency

<table>
<thead>
<tr>
<th></th>
<th>Elite</th>
<th>New Elite</th>
<th>Step</th>
<th>New Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.0504***</td>
<td>0.0356***</td>
<td>0.0448***</td>
<td>0.0610***</td>
</tr>
<tr>
<td>Number of policyholders</td>
<td>57063</td>
<td>8774</td>
<td>24139</td>
<td>2421</td>
</tr>
</tbody>
</table>

Table 4.9: The table shows the correlation estimates between $Y_{i2007}$ and $Y_{i2008}$ for all Elite, new Elite, all Step and new Step drivers, using a bivariate probit model. ***, **, * indicate significance at 1%, 5% and 10% level, respectively.

The omission of the gender variable does not explain the asymmetry, since the correlation is still significant when the gender variable is added to the model.

We do not find any significant correlation at all for the policyholders with an experience-rated insurance product. Table 4.8 shows the results for some levels of deductible, both in monetary units and in relation to the premium.

Moral hazard or learning

When we estimate our conditional bivariate probit model on $Y_{i2007}$ and $Y_{i2008}$, on our data set of all policyholders with an active automobile insurance policy in 2007 and 2008, without missing values in the conditional variables, we get the results in table 4.9.

The results show that we do not find the negative claim occurrence dependency, i.e. a claim is followed by fewer claims, as the theory of moral hazard predicts\(^{32}\), for any of

\(^{32}\)Abbring et al. (2003)
50 Evidence of Asymmetric Information

<table>
<thead>
<tr>
<th>Dependency of Claims and Change of coverage</th>
<th>$\rho$</th>
<th>$\Delta Z_i = 1$</th>
<th>New Elite</th>
<th>Step</th>
<th>New Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta Z_i = 1$</td>
<td>-0.00431</td>
<td>-0.00157</td>
<td>0.0249***</td>
<td>0.03617</td>
<td></td>
</tr>
<tr>
<td>$\Delta Z_i = -1$</td>
<td>0.00277</td>
<td>0.0149</td>
<td>-0.00567</td>
<td>0.0163</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.10: The table shows the correlation $\rho$ between claims in 2007 and change in coverage, i.e. level of deductible, in year 2008 for all Elite, new Elite, all Step and new Step drivers. $\Delta Z_i = 1$ investigates the increase of deductible or not, and $\Delta Z = -1$ investigates the decrease of deductible or not. ***, **, * indicate significance at 1%, 5% and 10% level, respectively.

In table 4.10 we see the estimated correlation between claims in year 2007 and the possible change of deductible. $\Delta Z_i = 1$ if policyholder $i$ increased the level of deductible between 2007 and 2008, then $\Delta Z_i = 0$ otherwise. We also test $\Delta Z_i = -1$ if policyholder $i$ decreased the level of deductible between 2007 and 2008, $\Delta Z_i = 0$ otherwise.

4.5 Discussion

As discussed in Chiappori and Salanie it is generally a delicate task to model the claim frequency, since omission of some variables can generate spurious correlations.\textsuperscript{33} In this paper we use the same covariates and models as the insurer uses for the pricing scheme of this particular automobile insurance coverage. Hence, we ought not to experience the same problems of spurious effects and therefore our results must be considered rather reliable. To a start, when we divide coverage into low and high based on monetary units of deductible, we generally do not find evidence of asymmetric information, not for repeated Elite drivers, and neither for repeated or new policyholders with experience-rated products. However, in table 4.5 with results from new Elite drivers, we are surprised to get a negative correlation for really high levels of $T$. This means that policyholders with a high deductible have more claims than the ones with a low deductible, in contrast to what we expect from the theory of adverse selection and moral hazard models. This could be explained by so called advantageous selection models, that assume that lower risk policyholder are both more cautious and risk averse.\textsuperscript{34} Fortunately this was explained by the pricing department of the insurer from which the

\textsuperscript{33}Chiappori and Salanié (2000)
\textsuperscript{34}de Meza and Webb (2001)
data originates; some young drivers are forced to have a high level of deductible in order to get an insurance at all, since they for different reasons are expected to have more claims.

When we state the deductible in relation to the paid premium, as in table 4.6 and 4.7, we find a strong positive risk-coverage correlation for both new and repeated Elite drivers, but not for the policyholders with an experience-rated product. One possible explanation for the lack of significance is that the number of observations of the new Step drivers (3861) are too few. But the number of observations of repeated Step drivers (53200) must be considered sufficient.

Another possible explanation is that we have not taken expectations on increase of future premium into consideration when truncating the claims. Predergast concludes that Bonus-malus systems as when experience-rating is used, and/or lack of information pooling within insurers, as on the Danish insurance market, constitute an implicit deductible.\textsuperscript{35} Hence, they can provide incentives for both self-protection and self-insurance, without explicit paragraphs.\textsuperscript{36}

Anyhow we see that the experience-rated pricing of the Step product is sufficiently good to remove the effects of any possible asymmetric information. These findings are coherent with Chiappori and Salanié, who do not find the risk-coverage correlation for an experience-rated automobile insurance product in French data.\textsuperscript{37} Our findings of a positive risk-coverage correlation for the fixed-priced product is also coherent with Cohen, who finds this correlation for drivers with three or more years or driving experience, with a fixed priced product.\textsuperscript{38} Even though we lack information of the number of years with driving license, in this paper, it is implicit since there is a criterion to be claims free for at least three years, to be able to purchase the fixed-price product. Hence the Elite drivers must have had their driving license for three or more years.

Since a positive risk-coverage correlation is a central prediction not only of adverse-selection models but also of moral hazard models, we introduced a test, fitted to the data available, to separate the two special cases. The results shows no evidence of moral hazard for any group of policyholders. Instead we found a positive claims occurrence correlation indicating that policyholders are unable to, in sufficient extent, change their risk behaviour, at least not from one year to the next. One can say that

\textsuperscript{35}Prendergast (1992)
\textsuperscript{36}Ehrlich and Becker (1972)
\textsuperscript{37}Chiappori and Salanié (2000)
\textsuperscript{38}Cohen (2005)
poor drivers remain so, for at least a while. This strengthens the arguments for applying experience-rated pricing.\textsuperscript{39} A positive claims occurrence dependency can also be an effect of unobserved heterogeneity, due to too few covariates used, according to Abbring, Chiappori, Heckman and Pinquet.\textsuperscript{40} However, Saito does not find that unobserved variables induce adverse selection.\textsuperscript{41} Hence, it is unclear to what extent the pricing scheme can be improved in order to deal with this.

When we investigate learning we find no significant correlation, except for $Step$ drivers when $\Delta Z_i = 1$, i.e. for those who increased their deductible between 2007 and 2008, and $\Delta Z_i = 0$ otherwise. This can be a way of reducing the renewal premium to compensate for the increase in premium, due to the claim. It also indicates that they are confident that the claim was coincidental and that they do not expect any claims in the following insurance period. One possible explanation is that repeated policyholders and the new $Elite$ drivers are already aware of their level of risk, and have purchased the insurance at a certain level of deductible based on the risk assessment of this. The $Step$ drivers take action on the occurrence of claims and change their deductible, while the $Elite$ drivers do not have to. The new $Step$ drivers might be more inexperienced and are less able to make a conclusion based on their claims experience, or perhaps they agree with the increased premium.

As said before, we note that the deductibles are generally of the same size as the premium, which is quite different from the Israeli market described in Cohen, where the regular deductible is set at approximately 50\% of the regular premium, high deductible is considered to be 90\% of the regular premium and very high 130\% of the regular premium.\textsuperscript{42} This means that almost all chosen deductibles of the Danish policyholders are to be considered high or very high according to Cohens gradation. Given this distinction, it is important for the insurer to keep an eye on the deductible levels; the deductible is often seen as a simple and efficient way to avoid small claims, since the processing claims are related to fixed costs and these are relatively large for small claims. However, if the levels of deductible causes adverse selection, this has to be taken into consideration. On the other hand, the insurer’s profit might be higher for insurance

\textsuperscript{39}For a survey in experience rating and credibility theory see Bühlmann and Gisler (2005). More recent developments within using claim migration are found in Englund et al. (2008) and Englund et al. (2009).
\textsuperscript{40}Abbring et al. (2003)
\textsuperscript{41}Saito (2006b)
\textsuperscript{42}Cohen (2005)
with high coverage, contrary to the prediction of competitive markets, and then this might compensate for the effects of adverse selection.

Rothschild and Stiglitz write that neither insurance firms nor their customers have to be perfectly informed about the differences in risk properties that exist among individuals, in order for adverse selection to exist; obviously the *Elite* drivers with high deductible are sufficiently aware of their risk.\(^{43}\)

Furthermore, we see that when the significance risk-coverage correlation is found our population is split so that most policyholders have a low deductible. This is in contrast to what Cohen finds on the Israeli market, but is coherent with the findings of Rothschild and Stiglitz saying that under adverse selection an equilibrium exists only if there are enough high risks (choosing low deductible).\(^{44}\) So, put in simple words, Cohen finds the few poor drivers, while we find the few good ones, which can be used for e.g. selecting which policyholders who might get a VIP treatment, or be offered additional products as in Englund, Nielsen and Tanggaard.\(^{45}\)

### 4.6 Conclusions

In this paper we find evidence of asymmetric information. We do not find evidence of moral hazard. The way the thresholds, for which the risk-coverage correlation is significant, separates the policyholders is coherent with the Rothschild-Stiglitz model. This model assumes adverse selection without moral hazard. Therefore the significant risk-coverage correlation in our investigation is assigned adverse selection. We find evidence of learning, but only for repeated policyholders with an experienced-rated product, for which adverse selection is not present, so it might be more of a reaction on an experienced claim to lower the total premium.

All this means that the *Elite* policyholders are sufficiently aware of their intrinsic risk and choose the amount of coverage, level of deductible, based on this. However, they are not able to change their risk behaviour sufficiently enough to influence the outcome in a significant way. The policyholders with the experience-rated product, are either unaware of their own risk, or risk averse. Or, the pricing scheme of this product

\(^{43}\)Rothschild and Stiglitz (1976)

\(^{44}\)Cohen (2005) and Rothschild and Stiglitz (1976)

\(^{45}\)Englund et al. (2009)
is simply sufficiently good to remove the effects of any possible asymmetric information.

We find a positive claim occurrence dependency, so called claim migration, which indicates that poor drivers remain bad for at least a while, i.e. policyholders are generally unable to change their risk behaviour, nor do they learn from their claims, at least not from one year to the next.
Asymmetric Information, Self-selection and Pricing of Insurance Contracts:
The Simple No-Claims Case

Abstract: This paper presents an optional bonus-malus contract based on \textit{a priori} risk classification of the underlying insurance contract. By inducing self-selection the purchase of the bonus-malus contract can be used as a screening device. This gives an even better pricing performance than both an experience rating scheme and a classic no-claims bonus system. We find additional improvements of 80\%, or more, compared to going from the flat rate to a covariate-based regression estimator. An application to the Danish automobile insurance market is considered.
Asymmetric Information and Self-selection

If one tells the truth, one is sure, sooner or later, to be found out.

Oscar Wilde

5.1 Introduction

In a world with two levels of risk and asymmetric information, where the insurance company cannot distinguish between customers from the two risk groups, there is no pooling equilibrium and there may not be an equilibrium at all.\(^1\) In a pooling equilibrium the lower risk policyholders subsidize the higher risk policyholders, and if the policyholders are aware of their risks and difference is too large between risk and price, the lower risks will not buy insurance. Thus, either the market for insurance breaks down or each type of policyholder buys a tailor-made insurance contract with payoff that caters to the specific riskiness of the policyholder. This idea of sufficiently tailor-made (differentiated) contracts may not fit exactly with what we see in current insurance markets. Insurance contracts offered by insurance companies are only to some extent differentiated, depending on the sophistication of the models, the available data and regulations.\(^2\) Few people would be happier than the actuaries\(^3\), if the self-evident truth "all men are created equal" also was true in the sense of their risks. Then it would be correct to apply the most simple risk estimator: the flat rate. However, in the sense of risk, there exist some properties which make us all unique as individuals, and which make the flat rate uncompetitive. This heterogeneity has always been a challenge when pricing individual insurances and therefore the actuaries have designed rating schemes (tariff structures) that tries to distribute the burden of claims fairly between the policyholders. This classification is based on \textit{a priori} variables. Such variables could be age, sex, residential area, marital status or other relevant variables.\(^4\) The most widely used models for covariate-based regression are generalized linear models (GLM).\(^5\) One way to improve the \textit{a priori} pricing is to make use of \textit{a posteriori} corrections based on

\(^1\)Rothschild and Stiglitz (1976)
\(^2\)In Allard et al. (1997) pooling equilibrium may exit if even the slightest distributional cost exists. Differences in risk aversion can also make policyholders with different risks accept pooling (at least to some extent) according to Rothschild and Stiglitz (1976).
\(^3\)At least their work on pricing models and estimators would be simplified to a great extent.
\(^4\)More on data in section A.1.
experience rating, such as credibility theory and bonus-malus systems (BMS). However, these methods have a limited impact when individual claim information is rare, as for new policyholders, and may take several years of observation before the precision is reasonable. Therefore we seek a new way to differentiate the policyholders by risk, already at the time of purchase of the insurance product.

This paper gives examples of an insurance contract, which may induce self-selection within the existing risk classes of the rating scheme. The new type of insurance contract has some features that are not found in contracts of standard BMS. Similarly to BMS the payoff that is offered to policyholders depends on the actual number of experienced claims during each insurance period, but the contract has a more general form of payoff than a standard contract with bonus. The payoff can be tailor-made to adopt individual policyholder’s needs or preferences. Implicit in this is the idea that insurance companies should offer a menu of contracts in such a way that self-selection is induced among the policyholders. With such a contract the insurer will be able to differentiate the net insurance premium for each period of time even more, thereby making it more competitive due to the separating effect. This is desirable since adverse selection can cause inefficiency in insurance contracts, which reduces the benefits of taking an insurance for the lower classes of risk, since the price will be too high in relation to the risk.

We use policyholder and claims data from a Danish insurance company to investigate the hypothesis that such a contract will in fact induce self-selecting behavior, in addition to the decision of buying an insurance, or not, at the differentiated prices of the rating scheme. We do this by examining how close we get in setting prices that are actuarially fair. We measure the deviations from fair valuation using the squared-error criterion: the lower sum of squared errors (SSE), the closer we are to an actuarially fair valuation. And, if we can show that a menu of the standard and the new contract, offered to the policyholders, gives a lower SSE (is better) than other existing contracts and pricing methods, then we claim that our contract actually succeeds in inducing self-selection.

The paper is arranged as follows. Section 5.2 put the present paper in its context,
gives a background to our pricing problem and serves as a small survey of previous literature on the subject. Section 5.3 describes a heterogeneity model and the new insurance contract is defined. In section 5.4 we compare the actuarial pricing ability of the new contract with other existing insurance contracts under equal conditions, i.e. without individual choice. Due to the findings of this study we continue to section 5.5 with a study, where we allow individual choice, based on an expected utility representation and private information. This section is followed by a discussion in section 5.6. Finally, section 5.7 concludes. In the Appendix, section A.1 describes the data process and section A.2 presents the credibility model and estimators.

5.2 Motivation

"I think my insurance premium is too high," said the indignant policyholder. The insurer answered with dry assertion: "Dear client, based on our risk evaluation it is correct". "I think you´re mistaken. I can bet my money on it," the policyholder continued. The insurer smiled, the policyholder just revealed that she was willing to bet on her own risk. She wouldn´t bet on it if she didn´t think she would benefit from it: "OK, if you´re right, we will reward you. But, if you´re wrong, you will reward us". The policyholder was thrilled with the new option to able to believe in and bet on herself: "It´s a deal!"

The reader might recognize the feeling of paying too much, for her insurance, in the conversation above. While the insurers argue that the insurance coverage is always active, due to the risk, i.e. the policyholders use the insurance all the time, the policyholders might argue that the insurance is only active in case of an accident followed by a claim. For policyholders, who have never reported a claim, the feeling of inactivity of the insurance might be extra evident. In this paper we will propose an opportunity for the policyholders to bet on their believes in being a good client; if they are right they get a dividend, if they are wrong they loose the entrance fee, the stake. Unlike many other bets – like football, boxing, horse racing etc. – the policyholder might have private information on, or might even control some of, the influencing factors in this bet.
When one party has access to information that is denied to the other, there is a problem of asymmetric information. This problem is present on all markets. Most of the time the seller does not know the buyer’s preferences for a product or service, nor the maximum price she will be willing to pay to acquire it. Similarly, it is normally not likely that the buyer has much information about the seller’s production technology or marginal costs. On the other hand, this asymmetry is irrelevant, in most cases. In a perfectly competitive market, the seller does not need a detailed knowledge of the buyer’s willingness to pay, because the seller has to charge the competitive price. In the same way the buyer does not benefit from information about the technology, since again all the information the buyer needs is included in the price. Hence, this type of asymmetric information is both extensive and unimportant. However, when the hidden information of one agent affects the payoff of the other party, the asymmetric information is indeed relevant and has important consequences for the existence and efficiency of competitive equilibrium.\footnote{Akerlof (1970) and Rothschild and Stiglitz (1976).} For instance, new cars are offered in different colors at the same price, which reveals that the seller is indifferent about the buyer’s preferences of color. In such situation a market equilibrium will still exist and be efficient as usual, since the buyer’s preferences does not directly affect the seller’s payoff. In the same way different levels of insurance coverage may be proposed to policyholders, reflecting asymmetric information about risk aversion. Differences in risk aversion itself does not affect the seller’s payoff, why the conclusions of Akerlof, Rothschild and Stiglitz do not apply, but the differences in true risk of the policyholders do have impact on the insurer’s payoff in a competitive setting, why the asymmetric information regarding the true risk of the policyholders is relevant.

Under adverse selection, without moral hazard, the degree of coverage chosen by the policyholder is based on the ex-ante assessment a policyholder makes of his riskiness. Each policyholder is fully described by a single number, e.g. his probability of accident, since neither the frequency nor the claim size are functions of the actions of the policyholders. It follows that high risks choose more coverage than low risks. In the presence of moral hazard, the causality works in the opposite direction caused by the unobservability of efforts to prevent accidents, hence the choice of coverage affects the efforts to prevent accidents. The probability of accidents can now be seen as a function of the level of care, which is unobservable by the insurer. Generous coverage reduces the expected cost of an accident for the policyholder and therefore the incentives for
safety. In this case, more coverage may lead to more accidents. In the end, both moral hazard and adverse selection predict a single-period positive correlation between risk and coverage within a risk class.

Empirical analyses to find evidence of asymmetric information have focused on the coverage-risk correlation, described above. While the correlation has been found in studies of some insurance markets, the findings have been rather diverse in the automobile insurance market. Three initial studies suggest the existence of coverage-risk correlation.10 But, their findings were challenged by following research.12 Chiappori and Salanie find no correlation between higher insurance coverage and more accidents in their French data.13 Dionne, Gouriéroux, and Vanasse reach a similar conclusion using a data set in Quebec, and so do Saito on the Japanese automobile insurance market.14 These scattered conclusions might of course be dependent on differences in specific data, definition of claim, statistical tests chosen and market structures. But, there is also empirical evidence that risk-averse policyholders might tend to buy more coverage and yet be more cautious, hence they are less risky.16 The effects of such preference-based selection can counteract, to some extent, the effects of “traditional” adverse selection, and the correlation between coverage and risk may therefore be of indeterminate sign. One could also argue that e.g. new policyholders and the insurer are equally uninformed regarding the policyholder’s risk.17 But, as time goes by both the insurer and the policyholder get wiser, yet not necessarily equally which can lead to ex post adverse selection. Furthermore, if the policyholder then changes insurer the asymmetry truly appears. Cohen finds evidence for such asymmetric information for new policyholders with three or more years of driving experience, which has been unobservable for the insurer.18

As a complement to all this previous work, the present paper focus on a possible solution or application to such an asymmetric information problem, and specifically adverse selection. We are interested in answering the questions: What if there is an...
5.3 The Model

informational asymmetry; how do we make use of it? What are the effects? Can we design a new contract and how does it behave, both as a pricing method and as a screening device? Is it simple enough to implement?

5.3 The Model

Our model consists of two parts. First, a model for insurance claims when policyholders are heterogeneous. Second, a model of the insurance contract that exhibits features useful for inducing self-selection among insurance customers.

Heterogeneous policyholders

Let us consider the number of claims, $N_{ij}$, for individual $i \in [1, ..., I]$ in insurance period $j \in [1, ..., J]$ on one particular coverage (product). The natural choice of insurance period is one year. We consider one coverage at a time and assume that the events in different coverages are independent. We assume that every policyholder has a latent individual risk profile $\Theta_i$, a realization of the random variable $\Theta_i$. The pairs $(\Theta_1, N_{1j}), (\Theta_2, N_{2j}), ..., (\Theta_I, N_{Ij})$ are independent random vectors, where $\Theta_1, \Theta_2, ...$ are iid and the two first moments are $E(\Theta_i) = \theta_0$ and $V(\Theta_i) = \tau^2$.19 The expectation of $N_{ij}$ is taken with respect to the latent variable $\Theta_i$, which is to be considered as a random effect for the insurer but known by the policyholder, and the coefficients $\nu_{ij}$ dependent on regression components (represented by a line-vector $x_{ij}$) and on the duration of the period $\omega_{ij}$:

$$\nu_{ij} = \omega_{ij} \exp(x_{ij}a) ; \ a \in \mathbb{R}^k,$$

where $a$ is a column-vector of parameters and where $k$ is the number of regression components. Hence, the number of claims $N_{ij}$, given $\Theta_i = \theta_i$, is assumed to be Poisson distributed with expectation $\nu_{ij}\theta_i$. We assume that the insurer has some prior knowledge about the individual, i.e. the regression components $x_{ij}$ also called the covariates. Such covariates could be age, sex, geographical area, or other relevant variables.20 When covariates are used to price insurances it is assumed that each individual with the same characteristics have the same risk. However, with an individual risk profile this

---

19 More on the credibility model and estimators in section A.2.
20 More on data in section A.1.
means that policyholders with the exact same set of covariates may differ in risk due to unobservable characteristics. When making decisions regarding the insurance the policyholder will take the conditional probability of claims \( p_{ij}(\theta_i) = \Pr(N_{ij} > 0 \mid \Theta_i = \theta_i) \) into consideration, while the insurer will pay attention to the unconditional probability of claims: \( p_{ij} = E[p_{ij}(\theta_i)] \), due to the assumed informational asymmetry.

The usual assumptions used in models on adverse selection are maintained in this paper: the individual information is costly to observe by the insurer, neither the frequency nor the claim size are functions of the actions of the policyholders, the provision of the insurance is costless and the insurer is risk-neutral while the policyholders are risk-averse, having identical twice-differentiable, increasing and strictly-concave utility functions \( U \).

Let \( Y_{ij} \) be the aggregated claim amount and \( S_{ij} \) be the severity of claims for individual \( i \), in insurance period \( j \). \( S_{ij} \) is assumed to be normally distributed with expectation \( m_{ij} \). If the probability of claims is observable by each policyholder and insurer, hence under symmetric information, the insurance contract is simple; in each period the insurer offers to cover all losses for each individual against an insurance premium \( \mu_{ij}(\theta_i) = E(Y_{ij} \mid \theta_i) \). We assume the frequency and severity of claims to be independent, why \( \mu_{ij}(\theta_i) = \nu_{ij}m_{ij}\theta_i \). Since we will only consider full coverage insurances in what comes next, the severity of claims will not affect the individual decision rules or choices, as can be seen in table 5.4 where the terminal wealth and expected utility for different scenarios are presented. Hence, we set \( S_{ij} = 1 \). In the following section we will propose a way to deal with the situation under asymmetric information, where the insurer cannot observe the latent individual risk profile \( \theta_i \), and thereby neither the conditional probability of claims \( p_{ij}(\theta_i) \).

**Insurance contract**

A standard insurance contract is priced at the actuarially fair value \( \mu_{ij} \). We now consider a more general contract, where, in addition to the premium \( \mu_{ij} \), the policyholder pays an entrance fee \( \pi_{ij} \). If the policyholder does not have any claims during the insurance period the insurer will pay an amount, a dividend, \( D_{ij} = \pi_{ij} + d_{ij} \). Mathematically the insurance contract is defined as \( D_{ij} = (\pi_{ij} + d_{ij}) 1_{\{N_{ij}=0\}} \), where \( 1_{\{N_{ij}=0\}} \) is the indicator for the event \( N_{ij} = 0 \). A standard insurance contract is, of course, defined by \( d_{ij} = \pi_{ij} = 0 \). In this sense the contract can be seen as an add-on to the standard insur-

\(^{21}\)More on the utility function in section 5.5.
5.3 The Model

ance contract. The entrance fee, $\pi_{ij}$, can be set either individually or collectively. This will be a strategic decision of the insurer and something we will discuss in the remaining parts of the paper. The dividend $d_{ij}$ is calculated via the actuarial net premium principle 22: $\pi_{ij} = \mathbb{E}[D_{ij}]$, 23 hence the initial price equals the expected outcome. 24 This contract has a binary nature; either you get the payoff or you do not.

Since the number of claims $N_{ij}$, given $\Theta_i = \theta_i$, is assumed to be Poisson distributed with expectation $\nu_{ij}\theta_i$, the dividend $D_{ij}$ is Bernoulli distributed with the conditional probability of claims, and $D_{ij} = 0, p_{ij}(\theta_i) = 1 - e^{-\nu_{ij}\theta_i}$. 25 This also means that the (unconditional) expectation of $p_{ij}(\theta_i)$, $p_{ij} = \mathbb{E}[p_{ij}(\theta_i)]$, is a function of $\nu_{ij}$, hence it is individual and so is the dividend and the return. The expected individual dividend is: 26

$$\mathbb{E}[D_{ij}] = \mathbb{E}[\mathbb{E}[D_{ij} | \Theta_i]] = (1 - p_{ij})(\pi_{ij} + d_{ij})$$

Note that the insurer will set the dividend based on $p_{ij}$ while the policyholder will evaluate the proposition based on $p_{ij}(\theta_i)$. By the actuarial net premium principle we get the expression of the individual dividend and return:

$$\mathbb{E}[D_{ij}] = \pi_{ij},$$

$$d_{ij} = \pi_{ij} \frac{p_{ij}}{1 - p_{ij}},$$

$$\frac{d_{ij}}{\pi_{ij}} = \frac{p_{ij}}{1 - p_{ij}}.$$

This means that the net return on the entrance fee $\pi_{ij}$ is individual but fixed, for $N_{ij} = 0$. We set the net return in the case of $N_{ij} = 0$ to $r_{ij}$, hence $r_{ij} = \frac{d_{ij}}{\pi_{ij}}$. Figure 5.1 shows the distribution of $r_{ij}$, based on the data described in section A.1. The net return of the add-on can be used when comparing with the return of other risky assets

---

22The net premium principle is justified if we can assume that risk is essentially nonexistent; if the insurer sells enough identically distributed and independent policies.

23The dividend can of course be calculated via any other principle, e.g. the Expected Value Premium Principle: $\pi_{ij} = (1 + \xi) \mathbb{E}[D_{ij}]$, for some $\xi > 0$. But, we want to keep the intuition simple and avoid discussions which might distract us from the goal, even though they might be interesting.

24Young (2004).

25The estimator of the conditional probability of claims and the credibility estimators are defined in section A.2.

26For more extensive calculations and derivations, see Appendix A.2.
5.4 Pricing contracts: simulated evidence

We will study pricing of the insurance contract by comparing actual claim histories for a sample of policyholders. The implied dividends to the policyholders are compared

Figure 5.1: This figure shows the distribution of estimated $\hat{r}_{ij}$, the net return given that $N_{ij} = 0$, for $j = 2004$. The data is described in section A.1 and the estimators are found in section A.2.

In order to illustrate the size or affect of the dividend, table 5.1 shows numerical examples for three different policyholders with an entrance fee of one additional yearly premium. We note that the net return is larger for an assumed high-risk policyholder than an assumed low-risk policyholder, in resemblance to the entrance fee, why high-risks simply have more at stake.
5.4 Pricing contracts: simulated evidence

<table>
<thead>
<tr>
<th>Policyholder</th>
<th>( \hat{\nu}_{ij} )</th>
<th>( \hat{\mu}^2_{ij} )</th>
<th>( \hat{r}_{ij} )</th>
<th>( \hat{\pi}_{ij} )</th>
<th>( \hat{\mu}^2_{ij} + \hat{\pi}_{ij} )</th>
<th>( n_{ij} )</th>
<th>( D_{ij} )</th>
<th>( \hat{\mu}<em>{ij} + \hat{\pi}</em>{ij} - D_{ij} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low risk</td>
<td>10%</td>
<td>3000</td>
<td>10.55%</td>
<td>3000</td>
<td>6000</td>
<td>0</td>
<td>3316</td>
<td>2684</td>
</tr>
<tr>
<td>Low risk</td>
<td>10%</td>
<td>3000</td>
<td>10.55%</td>
<td>3000</td>
<td>6000</td>
<td>1</td>
<td>0</td>
<td>6000</td>
</tr>
<tr>
<td>Medium risk</td>
<td>20%</td>
<td>5000</td>
<td>22.21%</td>
<td>5000</td>
<td>10 000</td>
<td>0</td>
<td>6110</td>
<td>3890</td>
</tr>
<tr>
<td>Medium risk</td>
<td>20%</td>
<td>5000</td>
<td>22.21%</td>
<td>5000</td>
<td>10 000</td>
<td>1</td>
<td>0</td>
<td>10 000</td>
</tr>
<tr>
<td>High risk</td>
<td>35%</td>
<td>6600</td>
<td>42.03%</td>
<td>6600</td>
<td>13 200</td>
<td>0</td>
<td>9374</td>
<td>3826</td>
</tr>
<tr>
<td>High risk</td>
<td>35%</td>
<td>6600</td>
<td>42.03%</td>
<td>6600</td>
<td>13 200</td>
<td>1</td>
<td>0</td>
<td>13 200</td>
</tr>
</tbody>
</table>

Table 5.1: This table shows numerical examples of a possible add-on arrangement and how this affects three different policyholders. \( \hat{\nu}_{ij} \) is the estimated expected number of claims, for policyholder \( i \) during insurance period \( j \). \( \hat{\mu}^2_{ij} \) is the estimated actuarial premium. \( \hat{r}_{ij} \) is the net return, if \( N_{ij} = 0 \), which can be seen as an odds of having claims and is based on the expected probability of claims. \( \hat{\pi}_{ij} \) is the entrance fee. \( n_{ij} \) is the experienced number of claims and \( D_{ij} \) is the dividend. This means that \( \hat{\mu}_{ij} + \hat{\pi}_{ij} \) is the gross premium and \( \hat{\mu}_{ij} + \hat{\pi}_{ij} - D_{ij} \) is the net premium. Estimators are defined in section A.2.

with the premium and the entrance fee. The accuracy of the pricing method is then assessed by estimating - across all individual policyholders - the sum of squared errors (SSE).

The idea behind this is that the insurer wants to break-even on each contract, under perfect competition. A too low premium will result in an economical loss and a too high premium will result in a loss of the policyholders for which another insurer can offer a lower premium. Clearly, the minimum SSE is attended in the somewhat unlikely and theoretical situation where each customer pays exactly his own claims costs to the insurance company. The results below should clearly be interpreted in light of this. Still, a low SSE seems to be a reasonable description of the insurance company’s optimization problem. We use the sum of squared errors as a measure of performance, since the quadratic error function indicates that it is equally bad for the insurer to price an insurance too high as well as too low, on a competitive market. With a too high premium the insurer looses policyholders and with a too low premium the insurer looses money.

Our data comes from \( I = 8337 \) policyholders from the personal lines of a Danish insurance company. The policyholders, under consideration, have one and only one automobile insurance, with full coverage and full duration for all four years of the period 2001 – 2004. More on the data process in section A.1

\(^{27}\)Bühlmann and Gisler (2005).
Table 5.2 shows the accumulative sum of the experienced and expected number of claims and the number of policyholder with no claims at all and with fewer claims than expected, for each year, respectively.

<table>
<thead>
<tr>
<th>year</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>∑<em>i { \sum</em>{j=2001}^{i} n_{ij} }</td>
<td>1604</td>
<td>3381</td>
<td>5264</td>
<td>7007</td>
</tr>
<tr>
<td>∑<em>i { \sum</em>{j=2001}^{i} \hat{\nu}_{ij} }</td>
<td>1644</td>
<td>3441</td>
<td>5264</td>
<td>6969</td>
</tr>
<tr>
<td>∑<em>i 1 { \sum</em>{j=2001}^{i} n_{ij} = 0 }</td>
<td>7128</td>
<td>6058</td>
<td>5165</td>
<td>4531</td>
</tr>
<tr>
<td>∑<em>i 1 { \sum</em>{j=2001}^{i} n_{ij} &lt; \sum_{j=2001}^{i} \hat{\nu}_{ij} }</td>
<td>7128</td>
<td>6068</td>
<td>5365</td>
<td>5153</td>
</tr>
</tbody>
</table>

Table 5.2: This table shows the accumulated number of experienced and expected number claims and the number of policyholders with no claims and less number of claims than expected, respectively, for each year of the period of 2001-2004.

Empirically, for this data set, we get the dispersion index \( \Delta = \frac{\text{Var}(N_{ij})}{E(N_{ij})} \hat{\Delta} = 0.280/0.210 = 1.33 > 1 \) which supports the correctness in assuming a conditional Poisson distribution, i.e. assuming an individual latent risk profile.\(^{28}\)

We use the information in the period \( j = 2001, \ldots, 2003 \) to estimate our parameters. To compare different pricing methods we calculate the SSE for a number of pricing methods and look for the one with the minimum SSE, in year \( j = 2004 \), see table 5.3. The SSE is defined as

\[
SSE(l, j = 2004) = \sum_{i} \left( \hat{\mu}_{ij}^l - n_{ij} \right)^2
\]

where \( l \) is an index for the pricing method resulting in an estimate of the insurance premium \( \hat{\mu}_{ij}^l \), based on the estimation data \( (j = 2001, \ldots, 2003) \) and which might be independent of \( i \) for the simplest methods. The investigated pricing methods are the following: the mean value estimator, the flat rate, where \( \hat{\mu}_{ij}^1 = \frac{\sum_{j=2001}^{i} n_{ij}}{\sum_{j=2001}^{i} \omega_{ij}} \) for all policyholders; the regression estimator \( \hat{\mu}_{ij}^2 = \hat{\nu}_{ij} \) is the estimated expected number of claims based on a covariate-based regression (GLM and Maximum likelihood) as the one performed at an insurance company. The experience rating \( \hat{\mu}_{ij}^3 = \hat{\mu}_{ij}^2 \hat{\theta}_i \) is based on the one

\(^{28}\)If \( N_{ij} \mid \Theta_i \sim Po(\nu_{ij} \Theta_i) \) then
\[ E(N_{ij}) = E(E(N_{ij} \mid \Theta_i)) = E(\nu_{ij} \Theta_i) = \nu_{ij} E(\Theta_i) \quad \text{and} \quad \text{Var}(N_{ij}) = E(\text{Var}(N_{ij} \mid \Theta_i)) + \text{Var}(E(N_{ij} \mid \Theta_i)) = E(\nu_{ij} \Theta_i) + \text{Var}(\nu_{ij} \Theta_i) = \nu_{ij} E(\Theta_i) + \nu_{ij}^2 \text{Var}(\Theta_i) \]
This gives us the dispersion index:
\[ \frac{\text{Var}(N_{ij})}{E(N_{ij})} = \frac{\nu_{ij} E(\Theta_i) + \nu_{ij}^2 \text{Var}(\Theta_i)}{\nu_{ij} E(\Theta_i)} > 1 \] for all distributions of stochastic \( \Theta_i > 0 \).
5.4 Pricing contracts: simulated evidence

<table>
<thead>
<tr>
<th>Validation year</th>
<th>Pricing method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\mu}_{ij}^1$</td>
</tr>
<tr>
<td>scaled $SSE$ $(l, j = 2004)$</td>
<td>1.0324</td>
</tr>
</tbody>
</table>

Table 5.3: This table shows the ratio of the sum of squared errors for each pricing method relative to the regression estimator $\hat{\mu}_{ij}^2$, indicating the performance of the different methods. The mean estimator $\hat{\mu}_{ij}^1$ gives every policyholder the same premium, a flat rate. The credibility estimator $\hat{\mu}_{ij}^3$, the classical no-claims BMS $\hat{\mu}_{ij}^4$ and the add-on $\hat{\mu}_{ij}^5$ all use the claim information in the present insurance period to adjust the premium, why they naturally should have a lower scaled $SSE$.

dimensional homogeneous credibility estimator. The credibility model and estimators are defined and described in section A.2 of this paper.\textsuperscript{29} A classical market-based no-claims BMS $\hat{\mu}_{ij}^1$, gives a percentage share of the premium in bonus, if some requirements are fulfilled. The percentage can be dependent on how many insurance products the policyholder has purchased but commonly, and in this case it is 10%. The bonus is given if the following criteria are fulfilled: the policy has a duration of at least three years, no claims are reported during the last three years and furthermore no claims are reported during the current insurance period. Normally, this requires that all policyholders actually pay $3 - 4\%$ more of the original actuarial premium up-front, for the portfolio to be balanced. For this subset the additional payment is $9.533\%$ (corresponds roughly to one monthly premium) extra up-front.\textsuperscript{30} Finally, we calculate the $SSE$ for a standard contract with an add-on $\hat{\mu}_{ij}^5$, for different amounts of entrance fee, where all policyholders are assumed to buy the add-on. This is to be able to compare with the other contracts, and the assumption will be removed in section 5.5.

Based on the described pricing methods and data, table 5.3 shows the sum of square errors divided by the $SSE$ of the regression estimator, as benchmark. The flat rate $\hat{\mu}_{ij}^1 = 0.2091$ (meaning approximately one claim every fifth year in average) based on the estimation data in $j = 2001, ..., 2003$, is the simplest possible estimator with an expected non-negative profit. The regression estimator $\hat{\mu}_{ij}^2$ is the prevailing pricing method today, why we have used it as benchmark in table 5.3 (we divided the $SSE$ of the different pricing methods with the $SSE$ of the regression estimator). The actual $SSE$ $(l = 2, j = 2004) = 2360$. The difference in performance of the flat rate and the regression estimator, respectively, is quite important since it illustrates the improvement

\textsuperscript{29}See also Bühlmann and Gisler (2005) for an excellent survey on credibility theory.

\textsuperscript{30}Due to the data selection every policyholder is included in the bonus system and has the opportunity of getting a bonus.
Figure 5.2: This figure shows how the scaled SSE of the add-on in \( j = 2004 \) varies with the entrance fee, for a cross-section of the data discussed in section A.1. Each 'o' represents one monthly insurance premium and the line is the scaled SSE of the classical no-claims bonus.

in performance of a pricing method definitely worth implementing (since all insurance companies have pricing departments).

The last three contracts all use individual claim information, and are therefore expected to perform better than the two first ones. In the last column the values of SSE are based on the optimal value of the entrance fee \( \hat{\pi}_{ij} = 3.357 \hat{\mu}_{ij}^2 \) for the add-on. We investigated the interval \( \frac{\hat{\pi}_{ij}}{\hat{\mu}_{ij}} \in [0, 5] \) with the tolerance \( 10^{-6} \). Figure 5.2 shows the scaled SSE as a function of \( \pi_{ij} \).
Table 5.4: The table shows the terminal wealth for a given choice of insurance dependent on claims \((N_{ij} > 0)\) or no-claims \((N_{ij} = 0)\), when purchasing a standard insurance with or without an add-on, and the expected utility for each case, respectively. \(W_{i0}\) is the initial wealth for policyholder \(i\) in insurance period \(j\), and \(\mu_{ij}\) is the insurance premium, \(\pi_{ij}\) is the entrance fee. \(d_{ij}\) is the net return. \(p_{ij}(\theta_i)\) is the conditional probability of having claims.

### 5.5 Individual Choice and Self-selection

In the previous section we saw that prices become more fair and (theoretically) more competitive by setting individual bonuses depending on claims (riskiness). However, the optimal amount of entrance fee of roughly three yearly premiums is unrealistic for implementation. In the experiment in the previous section all policyholders were offered the same contract. In this section we change this assumption. Each policyholder is offered a choice between two different insurance contracts: a standard contract (with \(d_{ij} = \pi_{ij} = 0\)) and one with an add-on (i.e. \(\pi_{ij} > 0\) and \(d_{ij} > 0\)).

We expect that this will induce self-selection among insurance policyholders within the same risk class of the rating scheme.

We alter the experiment from section 5.4 by assuming that the policyholders choose their preferred insurance contract as the one with highest expected utility.

Denote by \(W_{i0}\) the initial wealth of individual \(i\) prior to the decision about insurance coverage without reference to a particular year. Initial wealth consists of all the policyholder’s possessions including cash, financial assets and the object that is insured. The policyholder is assumed to have more wealth than the insured object and enough cash to pay insurance premiums and entrance fees, thus \(W_{i0} > \pi_{ij} + \mu_{ij}\).

Now our individual \(i\) will decide to insure if and only if the expected utility associated with an insurance is larger than the expected utility associated with not being insured:
\(E(U(W_{i0} - \mu_{ij})) \geq E(U(W_{i0}))\). We ignore the decision whether to buy or not buy insurance. Thus, there are only 2 different outcomes for the terminal wealth depending on the binary decision on whether to buy insurance with or without an add-on. The outcomes of the insurance decision and the claims outcome are summarized in table 5.4, which in addition has the expected utility under the two insurance decisions.

From the bottom panel of table 5.4 it follows that the decision is to buy insurance with add-on if:

\[
EU(\hat{W}_{i0} - \hat{\mu}_{ij} - \hat{\pi}_{ij} + D_{ij}) > U(\hat{W}_{i0} - \hat{\mu}_{ij}).
\] (5.2)

For a more simple notation later on, we want to express the decision rule as a function of the return on the add-on investment. The decision rule (5.2) can be rewritten as

\[
EU\left[\left(\hat{W}_{i0} - \hat{\mu}_{ij}\right)\left(1 + \frac{D_{ij} - \hat{\pi}_{ij}}{\hat{W}_{i0} - \hat{\mu}_{ij}}\right)\right] > U\left(\hat{W}_{i0} - \hat{\mu}_{ij}\right).
\] (5.3)

For our numerical study we set the (private) risk profile of the individual to the second last credibility estimate (based on most information until the final year), and assume the policyholder to have a Constant Relative Risk Aversion (CRRA) utility function. For \(W > 0\) the CRRA utility function is defined by:

\[
U(W) = \begin{cases} 
W^{1-\gamma} / (1 - \gamma), & \gamma \neq 1, \\
\ln(W), & \gamma = 0.
\end{cases}
\]

The choice of CRRA finds support by Friend and Blume, Pindyck and Szpiro.\(^{31}\) Friend and Blume conclude that the individuals coefficients of relative risk aversion are “on average well in excess of one and probably in excess of two” based on data on household asset holdings. Pindyck finds support for a relative risk aversion between 3 and 4 and Szpiro finds support for a relative risk aversion coefficient between 1.2 and 1.8 based on data on property/liability insurance. In our simple scenario we assume a relative risk aversion coefficient \(\gamma = 2\), and investigate a number of alternative scenarios where we alter the initial wealth and so on. Under CRRA utility the decision rule (5.3) is to buy the add-on if:

\[
EU\left[1 + \frac{D_{ij} - \hat{\pi}_{ij}}{\hat{W}_{i0} - \hat{\mu}_{ij}}\right] > U(1).
\] (5.4)

\(^{31}\)Friend and Blume (1975), Pindyck (1988) and Szpiro (1986)
We see that under the CRRA assumption the decision rule (5.2) does not only depend on \( \hat{W}_i \), but rather on the relative amount of wealth put under risk, which we set to \( \hat{\beta}_{ij} = \frac{\hat{\pi}_{ij}}{W_{i0} - \hat{\mu}_{ij}} \). So, instead of finding a particular level of wealth we simply seek the investment ratio \( \hat{\beta}_{ij} \). Now set \( R_{ij} = (D_{ij}/\hat{\pi}_{ij}) - 1 \) to be the net return of an insurance contract with an add-on. With the CRRA utility we now rewrite the decision rule (5.4) as:

\[
E \left[ \left( \hat{\beta}_{ij}(1 + R_{ij}) + (1 - \hat{\beta}_{ij}) \right)^{1-\gamma} \right] > 1.
\]

To compare the add-on with other alternative investments we could let \( R_{ij}^a \) be the return of any other (risky) investment and then the decision is to buy the add-on if:

\[
E \left[ \left( \hat{\beta}_{ij}(1 + R_{ij}) + (1 - \hat{\beta}_{ij}) \right)^{1-\gamma} \right] > E \left[ \left( \hat{\beta}_{ij}(1 + R_{ij}^a) + (1 - \hat{\beta}_{ij}) \right)^{1-\gamma} \right]
\]

but to keep the discussion as simple as possible we will set \( R_{ij}^a = 0 \) in our further study.

In our simulations we let the parameter \( \beta_{ij} \) vary to find the optimal ratio. If we take an arbitrary policyholder, with a fixed risk aversion coefficient, there are three variables affecting the decision rule of to buy or not to buy the add-on, based on CRRA: the conditional probability of claim \( p_{ij}(\theta_i) \), the net return \( R_{ij} \) and the investment ratio \( \beta_{ij} \). According to figure 5.3, we can find an interval, with a specific optimal \( \beta_{ij} \), that makes it a rational choice to buy the add-on, as long as the policyholder’s conditional probability of claim \( \hat{p}_{ij}(\theta_i) \) is less than what the insurer expect \( \hat{p}_{ij} \).

As suggested above the add-on can be designed in various ways. But, there are two basic approaches we want to investigate.

Either, all individuals are offered the same investment ratio \( \hat{\beta}_{ij} = \hat{\beta} \) and we seek the optimal investment ratio \( \tilde{\beta} \), measured in least SSE, for the whole portfolio.\(^{32}\) Figure 5.4 shows how many policyholders who choose to buy the add-on on different levels of \( \hat{\beta} \).

The optimal ratio for the portfolio, meaning that all policyholders are proposed with the same ratio, measured in least SSE, is summarized in table 5.5.

\(^{32}\)To calculate the SSE for \( l = 5 \) in equation 5.1 we calculate the entrance fee by multiplying the investment ratio by the different levels of assumed wealth and dividing by the expected claim severity, \( \hat{\pi}_{ij} = \frac{\hat{\mu}_{ij} W_{i0}}{\nu_{ij} m_{ij}} \hat{\nu}_{ij} \).
Figure 5.3: This figure shows how the net utility of equation 5.4 varies with the investment ratio $\beta$ and the conditional probability of claims $p_{ij} (\theta_i)$. If the net utility is positive, the policyholder chooses to buy the add-on. The example of the figure is based on an average policyholder with the estimated expected probability of claims $\hat{p}_{ij} = 0.1798$, and is offered a net return $\hat{r}_{ij} = 0.2193$ and has an assumed risk aversion coefficient $\gamma = 2$. Estimators used to estimate the probability of claims are defined in section A.2.

Table 5.5: This table shows the optimal ratio of wealth to invest, or bet, as entrance fee if all policyholders are proposed with the same investment ratio $\beta$ and are assumed to have the same initial wealth $W_{i0}$ and risk aversion coefficient ($\gamma = 2$).
Figure 5.4: This figure shows the demand curve of the add-on, i.e. the number of policyholders buying an add-on for different investment ratios based on cross-section of 8337 policyholders described in section A.1 and the decision rule (5.4). The demand curve is indifferent to initial wealth, in accordance with the theory.
Table 5.6: This table shows the scaled SSE of a standard insurance with add-on, when the estimated optimal investment ratio $\tilde{\beta}_{ij}$ of each individual is used as entrance fee, and each policyholder is assumed to have the same initial wealth $W_0$ and risk aversion coefficient ($\gamma = 2$). 5246 policyholders choose to buy the add-on at this optimal investment ratio. The scaled SSE uses the SSE of the regression estimator $\hat{\mu}_{ij}^2$ as benchmark. Estimators are defined in section A.2.

<table>
<thead>
<tr>
<th>Initial wealth ($W_0$) (1000 DKK)</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>scaled SSE (5, 2004)</td>
<td>0.9707</td>
<td>0.9427</td>
<td>0.8659</td>
<td>0.7623</td>
<td>4.125</td>
</tr>
</tbody>
</table>

buy an add-on at these portfolio optimal investment ratios is 18.84%, while it is 22.65% for those who do not buy the add-on.

Or, we can propose each individual their individually optimal investment ratio $\tilde{\beta}_{ij}$ and see how large or small the scaled SSE gets.\(^{33}\) In table 5.6 we see the scaled SSE for different assumptions on initial wealth, when every policyholder may invest their optimal entrance fee, seen as a ratio of their wealth. 5246 policyholders choose to buy the add-on at their $\tilde{\beta}_{2004}$.

Figure 5.5 shows the distribution of individually optimal ratio $\tilde{\beta}_{2004}$, independent of the magnitude of initial wealth.

The average experienced claim frequency in year 2004 for those policyholders who buy an add-on at their individually optimal investment ratio is 20.85%, while it is 21.00% for those who do not buy the add-on.

Another variation, in relation to the one in section 5.4, would be to offer the policyholders an entrance fee where the amount is in relation to their insurance premium, seen in table 5.7. We get the following performance results, for different initial wealth, still with the regression estimator as benchmark:

The average experienced claim frequency in year 2004 for those policyholders who buy an add-on when the entrance fee is an portfolio optimal ratio of the individually actuarial premium is 18.79%, while it is 23.16% for those who do not buy the add-on.

\(^{33}\)Since we do not know the wealth of each policyholder in practise we would simple let them choose the amount put at risk by themself.
5.5 Individual Choice and Self-selection

Figure 5.5: This figure shows the distribution of the estimated optimal investment ratio for each individual $\tilde{\beta}_i$, in the cross-section of policyholder described in section A.1, independent of assumptions of the magnitude of initial wealth.

<table>
<thead>
<tr>
<th>Initial wealth ($W_{i0}$) (1000 DKK)</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\pi}_{ij}$</td>
<td>0.06366</td>
<td>0.2851</td>
<td>0.8576</td>
<td>1.605</td>
<td>3.566</td>
</tr>
<tr>
<td>$\hat{\mu}_{ij}^2$</td>
<td>0.9837</td>
<td>0.9560</td>
<td>0.8948</td>
<td>0.8147</td>
<td>0.6046</td>
</tr>
<tr>
<td>scaled $SSE$ ($5, 2004$)</td>
<td>0.9837</td>
<td>0.9560</td>
<td>0.8948</td>
<td>0.8147</td>
<td>0.6046</td>
</tr>
<tr>
<td>number of buyers</td>
<td>4781</td>
<td>4147</td>
<td>4010</td>
<td>4178</td>
<td>5148</td>
</tr>
</tbody>
</table>

Table 5.7: This table shows the scaled SSE, with the SSE of $\hat{\mu}_{ij}^2$ as benchmark, for different initial wealth $W_{i0}$, when the entrance fee is a multiple of the individual insurance premium. The multiple is the same for each individual, while the insurance premium of course is individual. The SSE of the add-on is now based on the individual choice of the policyholders.
5.6 Discussion

As seen in the previous sections this type of add-on to the standard insurance can be designed in various ways; both the criterion of getting a dividend and the size of the dividend can easily be changed and adapted to a specific business line or area. For instance, no-claims $N_{ij} = 0$, are not that common, or relevant as a criterion, for policyholders in areas with high claim frequency, e.g. glass insurance within commercial transport. In such areas of business a criterion based on a boundary of number or severity of claims, fixed or in relation to the expectations, might be more appropriate. This will of course affect the return, but that is up to the interested reader to investigate.

We have wanted to keep the model as simple and as descriptive as possible in this paper, hence we kept the criterion of no-claims. The size of the entrance fee $\hat{\pi}_{ij}$, or dividend $d_{ij}$, will be a strategic decision based on policyholder insight or business strategy of the insurer, or by the policyholder herself. Either one will just be a linear function of the other (since $\hat{\rho}_{ij}$ is individually fixed). The size of $\hat{\pi}_{ij}$ will decide if the add-on is to be considered as a part of the betting budget, or an important part of the extended household balance sheet. This will of course be affected by the preferences and risk perception of the policyholders, and in an asset portfolio consideration the add-on can be compared to the return of other risky assets. The policyholders might choose other types of investments, e.g. stock, if the offered net return $\hat{\rho}_{ij}$ is considered too low. Note, however, that the dividend of the add-on is most likely not correlated to the ordinary market risks or assets, as stock and bonds, but to other ones as the insured objective (car, house or boat etc.) or the insured person.

When comparing the different pricing methods, in the initial study without individual choice (table 5.3), we see that the credibility estimator $\hat{\mu}_{ij}^3$ gives an additional improvement of 25%, compared to going from the flat rate $\hat{\mu}_{ij}^1$ to regression $\hat{\mu}_{ij}^2$, but it is not as good as we had expected. Earlier studies show a potential of the credibility estimator of the same magnitude as going from the flat rate to the regression-based estimator, but this is of course also dependent on type of coverage and data set.$^{34}$ However, one could raise questions regarding the amount of information; four years of information is perhaps too few to fully benefit from the credibility approach.

The performance of the classical no-claims bonus $\hat{\mu}_{ij}^4$ is however surprisingly good.

$^{34}$Englund et al. (2008) and Englund et al. (2009).
5.6 Discussion

(an additional improvement of 52% compared with implementing the regression) and at the same time rather simple. It gives an improvement of the same magnitude as going from flat rate to regression estimator. This type of classical bonus systems are however afflicted with some drawbacks such as fairness and stability aspects. Fairness: Giving a fixed percentage in bonus is advantageous for the expected low-risk policyholders and disadvantageous for the expected high-risks compared to their odds of reporting claims, however it can be justified if, or at least explained by, the insurer has a strategy towards expected low-risks. Stability: If all customers qualify for the bonus system (purchase the required number of products etc.), then the system would not be stable, in the sense that the insurer will lose money if the initial extra payment is not reestimated.

Remember that one of the fundamental principles of insurance is that the policyholder should pay for the expected transferred risk, not the outcome. The add-on deals with this while it retains the advantages of simplicity, low requirements on information, and it can perform even better than both the credibility estimator, \( \hat{\mu}_{ij}^3 \), and the 10% no-claims bonus system, \( \hat{\mu}_{ij}^4 \), depending on the size of \( \hat{\pi}_{ij} \). The scaled SSE of the add-on with the optimal entrance fee is exceptionally low, almost 20 times the improvement of implementing the regression estimation (instead of the flat rate), meaning we have found a fantastic pricing method for competitive markets. However, we doubt that any policyholder will put up the optimal entrance fee of more than three annual insurance premiums up front and risk it all. Even though the positive net return in average is

\[
\frac{1}{I} \sum_{i} \hat{r}_{2004} = 22.25\% ,
\]

in case of \( N_{2004} = 0 \), see figure 5.1, and the add-on thereby can be considered a rather good investment for more than half of the policyholders.\(^{35}\)

According to table 5.2 54.35% of the policyholders do not report any claim during the period of 2001-2004. However, an entrance fee of this magnitude could be a good alternative to an increased deductible.\(^{36}\) Moreover, for a small amount of entrance fees we can get quite good results. In figure 5.2 we see that with an entrance fee of (less than) one monthly insurance premium (\( \frac{1}{12} = 8.333\% \)) we get a lower scaled SSE than for the classical no-claims bonus (where the additional payment is 9.533%). Hence, we get a more fair pricing, just by giving the policyholder an individual percentage share of bonus, based on their expected risk, instead of a collective fixed rate for every policyholder.

When we allow individual choice in section 5.5 we are interested in the performance

\(^{35}\)This should of course be seen in the perspective of each individuals investment strategy.

\(^{36}\)For a broader discussion on how to set the level of deductibles, coinsurance and design of paragraphs to force the policyholder to self-protection and self-insurance, see Ehrlich and Becker (1972).
Asymmetric Information and Self-selection

of the add-on both as a pricing method and as a separating (self-selection) mechanism. In table 5.5 we see that the optimal investment ratio (ratio of wealth at risk) for the entire portfolio is decreasing, while the entrance fee is increasing with an increasing wealth. This results in a decreasing SSE since each policyholder bears more of her own risk. The increasing number of policyholders buying the add-on is due to the decreasing $\beta$, while the risk aversion is kept constant. Notably is however that we already with an entrance fee of 620 DKK get an additional improvement of the SSE of 80% compared going from the flat rate to the regression estimator. Just by adding individual choice!

In table 5.6, where 5246 policyholders buy the add-on at their individually optimal ratio of wealth at risk (see figure 5.5), we get an even better pricing method than in table 5.5 for all assumptions on initial wealth, e.g. an additional improvement of 88% for $W_{i0} = 10 000 \, DKK$, but $W_{i0} = 1 000 000 \, DKK$. For this initial wealth there obviously has to be some restrictions for the entrance fee to be a good pricing method. Such restriction could for instance be that the net dividend $d_{ij}$ may not exceed the actuarial premium, or else the squared error will increase the larger the dividend is. To be kept in mind is also that under the assumptions made there is only one risky asset, the add-on, to invest in and the wealthy policyholders are in that way forced to put all their investments in this asset. This assumption might not fit too well with reality. For implementational reasons the insurer might offer the policyholder with a stepwise function, instead of a continuum of eligible entrance fees. The precision of the function may depend on the precision of risk perception of the policyholders and the costs of implementation and systems maintenance and so on.

To link to the investigations in section 5.4, and since it might be easier to relate to the insurance premium than the initial wealth both for the policyholder and the insurer, we also investigated the entrance fee expressed as a ratio of the actuarial premium, but now allowing for individual choice. The results are found in table 5.7. Generally the SSE is somewhat higher than in table 5.6 and 5.5 but the number of policyholders buying the add-on is larger than in table 5.5. All SSE are less than for the present no-claims bonus system, except for $W_{i0} = 10 000 \, DKK$, and the performance is increasing for the assumptions of larger initial wealth, but so is the entrance fee.

So, we have found that each policyholder has an optimal entrance fee, depending on initial wealth, individual risk profile and risk aversion. Since the insurer usually does not know the initial wealth or risk aversion of each individual the insurer may let the
policyholder choose the size of entrance fee, by herself, to get an optimal affect of the self-selection mechanism when pricing insurance. The more the individual can choose by herself the larger the price-coverage differentiation becomes and thereby also the competitiveness (due to fair pricing). However, the add-on will not be as effective as a screening device.

5.7 Conclusions

In this paper we have proposed an alternative way to turn a problem of asymmetric information into a solution of how to price insurance in groups with heterogeneous risk. The design of the proposed add-on can vary in many ways, but we have at least showed that there might be both entrance fees and dividend at realistic levels.

Under the assumptions made we have found that the insurer easily could improve the pricing by implementing a bonus system where every policyholder has an opportunity of a bonus with an individual bonus rate, in contrast to the present fixed ones for a limited subset of policyholders.

An even larger improvement in pricing performance is obtained by letting the policyholders choose, to buy the add-on or not, by themselves. A third and last improvement is obtained by letting the policyholders choose the size of entrance fee, based on their private information on risk, preferences and wealth. However, the more the price is differentiated the less the add-on separates the policyholders and vice versa.

Note that all the analysis of this paper is based on real data, but the decisions of the policyholders are simulated. Naturally, it would have been more interesting to investigate the performance of the add-on based on decisions made by real policyholders, but that will be the subject of another paper.
Appendix A

Appendix

A.1 Data Process

Out of well over 375,000 policyholder in the personal lines portfolio of a Danish insurance company, we choose the ones with one and only one automobile insurance, with full coverage, full duration for all four years of the period 2001 – 2004 and no experience rating. In order to compare the classical bonus, with a fixed bonus rate of 10%, with the add-on, with individually set bonus percentage, as fair as possible, each policyholder has to fulfill the criteria of the classical bonus, to the extent that the bonus is only dependent on the claim experience. Therefore we also restrict our data set to policyholders who besides the automobile insurance also have purchased Building and Personal Property insurance, with full duration in the period 2001 – 2004. The data is processed in order to remove policyholders with missing values and obvious outliers. This altogether resulted in a rather diminished data set of \( I = 8337 \) policyholders. The data set contains the variables used by the pricing department at the Danish insurance company, for each of the years during the period. The covariates are for instance: policyholder’s demographic characteristics such as age, gender, geography; policyholder’s car characteristics: brand, size of engine, model year, initial value of the car, commercial vehicle or not, main vehicle or not; period covered: the length (duration) of the period covered by the purchased policy (in this case each policyholder has full duration per year).

In our data set there is a difference between the total expected and total experienced number of claims, which explained by the pricing procedure. The rating scheme is
estimated on all policyholders of the automobile portfolio, while we use only a subset of these. One might suspect that policyholders with full duration in four consecutive years, and who have purchased at least three insurance product, have a better claim history than we expected, even though total duration (seniority) is not a significant variable in the rating scheme. Due to this difference, and our interest to investigate the performance of the pricing methods, we scaled the original claim frequency estimates of the regression, before we perform the analyses of this paper. The regression estimates are scaled with a factor $\frac{\sum_i \sum_j n_{ij}}{\sum_i \sum_j \nu_{ij}} = 0.9164$, where $i \in [1, 8337]$ and $j \in [2001, 2003]$.

### A.2 Credibility Model and Estimators

We use the same definition and treatment of standardized experienced claim frequency with prior knowledge as Bühlmann and Gisler: $F_{ij} = \frac{N_{ij}}{\nu_{ij}}$.\(^1\)

Model assumptions:

(i) Conditionally, given $\Theta_i$, the random variable $N_{ij}$ are independent and Poisson distributed with expected value $E[N_{ij} | \Theta_i] = \nu_{ij}\Theta_i$. The conditional first and second order moments are: $E[F_{ij} | \Theta_i] = \Theta_i$ and $Var[F_{ij} | \Theta_i] = \sigma^2(\Theta_i)\nu_{ij}$.

(ii) The pairs $(\Theta_1, N_{1j}), (\Theta_2, N_{2j}), \ldots, (\Theta_I, N_{Ij})$ are independent random vectors, where $\Theta_1, \Theta_2, \ldots$ are iid with $E(\Theta_i) = \theta_0$ and $V(\Theta_i) = \tau^2$.

The interpretation of the parameters are as follows. $\theta_0$ is the collective risk profile, $\sigma^2(\Theta_i)$ is the variance within individual risk, and $\sigma^2$ is the average of $\sigma^2(\Theta_i)$. $\tau^2$ is the variance between individual risk profiles.

**Estimators**

With this standardization we get that the best linear estimator of $\Theta_i$ is

$$\hat{\Theta}_i = \hat{\theta}_0 + \hat{\alpha}_i \left( F_i - \hat{\theta}_0 \right)$$

\(^1\)Bühlmann and Gisler (2005)
or seen as a weighted mean between individual and collective data

\[
\hat{\Theta}_i = \hat{\alpha}_i F_i + (1 - \hat{\alpha}_i) \hat{\theta}_0
\]

where

\[
\hat{\theta}_0 = \frac{\sum_{i=1}^{I} \hat{\alpha}_i F_i}{\sum_{i=1}^{I} \hat{\alpha}_i}, \quad \hat{\alpha}_i = \frac{\hat{\nu}_i}{\hat{\nu}_i + \hat{\sigma}^2}
\]

\[
F_i = \frac{\sum_{j=1}^{J} N_{ij}}{\sum_{j=1}^{J} \hat{\nu}_{ij}}, \quad \hat{\nu}_i = \sum_{j=1}^{J} \hat{\nu}_{ij}
\]

The estimate \(\hat{\theta}_0\) can be interpreted as a calibration factor; if the experienced number of claims is larger than expected, the estimate of \(\hat{\theta}_0\) is larger than 1, and vice versa.

For the estimation of \(\tau^2\) we use an estimator proposed by DeVylder. \(^2\) This is justified by the comparison of several estimators which found that DeVylders estimator is the best one to use for standardized claim frequencies.\(^3\) The estimator has the following expression.

\[
\hat{\tau}^2 = \hat{c} \left( \frac{1}{I-1} \sum_{i=1}^{I} \frac{\hat{\nu}_i}{\hat{\nu}} (F_i - F)^2 - \frac{\hat{\sigma}^2}{\hat{\nu}} \right)
\]

\[
\hat{c} = (I-1) \left( \frac{1}{\hat{\nu}} \sum_{i=1}^{I} \hat{\nu}_i \left( 1 - \frac{\hat{\nu}_i}{\hat{\nu}} \right) \right)^{-1}
\]

\[
\hat{\sigma}^2 = \frac{1}{J-I} \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{\nu}_{ij} (F_{ij} - F_i)^2
\]

where \(\hat{\nu} = \sum_{i=1}^{I} \hat{\nu}_i\) and \(F = \hat{\nu}^{-1} \sum_{i=1}^{I} \sum_{j=1}^{J} N_{ij} \).\(^4\) We now have estimators for each parameter in use. In the present data set we get the following estimates of the structural parameters

\[
\hat{\theta}_0 = 1.003, \quad \hat{\sigma}^2 = 1.256, \quad \hat{\tau}^2 = 0.3267
\]

When estimating \(p_{ij} = E[p_{ij}(\Theta_i)]\) we approximate \(p_{ij}(\Theta_i) = 1 - e^{-\nu_{ij}\Theta_i}\) by its second order Taylor approximation, around \(\Theta_i = \theta_0\):

\[
p_{ij}(\Theta_i) = p_{ij}(\theta_0) + \frac{p'_{ij}(\theta_0)}{1!} (\Theta_i - \theta_0) + \frac{p''_{ij}(\theta_0)}{2!} (\Theta_i - \theta_0)^2 + R_3(\Theta_i)
\]

\(^2\)De Vylder (1978)
\(^3\)Dubey and Gisler (1981)
\(^4\)See Englund et al. (2008) and Englund et al. (2009) for multidimensional credibility models and estimators.
where $R_3(\Theta_i) = \frac{p''_ij(\xi)}{3!} (\Theta_i - \theta_0)^3$, $\theta_0 \leq \xi \leq \Theta_i$, such that

$$p_{ij} = E[p_{ij}(\Theta_i)] = 1 - e^{-\nu_{ij}\theta_0} + \nu_{ij} e^{-\nu_{ij}\theta_0} \left( (\Theta_i - \theta_0) - \frac{\nu_{ij}^2 e^{-\nu_{ij}\theta_0}}{2!} E[(\Theta_i - \theta_0)^2] + E[R_3(\Theta_i)] \right)$$

where $E[R_3(\Theta_i)] = \frac{\nu_{ij}^3 e^{-\nu_{ij}\xi}}{3!} E[(\Theta_i - \theta_0)^3]$. In our sample where the maximum individual expected number of claims is $\hat{\nu}_{ij} = 0.8302$ and $\frac{1}{t} \sum (\hat{\theta}_i - \hat{\theta}_0)^3 = 0.01834$, this gives us a remainder term $\hat{R}_3(\Theta_i) < 0.0017490$, why the second order approximation is really sufficient for our purpose. We therefore estimate $\hat{p}_{ij} = 1 - e^{-\hat{\nu}_{ij}\hat{\theta}_0} - \frac{\hat{\nu}_{ij}^2 e^{-\hat{\nu}_{ij}\hat{\theta}_0}}{2} \hat{\tau}^2$. 
Bibliography


