Multivariate credibility with application to cross-selling financial services products

Fredrik Thuring
Faculty of Actuarial Science
Sir John Cass Business School, City University, London, UK
Supervisors: Prof. Vladimir K. Kaishev and Prof. Jens Perch Nielsen

A thesis submitted for the degree of
PhilosophiæDoctor (PhD)

2012 December

To V.

Abstract

In this thesis, methods that are capable of improving the revenue and profitability of a financial services company are presented. Of particular interest is the use of customer specific information for pricing insurance products and segmenting a customer population based on the expected profitability of the customers. A prerequisite is the possibility for customers to have many different financial services products from the same provider. The thesis presents multivariate credibility models for how customer specific information from one (or many) financial services products is related to customer specific information from another financial services product. The models are foremost applied to the context of cross-selling (selling additional products to existing customers) where customer specific information from the offered cross-sale product is not available before the sale. As products are related, it is reasonable to use an appropriate (credible) amount of customer specific information from another product (or products), for estimating the profitability expected to emerge from the offered cross-sale product. In four separate but related articles, it is shown that having appropriate models for pricing and customer segmentation is of great importance for a financial services company aiming at running a profitable and growing business.

Acknowledgements

I take the opportunity to thank my supervisors Professor Vladimir Kaishev, for the many enjoyable hours spent discussing and writing the different manuscripts, and Professor Jens Perch Nielsen, for his encouragement, our discussions and his ability to manage the propulsion of my PhD-project.

I would also like to thank my co-supervisors Professor Mogens Steffensen, for the many informal and interesting meetings, and Professor Montserrat Guillén, for making me feel welcomed at UB and for the productive atmosphere during our work. I also thank Dr. Catalina Bolancé for our collaboration at UB.

At Codan Forsikring A/S, there are many people to acknowledge. My former manager Dennis and my current manager Per, for helping me balance the work load. Martin, for the countless hours spent discussing literature and programming code. I also thank Josef, Kristoffer, Jim and Magnus for helping me with theoretical challenges, discussions about practical implications and data extraction. Furthermore I thank my family and friends for their interest in my progress and their belief in me. Finally and especially I thank Hillevi, for her support, understanding and love.

Contents

1

Introduction

This dissertation presents methodologies for pricing and direct marketing of financial services products (insurance cover, mortgage contracts, general loans, etc.), provided by financial institutions (insurance companies and banks), with particular reference to the profit generated by each customer. The problem of setting the correct price, to a specific customer with respect to a specific product, is of paramount importance for any financial institution aiming at running a profitable business. Additionally, companies that are able to identify and target the right existing customers, to offer them additional financial services products, will benefit hugely in terms of growth of revenue and market share. Direct marketing to existing customers by offering them additional products or services is referred to as cross-selling and constitutes the main track of research in this thesis.

Most financial services products differ from conventional retail products in the way that the cost, associated with a specific customer with respect to a specific product, is stochastic and becomes known to the financial institutions after some (also stochastic) time. For example, an insurance company offers insurance cover, making them liable to give economic compensation to cover eventual losses, related to events covered by the insurance policy. Banks can provide financial means for its customers to make investments or purchases, however if the investments fail or the customers are unable to repay their dept, the banks suffer economic losses. It is crucial for financial institutions to be able to segment its customers with respect to the probability and size of these future losses as well as to price its liabilities towards its customers.

1. INTRODUCTION

Successful cross-selling of financial services products is highly dependent on the ability, of the financial institution, to segment its customers with respect to the customers' potential for generating future profits. Normally, financial institutions have hundreds of thousands of customers in its database and since cross-selling usually involves personal communication with the customers, it is usually a formidable task to approach all of them. Instead, financial institutions carefully select a subset of customers to approach, to whom certain additional financial products should be offered. Approaching a described subset of customers is usually referred to as launching a cross-sale campaign and obviously, the quality of the selection methodology affects the financial result from the campaign.

Prior to launching a cross-sale campaign, it is common practice to segment the customers of a financial institution with respect to their estimated probability of purchasing the offered product. In such cases, the estimated probability is the outcome of a binary response model (normally a logistic regression model) evaluated with observable explanatory factors of the customers in the database of the financial institution. I.e, customers with high estimated purchase probability is selected to be approached. One of the main research findings in this dissertation is that this frequently used and established methodology of segmenting customers only based on the estimated probability of purchase can be challenged and improved, especially for the financial services industry.

A prerequisite for the research presented in this dissertation, is that customers of a financial institution have (or at least could have) multiple financial services products from the same provider. Additionally, the stochastic events (insurance claims, loan defaults, etc.), associated with the financial services products of a specific customer, are related. I.e., for a specific customer, the occurrence of events associated with one of his or her purchased products is correlated with the occurrence of events associated with another purchased product. All studies in the thesis are limited to only the occurrence of these mentioned events, thereby assuming that the corresponding size of the generated losses are unrelated. This is an assumption that could be challenged but is made in order to limit the scope of the thesis and thereby left for future research. Nonetheless, the assumed correlation between the occurrence of the stochastic events, from multiple products of the same customer, is imposed in the described models in the following chapters. Taking this correlation into consideration, when segmenting the database of a financial institution, the overall profit from a particular cross-sale campaign can be improved significantly.

In this thesis, the number of occurred events, generated from correlated financial services products of a specific customer, is related to some a priori notion of the expected number of events. I.e throughout the dissertation, it is assumed that financial institutions have methodologies for assessing an a priori frequency or probability of occurrence of the events. In the most simple cases, this assessment can be made from only a qualitative analysis of the customers but is normally made by evaluating a previously estimated model with certain characteristics (explanatory factors) of the customers. For example, it is assumed that a financial institution is able to assess the a priori expected number (or probability) of events $\hat{\lambda}_i$, associated with a specific customer i, as $\hat{\lambda}_i = \hat{f}(\mathcal{Y}_i)$, where $\mathcal{Y}_i$ is an appropriate set of characteristics for the customer i and $\hat{f}$ is the previously estimated regression function.

Another important assumption, that is made implicitly trough out the thesis, is that the success of a cross-sale approach is independent of the risk characteristics of an individual. I.e. customers identified as having low risk characteristics have the same propensity to purchase an offered product as those associated with high risk characteristics. In real life applications of the presented models, this assumption would probably fail since the risk characteristics in many cases affects the price via experience rating. Therefore customers identified as having low risk characteristics would be harder to cross-sell to since the financial company, at which they currently have the product, might be giving an experience rated discount. Not dealing with these kind of correlations presents a risk of anti-selection, especially when models for sales probability and risk characteristics are combined, and is neglected in the thesis but should be considered prior to any real life implementation.

In most cases, a financial services product is associated with a certain duration for which the product is valid, for example an insurance policy is usually bought as a yearly cover and a mortgage (or other loan) is associated with a fixed interest rate during a specific time period. The a priori estimate $\hat{\lambda}_i$ is related to the corresponding time period of the product, for example for an insurance company, $\hat{\lambda}_i$ is usually the a priori expected number of claims associated with a certain customer i for a certain insurance product. Correspondingly for a bank, $\hat{\lambda}_i$ is the a priori probability that the customer i will default on a certain loan product from the bank. By comparing the observed number of events n_i (available to the financial institution at the end of the time period) to the a priori expected number $\hat{\lambda}_i$, the financial institution is able to assess if a specific customer i was associated with more or less events than a priori expected, during that time period. It should be noted that n_i is a realisation of the discrete stochastic variable N_i. If the customer has had a certain product with the same provider during multiple time periods, the financial institution is able to compare the total number of events related to customer i, $\sum_{j=1}^{J_i} n_{ij}$, to the total number of a priori expected events $\sum_{j=1}^{J_i} \hat{\lambda}_{ij}$, where J_i is the number of time periods.

The deviation between the a priori expected number of events $\hat{\lambda}_i$ and the corresponding actual number of events N_i is central in the proposed new methodology for cross-selling financial services products. In this dissertation, it is argued that a multivariate credibility model, based on the deviation between $\hat{\lambda}_i$ and n_i, is useful as a complement to, or in combination with, the established methodology of segmenting customers only based on the estimated probability of purchase. The multivariate credibility model, elaborated on in this thesis, is the multivariate Bülhmann-Straub model and throughout the thesis the term multivariate is interpreted as multiple financial services products.

A multivariate credibility model is suitable in cross-selling since it allows for correlation between the entering variables (in this case the events associated with multiple financial services products). Additionally, using a multivariate credibility model, information from different sources is weighted based on the relevance (or credibility) of that information. In chapters 3, 4 and 5 it will be shown that a "credible amount" of information about $\hat{\lambda}_i$ and n_i for one financial services product can be used to estimate certain quantities, related to the profit, for another financial services product. When cross-selling financial services products (or products in general), the approached customers already have at least one existing product from the company. One of the scientific contributions from this dissertation is to show how information from an existing product (or set of products) can be used in estimating the profit that is expected to emerge from approaching a specific customer to offer a specific additional product.

The rest of this dissertation consist of four research papers which are self contained with their own introductions, where related prior work and references are described.

Consequently, reference to prior research is absent in this introduction. The rest of the introduction will give a description of the four papers and their contribution.

Chapter 2 presents the paper *Multidimensional Credibility with Time Effects - An Application to Commercial Business Lines*. The chapter focuses on improving the current pricing scheme of a commercial lines insurance business, by altering the a priori claim frequency estimate using experience rating. Experience rating is introduced via a multivariate credibility model, with or without a dependence of the age of the claim information. A definition is found in the paper, of a customer specific variable which is dependent of the deviation between the a priori expected number of events $\hat{\lambda}_i$ and the corresponding actual number of events N_i. This definition is used in the following chapters of the dissertation. The paper presents a multivariate credibility estimator of the customer specific risk profile θ_i, which takes into account the dependence in claim frequency between the different insurance products. To produce an improved claim frequency estimate (compared to the a priori estimate $\hat{\lambda}_i$), the estimate of the risk profile $\hat{\theta}_i$ is multiplied with the a priori claim frequency estimate resulting in a posterior claim frequency estimate dependent on the individual claims experience. Different credibility estimators are evaluated using information from different number (and combinations) of the insurance products and conclusions are drawn from an out-of-sample validation study. The contribution from the paper is a methodology for improving insurance pricing, with special reference to claim prediction, using experience from multiple insurance products via a multivariate credibility model.

Chapter 3 presents the paper *A credibility method for profitable cross-selling of insurance products*. This is the first of three consecutive chapters where the (time-independent) model from Chapter 2, is used for cross-selling financial services products. The study is on the simple cross-sale case where customers of an insurance company have only one insurance product with the company, which seeks to approach a subset of them to offer a second one. In contrast to Chapter 2, the objective is not to alter an a priori claim frequency estimate but rather to use the estimate of the risk profile $\hat{\theta}_i$ by itself, to segment the customer population with, and approach customers with a low risk profile estimate. The credibility estimator for this particular situation is shown to be a special case of the bivariate credibility estimator from Chapter 2. The methodology is evaluated in a real data study of personal lines insurance customers who all (at the time of data collection) were in possession of three different insurance products from an insurance provider. In the data study, it is imagined that the customers are in possession of only one of the three products and the methodology is used to evaluate if information about $\hat{\lambda}_i$ and n_i, with respect to this product, can be used for segmenting the population, with respect to deviation between $\hat{\lambda}_i$ and n_i for another product. From the results, it is concluded that the correlation between the different insurance products is captured by the credibility estimator and that the methodology is able to differentiate the customer population with respect to profit for a product not belonging to the customers.

Chapter 4 presents the paper *Selecting prospects for cross-selling financial products using multivariate credibility*. This chapter follows logically from Chapter 3 by extending the methodology to make use of the fact that some customers of financial institutions have multiple products with the provider even before being approached for a cross-sale attempt. The paper presents a solution to this problem, by generalising the methodology from Chapter 3 and using information about $\hat{\lambda}_i$ and n_i from all the existing products of a specific customer. The data validation study is performed in a similar way as in Chapter 3 however the results are presented somewhat differently to underline the practical implications of using the methodology. It is concluded that the segmentation is improved by using information from multiple products in comparison to using information from only one and that it is easier to identify a small subset to avoid to contact (due to poor profitability) than a small subset to approach.

Chapter 5 presents the paper *Optimal customer selection for cross-selling of financial services products*. In this chapter, a new model is presented which combines the model from Chapter 3 with a model for the probability of a customer purchasing an offered product. With this model, the financial institution is able to identify an optimal subset of customers to approach, taking both the estimated probability of purchase and the estimated risk profile (with respect to the offered product) into account. Additionally, in order to replicate a real cross-sale campaign, each cross-sale approach is associated with a cost, for the financial institution, representing costs of staffing and administrative expenses related to running the campaign. The stochastic profit variable, from each customer with respect to the cross-sale product, is thoroughly described by deriving distributional properties of the variable. Also distributional properties for the sum of the stochastic profits, from a subset of the customers, are derived. In the data study, the model is tested on real data from a previously run cross-sale campaign, which was conducted using the methodology of segmenting customers based only on the estimated probability of purchase. It is shown that by segmenting the customers based on the expected profit, rather than by only the estimated sales probability, the profit from the cross-sale campaign, realised by the financial institution, would double.

2

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.

2.1 Introduction

This chapter is a revised version of the paper Englund et al. (2008).
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, Guillén, Gustafsson, Nielsen and Nielsen (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), Zehnwirth (1985), see also Halliwell (1996), 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 (1975a; 1975b) 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 Section 2.2 we state the credibility model and the estimators in our multidimensional setup. The model is generalized in Section 2.3 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 Section 2.4, from which we, in Section 2.5, draw conclusions concerning the predictability of the credibility estimators.

2.2 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.

2.2.1 The multidimensional credibility model

The number of insurance claims N_{ijk} is a stochastic variable and we have an a priori expected number of claims λ_{ijk} for individual $i \in (1, \ldots, I)$, calendar year $j \in (1, \ldots, J)$ and coverage $k \in (1, \ldots, K)$. A dot, '$\cdot$', indicates summation over that index. We define the standardized number of claims $\mathbf{F}_{ij} = \begin{bmatrix} N_{ij1}\lambda_{ij1}^{-1}, & N_{ij2}\lambda_{ij2}^{-1}, & \ldots, & N_{ijK}\lambda_{ijK}^{-1} \end{bmatrix}^T$.

Model assumptions

(*i*) *Given the vector of individual risk parameters* $\mathbf{\Theta}_i$, *all random variables* N_{ijk}, *representing the number of insurance claims, are independent and Poisson-distributed with expected value* $\mathbb{E}\left[\mathbf{N}_{ij} \mid \mathbf{\Theta}_i\right] = \begin{bmatrix} \lambda_{ij1}\Theta_{i1}, & \lambda_{ij2}\Theta_{i2}, & \ldots, & \lambda_{ijK}\Theta_{iK} \end{bmatrix}^T$. *The* *expected value and (co-) variance of* $\mathbf{F}_{ij}$*, conditional on* $\mathbf{\Theta}_i$ *are:*

$$\begin{aligned}
\mathbb{E}\left[\mathbf{F}_{ij} \mid \mathbf{\Theta}_i\right] &= \mathbf{\Theta}_i \\
Cov\left[\mathbf{F}_{ij}, \mathbf{F}_{ij}^T \mid \mathbf{\Theta}_i\right] &= \mathbf{S}_{ij}\left(\mathbf{\Theta}_i\right), \text{ with diagonal elements } Var\left[F_{ijk} \mid \Theta_{ik}\right] = \frac{\sigma_k^2\left(\Theta_{ik}\right)}{\lambda_{ijk}} \\
Cov\left[\mathbf{\Theta}_i, \mathbf{\Theta}_i^T\right] &= \mathbf{T}
\end{aligned}$$

(*ii*) *The pairs* $(\mathbf{\Theta}_1, \mathbf{N}_{1j}), (\mathbf{\Theta}_2, \mathbf{N}_{2j}), ..., (\mathbf{\Theta}_I, \mathbf{N}_{Ij})$ *are independent, and* $\mathbf{\Theta}_1, \mathbf{\Theta}_2, ..., \mathbf{\Theta}_I$ *are independent and identically distributed with*
$\mathbb{E}\left[\mathbf{\Theta}_i\right] = \left[\ \mathbb{E}\left[\Theta_{01}\right], \quad \mathbb{E}\left[\Theta_{02}\right], \quad \ldots, \quad \mathbb{E}\left[\Theta_{0K}\right]\ \right]^T = \theta_0.$

The notation $\mathbf{\Theta}_i$ and $\mathbf{N}_{ij}$ represents vectors of length K, containing the individual risk parameters Θ_{ik} and the individual number of insurance claims N_{ijk}, with $k \in (1, \ldots, K)$, respectively. Furthermore we define $\mathbf{S}_{ij} = \mathbb{E}\left[\mathbf{S}_{ij}\left(\mathbf{\Theta}_i\right)\right]$ and $\mathbf{S}_i = \mathbb{E}\left[\mathbf{S}_{i\cdot}\left(\mathbf{\Theta}_i\right)\right]$.

2.2.2 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 $\mathbf{\Theta}_i$, given the observed experience, is:

$$\hat{\mathbf{\Theta}}_i = \mathbb{E}\left[\mathbf{\Theta}_i\right] + Cov\left[\mathbf{\Theta}_i, \mathbf{F}_{i\cdot}\right] Cov\left[\mathbf{F}_{i\cdot}, \mathbf{F}_{i\cdot}^T\right]^{-1}\left(\mathbf{F}_{i\cdot} - \mathbb{E}\left[\mathbf{F}_{i\cdot}\right]\right) \tag{2.1}$$

where

$$\begin{aligned}
Cov\left[\mathbf{\Theta}_i, \mathbf{F}_{i\cdot}\right] &= \mathbb{E}\left[Cov\left[\mathbf{\Theta}_i, \mathbf{F}_{i\cdot}\right] \mid \mathbf{\Theta}_i\right] + Cov\left[\mathbb{E}\left[\mathbf{\Theta}_i \mid \mathbf{\Theta}_i\right], \mathbb{E}\left[\mathbf{F}_{i\cdot} \mid \mathbf{\Theta}_i\right]\right] \\
&= 0 + Cov\left[\mathbf{\Theta}_i, \mathbf{\Theta}_i^T\right] = \mathbf{T}
\end{aligned}$$

and

$$\begin{aligned}
Cov\left[\mathbf{F}_{i\cdot}, \mathbf{F}_{i\cdot}^T\right] &= \mathbb{E}\left[Cov\left[\mathbf{F}_{i\cdot}, \mathbf{F}_{i\cdot}^T\right] \mid \mathbf{\Theta}_i\right] + Cov\left[\mathbb{E}\left[\mathbf{F}_{i\cdot} \mid \mathbf{\Theta}_i\right], \mathbb{E}\left[\mathbf{F}_{i\cdot} \mid \mathbf{\Theta}_i\right]^T\right] \\
&= \mathbb{E}\left[\mathbf{S}_{i\cdot}\left(\mathbf{\Theta}_i\right)\right] + Cov\left[\mathbf{\Theta}_i, \mathbf{\Theta}_i^T\right] = \mathbf{S}\mathbf{W}_i^{-1} + \mathbf{T}
\end{aligned}$$

where we have used the notation $\mathbf{S}_i = \mathbf{S}\mathbf{W}_i^{-1}$. The vector of standardized number of claims is $\mathbf{F}_{i\cdot} = \left[\ F_{i\cdot 1}, \quad F_{i\cdot 2}, \quad \cdots, \quad F_{i\cdot K}\ \right]^T$, where $F_{i\cdot k} = (\sum\limits_{j=1}^{J} \lambda_{ijk})^{-1} \sum\limits_{j=1}^{J} \lambda_{ijk} F_{ijk}$. The weight matrix $\mathbf{W}_i$ is a diagonal matrix with $\lambda_{i\cdot k} = \sum_{j=1}^{J} \lambda_{ijk}$ in the k^{th} diagonal element. The credibility weight α_i takes the following expression, $\widehat{\alpha}_i = \mathbf{T}(\mathbf{S}\mathbf{W}_i^{-1} + \mathbf{T})^{-1} = \mathbf{T}\mathbf{W}_i(\mathbf{T}\mathbf{W}_i + \mathbf{S})^{-1}$, where the diagonal matrix $\mathbf{S}$ contains elements $\sigma_k^2 = \mathbb{E}\left[\sigma_k^2(\Theta_{ik})\right]$. The expression for the credibility weight makes us 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:

$$\begin{aligned} \hat{\boldsymbol{\Theta}}_i &= \widehat{\alpha}_i \mathbf{F}_{i\cdot} + (\mathbf{I} - \widehat{\alpha}_i)\,\theta_0 \\ &= \theta_0 + \widehat{\alpha}_i\,(\mathbf{F}_{i\cdot} - \theta_0) \end{aligned} \tag{2.2}$$

$\mathbf{I}$ is the identity matrix. Note that the one-dimensional credibility estimator is a trivial special case of (2.2).

2.2.3 Estimation of the parameters

The estimation of the parameters θ_0, $\mathbf{T}$ and $\mathbf{S}$ is inspired by the elegant estimators presented by Bühlmann and Gisler (2005) for a different but related multivariate credibility problem. We estimate the elements of θ_0 by $\hat{\theta}_{0k} = (\sum_{i=1}^{I} \widehat{\alpha}_{ik})^{-1} \sum_{i=1}^{I} \widehat{\alpha}_{ik} F_{i\cdot k}$. The estimator of $\mathbf{S}$ is a diagonal matrix with $\hat{\sigma}_k^2$ as k^{th} diagonal:

$$\hat{\sigma}_k^2 = \frac{1}{J_{\cdot k} - I_k} \sum_{i=1}^{I} \sum_{j=1}^{J} \lambda_{ijk} \left(F_{ijk} - F_{i\cdot k}\right)^2.$$

$J_{\cdot k}$ 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 $\sigma_k^2 = \theta_{0k}$. The estimation of θ_{0k} follows the estimation of σ_k^2, due to the estimation of the credibility weights, wherefore we use two different estimators for this. The preliminary estimator of $\mathbf{T}$, $\widetilde{\mathbf{T}}$, has diagonal

$$\widetilde{\tau}_{kk}^2 = c_k \left(\frac{1}{I_k - 1} \sum_{i=1}^{I} \frac{\lambda_{i\cdot k}}{\lambda_{\cdot\cdot k}} \left(F_{i\cdot k} - F_{\cdot\cdot k}\right)^2 - \frac{\sigma_k^2}{\lambda_{\cdot\cdot k}} \right),\ F_{\cdot\cdot k} = \left(\sum_{i=1}^{I} \lambda_{i\cdot k} \right)^{-1} \sum_{i=1}^{I} \lambda_{i\cdot k} F_{i\cdot k}$$

and for $k \neq k'$:

$$\widetilde{\tau}^2_{kk'} = \frac{c_k}{I_{kk'} - 1} \sum_{i=1}^{I} \frac{\lambda_{i \cdot k}}{\lambda_{\cdot \cdot k}} \left(F_{i \cdot k} - F_{\cdot \cdot k} \right) \left(F_{i \cdot k'} - F_{\cdot \cdot k'} \right).$$

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

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

$$\overline{\tau}^2_{kk'} = signum \left(\frac{\widetilde{\tau}^2_{kk'} + \widetilde{\tau}^2_{k'k}}{2} \right) \min \left(\frac{\left| \widetilde{\tau}^2_{kk'} + \widetilde{\tau}^2_{k'k} \right|}{2}, \sqrt{\widetilde{\tau}^2_{kk} \widetilde{\tau}^2_{k'k'}} \right) \tag{2.3}$$

Even the corrected estimator $\overline{\mathbf{T}}$ resulting from (2.3) 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 $\overline{\mathbf{T}}$, put a floor of zero on the diagonal eigenvalue matrix, and compose the new eigenvalue matrix and we reconstruct the final estimator $\hat{\mathbf{T}}$. This estimator is more robust to calculate than the original estimator of Englund et al. (2008). The only parameter left to estimate now is λ_{ijk}, which we assume to be estimated from some pricing model developed by the company. This can be done at various levels of sophistication.

2.3 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 $\boldsymbol{\Theta}_i$, we now suppose that the risk characteristics of year s are given by an unknown parameter $\boldsymbol{\Theta}_{is}$, and that the dependence between $\boldsymbol{\Theta}_{is}$ and $\boldsymbol{\Theta}_{it}$ 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).

2.3.1 The time-dependent model and estimator

As a model for the time-dependent credibility estimator we generalize the model stated in Section 2.2.1 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 Θ_{ijk}, with expected value $\mathbb{E}\left[N_{ijk} \mid \Theta_{ijk}\right] = \lambda_{ijk}\Theta_{ijk}$ and the process $\{\Theta_{ijk}\}_{j=1,\dots,J}$ to be an auto-regressive process with lag 1, with the covariance structure: $Cov\left[\Theta_{ijk}, \Theta^T_{i'j'k'}\right] = \tau^2_{kk'}\left(\rho_{kk'}\right)^{|j-j'|}$ if $i = i'$ and $Cov\left[\Theta_{ijk}, \Theta^T_{i'j'k'}\right] = 0$ otherwise, with $\tau^2_{kk'} = \tau^2_{k'k}$ and $|\rho_{kk'}| \leq 1$. Correspondingly to Section 2.2.1 we further assume that $\mathbb{E}\left[\mathbf{F}_{ij} \mid \mathbf{\Theta}_{ij}\right] = \mathbf{\Theta}_{ij}$, $\mathbb{E}\left[\mathbf{\Theta}_{ij}\right] = \theta_0$ and $Cov\left[\mathbf{F}_{ij}, \mathbf{F}^T_{ij} \mid \mathbf{\Theta}_{ij}\right] = \mathbf{S}_{ij}\left(\mathbf{\Theta}_{ij}\right)$ with diagonal elements $Var\left[F_{ijk} \mid \Theta_{ijk}\right] = \frac{\sigma^2_k(\Theta_{ijk})}{\lambda_{ijk}}$.

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

$$\hat{\mathbf{\Theta}}_{i(j+1)} = \mathbf{a}\left(\hat{\alpha}_{ij}\mathbf{F}_{ij} + \left(\mathbf{I} - \hat{\alpha}_{ij}\right)\hat{\mathbf{\Theta}}_{ij}\right) + \left(\mathbf{I} - \mathbf{a}\right)\theta_0 \tag{2.4}$$

where $\mathbf{F}_{ij} = \begin{bmatrix} F_{ij1}, & F_{ij2}, & \cdots, & F_{ijK} \end{bmatrix}^T$, as previously noted. $\mathbf{a}$ is a diagonal matrix of elements in $[0, 1)$ driving the stationary auto-regressive processes. The credibility weight is $\hat{\alpha}_{ij} = \mathbf{T}_{ij}\mathbf{W}_{ij}\left(\mathbf{T}_{ij}\mathbf{W}_{ij} + \mathbf{S}\right)^{-1}$. Similarly to (2.2) the matrix $\mathbf{S}$ is diagonal with elements $\sigma^2_k = \mathbb{E}\left[\sigma^2_k\left(\Theta_{ijk}\right)\right]$ and the matrix $\mathbf{W}_{ij}$ is diagonal with λ_{ijk}, in the k^{th} diagonal element. The updating matrix $\mathbf{T}_{i(j+1)}$ is

$$\mathbf{T}_{i(j+1)} = \mathbf{a}^2\left(\mathbf{I} - \hat{\alpha}_{ij}\right)\mathbf{T}_{ij} + \left(\mathbf{I} - \mathbf{a}^2\right)\mathbf{T}_1$$

which is the multivariate equivalent of (18) in Sundt (1981). The starting values for the recursion are $\hat{\mathbf{\Theta}}_{i1} = \hat{\mathbf{\Theta}}_0$ and $\mathbf{T}_1 = \hat{\mathbf{T}}$, as in Section 2.2.2. The elements in $\mathbf{a}$ are estimated by performing a minimization of the residual sum of squares (see Subsection 2.4.2) based on the one-dimensional time-dependent credibility estimator for each coverage. Note that the estimator deals with individual missing values automatically.

2.4 An Application to Danish Commercial Business Lines

In this section we present an extensive study of a 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 practise. 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.

2.4.1 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 ω_{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 from a collateral data set of the same insurance company. We have up to eight years of information. Only few policyholders have eight years of information for all four coverages. In Table 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.

Table 1

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

Insurance coverage	**Total number of expected claims**	**Total number of reported claims**
Building - Fire	809	787
Building - Glass	7455	7797
Building - Other	1980	2004
Building - Water	2763	2731

2.4.2 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 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 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 $\hat{N}_{ijk} = \widehat{\bar{\lambda}}_k$, $\hat{N}_{ijk} = \hat{\lambda}_{ijk}$ or $\hat{N}_{ijk} = \hat{\lambda}_{ijk}\hat{\theta}_{ik}$. The estimate $\widehat{\bar{\lambda}}_k$ is calculated with the estimator
$\bar{\lambda}_k = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ijk}(\sum_{i=1}^{I}\sum_{j=1}^{J} \omega_{ijk})^{-1}$, 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. I.e. we use as estimation data all customers' individual claims experience in all but their last year of policy duration. All customers' last year of policy duration then forms the validation data set, hence we are performing an out-of-sample validation study. The results are presented in Figure 2.1. 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 2 we present a sheet to make he interpretation of Figure 2.1 easier.

Table 3 contains the resulting SS values represented in Figure 2.1.

The numbers on the horizontal axis of the subfigures in Figure 2.1 are keyed to Table 2.

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

Table 2

Table to help interpret Figure 2.1. MVE stands for mean value estimator, Present for the present estimator, (t)CE for credibility estimators with and without time dependence. The coverage names in the parentheses show the additional coverages used.

x-axis	Building - Fire	Building - Glass	Building - Other	Building - Water
1	MVE	MVE	MVE	MVE
2	Present	Present	Present	Present
3	1d CE	1d CE	1d CE	1d CE
4	2d CE (Glass)	2d CE (Fire)	2d CE (Fire)	2d CE (Fire)
5	2d CE (Other)	2d CE (Other)	2d CE (Glass)	2d CE (Glass)
6	2d CE (Water)	2d CE (Water)	2d CE (Water)	2d CE (Other)
7	3d CE (Glass, Other)	3d CE (Fire, Other)	3d CE (Fire, Glass)	3d CE (Fire, Glass)
8	3d CE (Glass, Water)	3d CE (Fire, Water)	3d CE (Fire, Water)	3d CE (Fire, Other)
9	3d CE (Other, Water)	3d CE (Other, Water)	3d CE (Glass, Water)	3d CE (Glass, Other)
10	4d CE (all)	4d CE (all)	4d CE (all)	4d CE (all)
11	1d tCE	1d tCE	1d tCE	1d tCE
12	2d tCE (Glass)	2d tCE (Fire)	2d tCE (Fire)	2d tCE (Fire)
13	2d tCE (Other)	2d tCE (Other)	2d tCE (Glass)	2d tCE (Glass)
14	2d tCE (Water)	2d tCE (Water)	2d tCE (Water)	2d tCE (Other)
15	3d tCE (Glass, Other)	3d tCE (Fire, Other)	3d tCE (Fire, Glass)	3d tCE (Fire, Glass)
16	3d tCE (Glass, Water)	3d tCE (Fire, Water)	3d tCE (Fire, Water)	3d tCE (Fire, Other)
17	3d tCE (Other, Water)	3d tCE (Other, Water)	3d tCE (Glass, Water)	3d tCE (Glass, Other)
18	4d tCE (all)	4d tCE (all)	4d tCE (all)	4d tCE (all)

Table 3

Table showing the resulting SS values of the validation analysis. The values are represented in Figure 2.1.

Estimator	1	2	3	4	5	6	7	8	9
Building - Fire	138.97	127.38	124.96	125.02	124.91	124.45	124.93	124.37	124.51
Building - Glass	2637.1	2061.2	1640.2	1641.7	1629.6	1648.8	1630.9	1649.7	1639.1
Building - Other	830.12	744.38	637.94	634.70	635.61	637.86	632.83	634.87	636.18
Building - Water	376.82	313.42	304.38	304.07	304.12	304.38	303.81	304.04	304.07
	10	11	12	13	14	15	16	17	18
Building - Fire	124.39	125.02	125.08	124.96	124.65	124.99	124.56	124.71	124.59
Building - Glass	1640.0	2116.7	1715.6	1704.3	1718.2	1703.5	1716.7	1705.3	1704.7
Building - Other	633.54	674.46	664.46	663.24	667.76	660.48	664.66	663.69	661.05
Building - Water	303.70	303.30	301.28	301.68	301.71	301.28	301.28	301.74	301.31

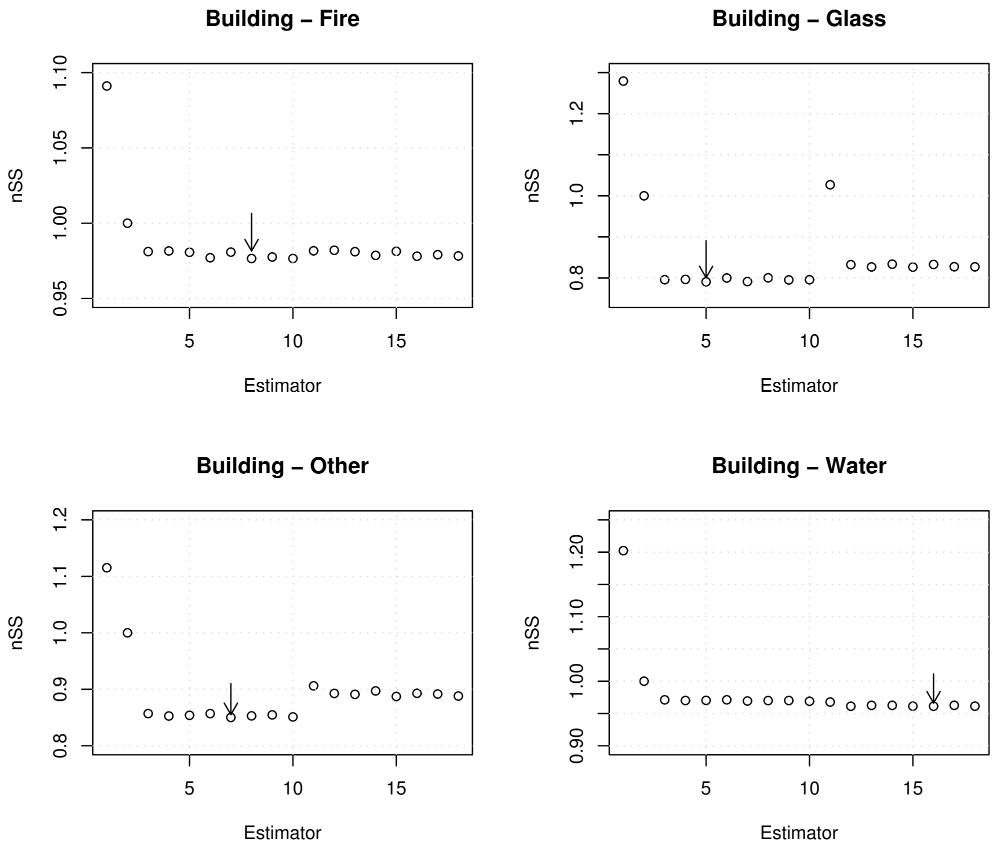

Figure 2.1: The normalized Sum of Squares (nSS) values for the different estimators.

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

According to Figure 2.1 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 2.1. The best estimators for the four different coverages are:

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.

2.5 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. 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.

References

Bühlmann, H., 1967, Experience rating and credibility, *Astin Bulletin*, 4(3), 199-207.

Bühlmann, H. and A. Gisler, 2005, *A Course in Credibility Theory and its Applications* (Berlin, Germany: Springer Verlag).

Bühlmann, H. and E. Straub, 1970, Glaubwürdigkeit für Schadensätze, *Bulletin of Swiss Association of Actuaries*, 70(1), 111-133.

De Vylder, Fl., 1976, Geometrical credibility, *Scandinavian Actuarial Journal*, 121-149.

De Vylder, Fl., 1977, Iterative credibility, *Bulletin of Swiss Association of Actuaries*, 77, 25-33.

De Vylder, Fl., 1978, Parameter estimation in credibility theory,*Astin Bulletin*, 10(1), 99-112.

Englund, M., M. Guillén, J. Gustafsson, L. H. Nielsen, and J. P. Nielsen, 2008, Multivariate Latent Risk: A Credibility Approach, *Astin Bulletin*, 38(1), 137-146.

Englund, M., Gustafsson, J., Nielsen, J.P. & Thuring, F. (2009). Multidimensional Credibility with Time Effects - An Application to Commercial Business Lines, *The Journal of Risk and Insurance*, 76, 443-453.

Gerber, H. U. and D. A. Jones, 1975a, Credibility formulæ with geometric weights, *Transaction of the Society of Actuaries*, 27, 39-52.

Gerber, H. U. and D. A. Jones, 1975b, Credibility formulas of the updating type, in: P. M. Kahn, ed., *Credibility: Theory and applications* (New York: Academic Press).

Greig, R., 1999, Random Effects Linear Statistical Models and Bühlmann-Straub Credibility, *CAS Forum, Winter 1999*, 387-404.

Halliwell, L., 1996, Statistical Models and Credibility, *CAS Forum, Winter 1996*, 61-152.

Hachemeister, C. A., 1975, Credibility for regression models with an application to trend, in: P. M. Kahn, ed., *Credibility: Theory and applications* (New York: Academic Press).

Jewell, W. S., 1974, Exact multidimensional credibility, *Bulletin of Swiss Association of Actuaries*, 74, 193-214.

Kremer, E., 1982, Credibility for some evolutionary models, *Scandinavian Actuarial Journal*, 129-142.

Mehra, R. K., 1975, Credibility theory and Kalman filtering with extensions, *Report RM 75-64, International Institute for Applied Systems Analysis, Schloss Laxenburg, Austria.*

Mowbray, A. H., 1914, How extensive a payroll exposure is necessary to give a dependable pure premium?, *Proceedings of the Casualty Actuarial Society*, 1, 24-30

Sundt, B., 1979, A hierarchical credibility regression model, *Scandinavian Actuarial Journal*, 107-114.

Sundt, B., 1981, Recursive credibility estimation, *Scandinavian Actuarial Journal*, 3-22.

Whitney, A. W., 1918, The theory of experience rating, *Proceedings of the Casualty Actuarial Society*, 4, 274-292.

Zehnwirth, B., 1985, Linear filtering and recursive credibility estimation, *ASTIN Bulletin*, 15(1), 19-38.

3

A credibility method for profitable cross-selling of insurance products

Abstract
A method is presented for identifying an expected profitable set of customers, to offer them an additional insurance product, by estimating a customer specific latent risk profile, for the additional product, by using the customer specific available data for an existing insurance product of the specific customer. For the purpose, a multivariate credibility estimator is considered and we investigate the effect of assuming that one (of two) insurance products is inactive (without available claims information) when estimating the latent risk profile. Instead, available customer specific claims information from the active existing insurance product is used to estimate the risk profile and thereafter assess whether or not to include a specific customer in an expected profitable set of customers. The method is tested using a large real data set from a Danish insurance company and it is shown that sets of customers, with up to 36% less claims than a priori expected, are produced as a result of the method. It is therefore argued that the proposed method could be considered, by an insurance company, when cross-selling insurance products to existing customers.

3.1 Introduction

This chapter is a revised version of the paper Thuring (2012).
Many marketers of consumer products have noticed that, as part of their marketing campaign, they can offer insurance cover for their products for sale, such as free car insurance for a new car or specific insurance for a new piece of home electronics. Often the insurance cover is not provided by the marketers themselves but through a partnership agreement with an insurance provider, which may see this as one of its distribution channels. As a consequence, consumers will have different insurance cover for many of their products, most probably from different insurance providers, and except from losing the possibility to get a bundling discount, on insurances from the same insurer, the consumers will experience few negative effects with having multiple insurance providers. However, from an insurance company's point of view, providing only a single or few insurance products to a customer is seldom desirable since such customers are more likely to cancel their existing business with the company in favour of a competitor, see Kamakura et al. (2003) for a general discussion about cross-selling as a method for retaining customers and Brockett et al. (2008) for an overview on how much time is left to stop total customer defection. Hence, insurance companies would be interested in developing their sales methods for increasing the number of products for their existing customers.

Increasing the number of products of a company's existing customers is referred to as cross-selling. In most cases this means personal communication, often through call-centres, with the customers for which the expected demand for a certain product is high. In this paper it is argued that for some businesses, especially insurance business, there is an alternative to this sales driven cross-selling approach. Unlike conventional retail products, insurance products are associated with costs that are stochastic and determined at a stochastic time interval after a sale has been made. This stochasticity implies that, from an insurer's point of view, also the profitability for a certain customer is stochastic. However, the profitability might be predictable and hence reveal sets of customers which are preferable for the insurance company to extend the existing business with. This paper contributes with a method for such profitability predictions not found in either the marketing or the actuarial literature. In the data study in Section

3.4, it is shown that the proposed method produces sets of customers with up to 36% less claims than expected. While this is just one example for one particular data set it still suggests that the method would be useful in practice.

The marketing literature on cross-sale models focuses primarily on various ways to model the demand for a certain cross-sale product amongst a company's customers. Often different regression models are evaluated based on data for the sales response of past cross-sale attempts, where patterns of the customers with high demand is sought after. One of the first efforts to model a cross-sale opportunity formally is Kamakura et al. (1991), where a latent trait model is presented for the probability that a consumer would use a particular product or service, based on their ownership of other products or services. Another study is made by Knott et al. (2002) where a comparison is made of four different models for the probability of a successful cross-sale. Kamakura et al. (2003) discuss reasons why cross-selling is crucial for financial services (such as banks and insurance companies) and present a predictive model for whether or not customers satisfy their needs for financial services elsewhere. They argue that when a customer acquires more products or services from the same company, the switching cost of the customer increases and thereby minimises the risk of the customer leaving for a competitor. In Li et al. (2005) a natural ordering in which to present different products to a customer is investigated. They model the development over time for customer demand of multiple products and apply latent trait analysis to position financial services at correct time points within the customer lifetime.

The cross-sale method presented in this paper uses developments in multivariate credibility theory, for calculating a customer i's expected profitability of the cross-sale insurance product k with available data from another insurance product k'. The actuarial research branch of credibility theory investigates how collective and individual information should be weighted to produce a fair insurance premium for each individual. The literature on the subject is rich, dating back to early papers by Mowbray (1914) and Whitney (1918) which are the first studies of what later became known as credibility theory. Pioneering papers on credibility theory are Bühlmann (1967) and Bühlmann & Straub (1970) where in the latter paper the Bühlmann-Straub credibility estimator is derived. Credibility estimators and Bayesian statistics are investigated in e.g. Bailey (1950), Jewell (1974) and Gangopadhyay & Gau (2007). This paper uses developments of multivariate credibility found in Englund et al. (2008) and Englund et al. (2009), both papers model frequency of insurance claims from correlated business lines. Multivariate credibility models are also found in Venter (1985), describing multivariate credibility models in a hierarchical framework, and Jewel (1989), investigating multivariate predictions of first and second order moments in a credibility setting. More recent references are Frees (2003), who applies multivariate credibility models for predicting aggregate loss, and Bühlmann & Gisler (2005), which is one of the standard references in credibility theory.

The structure of the paper is as follows. In Section 3.2 the credibility model is described and the estimator is presented for the case of complete data for both products. In Section 3.3 the multivariate credibility estimator for cross-selling is presented for the case of unavailable information for the cross-sale product k. In Section 3.4 the cross-sale method is tested and analysed on a large data set from the personal lines of business of a Danish insurance company and concluding remarks are found in Section 3.5.

3.2 The credibility model and estimator

We use the model from Englund et al. (2008) and estimation following Englund et al. (2009). We consider insurance customers $i = 1, \ldots, I$ in time periods $j = 1, \ldots, J_i$ with insurance products k' and k, for convenience we will use the index $l \in k', k$ and $r \in k', k$ for insurance products in general. The insurance customer i is characterised by his/her individual risk profile θ_{il} which is a realisation of the independent and identically distributed random variable Θ_{il}, with $\mathbb{E}\left[\Theta_{il}\right] = \theta_{0l}$ and $Cov\left[\Theta_{il}, \Theta_{ir}\right] = \tau_{lr}^2$ with $l, r \in \{k', k\}$, θ_{0l} is often called the collective risk profile. The number of insurance claims N_{ijl} is assumed to be a Poisson distributed random variable with conditional expectation $\mathbb{E}\left[N_{ijl} \mid \Theta_{il}\right] = \lambda_{ijl}\Theta_{il}$ and the pairs $\left(\Theta_{1l}, N_{1jl}\right), \left(\Theta_{2l}, N_{2jl}\right), \ldots, \left(\Theta_{Il}, N_{Ijl}\right)$ are independent. We have a priori expected number of claims $\lambda_{ijl} = e_{ijl} g_l\left(\mathcal{Y}_{ijl}\right)$ of customer i in period j and product $l \in k', k$, which depends on the exposure e_{ijl}, a regression function g_l and of a set of explanatory tariff variables $\mathcal{Y}_{ijl}$ characterising the customer and the insured object. Note that g_l is common for all customers i and time periods j and is estimated based on collateral data from the insurance company. We assume that e_{ijl} can take values between $[0;1]$, where $e_{ijl}=0$ means that the l-th product is not active for customer i in time period j and correspondingly, $e_{ijl}=1$ means that the product l of customer i is active during the entire time period j. We define $F_{ijl}=\frac{N_{ijl}}{\lambda_{ijl}}$, which is a measure of the deviation between the a priori expected number of claims λ_{ijl} and the actual number of claims N_{ijl} . Further we assume that the insurance premium P_{ijl} is proportional to λ_{ijl} and that the claim severities $X_{ijl}^{(\nu)}$, $\nu=1,2,\ldots,N_{ijl}$ are independent and also independent of N_{ijl} with $\mathbb{E}\left[X_{ijl}^{(\nu)}\right]=h_l\left(\mathcal{X}_{ijl}\right)$. Analogous to the claim frequency, $\mathcal{X}_{ijl}$ is a set of explanatory variables and h_l a regression function. Note that $\mathbb{E}\left[F_{ijl}\mid\Theta_{il}\right]=\Theta_{il}$ and, under the stated assumptions, the lower the individual risk profile θ_{il} is, the higher the profitability is. We assume a conditional covariance structure of F_{ijl} as $Var\left[F_{ijl}\mid\Theta_{il}\right]=\frac{\sigma_l^2(\Theta_{il})}{\lambda_{ijl}}$ and $Cov\left[F_{ijk'},F_{ijk}\mid\Theta_{ik'},\Theta_{ik}\right]=0$, where $\sigma_l^2\left(\Theta_{il}\right)$ is the variance within an individual customer i, for $l\in k',k$. With $F_{i\cdot l}=\frac{\sum_{j=1}^{J_i}N_{ijl}}{\sum_{j=1}^{J_i}\lambda_{ijl}}$ and $\lambda_{i\cdot l}=\sum_{j=1}^{J_i}\lambda_{ijl}$ we get $Var\left[F_{i\cdot l}\mid\Theta_{il}\right]=\frac{\sigma_l^2(\Theta_{il})}{\lambda_{i\cdot l}}$ and $Cov\left[F_{i\cdot k'},F_{i\cdot k}\mid\Theta_{ik'},\Theta_{ik}\right]=0$. Since we consider the two-dimensional case with the specific insurance products k' and k, under the stated model assumptions, the multivariate credibility estimator of $\theta_i=[\theta_{ik'},\theta_{ik}]'$ is (see Englund et al., 2009 and Bühlmann & Gisler 2005, p. 181)

$$\hat{\theta}_i=\theta_0+\alpha_i\left(F_{i\cdot}-\theta_0\right) \tag{3.1}$$

with $\hat{\theta}_i=\left[\hat{\theta}_{ik'},\hat{\theta}_{ik}\right]'$, $\theta_0=\left[\theta_{0k'},\theta_{0k}\right]'$, $F_{i\cdot}=\left[F_{i\cdot k'},F_{i\cdot k}\right]'$ and $\alpha_i=\begin{pmatrix}\alpha_{ik'k'} & \alpha_{ik'k}\\ \alpha_{ikk'} & \alpha_{ikk}\end{pmatrix}$. The credibility weight $\alpha_i=T\Lambda_i(T\Lambda_i+S)^{-1}$ where $T=\begin{pmatrix}\tau_{k'k'}^2 & \tau_{k'k}^2\\ \tau_{kk'}^2 & \tau_{kk}^2\end{pmatrix}$, $\Lambda=\begin{pmatrix}\lambda_{i\cdot k'} & 0\\ 0 & \lambda_{i\cdot k}\end{pmatrix}$ and $S=\begin{pmatrix}\sigma_{k'}^2 & 0\\ 0 & \sigma_k^2\end{pmatrix}$, see Englund et al. (2009). The parameters $\sigma_{k'}^2$ and σ_k^2 are equal to $\mathbb{E}\left[\sigma_{k'}^2\left(\Theta_{ik'}\right)\right]$ and $\mathbb{E}\left[\sigma_k^2\left(\Theta_{ik}\right)\right]$, respectively. We are considering a homogeneous credibility estimator and we therefore need an estimator for the collective risk profiles $\theta_{0k'}$ and θ_{0k}. An unbiased estimator is found in Bühlmann & Gisler (2005 p. 183) as $\hat{\theta}_0=\left(\sum_{i=1}^{I}\alpha_i\right)^{-1}\sum_{i=1}^{I}\alpha_iF_{i\cdot}$. Performing the matrix multiplication in (3.1) gives the multivariate credibility estimator of θ_{ik}, for the specific product k, based on $F_{i\cdot k'}$ and $F_{i\cdot k}$ as,

$$\hat{\theta}_{ik}=\theta_{0k}+\alpha_{ikk'}\left(F_{i\cdot k'}-\theta_{0k'}\right)+\alpha_{ikk}\left(F_{i\cdot k}-\theta_{0k}\right). \tag{3.2}$$

For the estimation procedure of the parameter matrices S and T see e.g. Bühlmann & Gisler (2005) pp. 185-186 or Englund et al. (2009).

3.3 Cross-selling with the credibility estimator

We are interested in cross-selling an insurance product k to a set Φ of customers already having another insurance product k' from the insurance company. The hypothesis is that an estimator $\hat{\theta}_{ikk'}$, of the risk profile θ_{ik}, can be obtained, based only on the available data for $F_{i\cdot k'}$ with respect to the existing product k', and that a profitable cross-sale set Φ^* would consist of customers with as low $\hat{\theta}_{ikk'}$ as possible. Prior to cross-selling product k to the i-th customer, product k is inactive and no claims have been reported i.e. $n_{ijk} = 0$. Also the exposure $e_{ijk} = 0$, with respect to the cross-sale product k, which leads to $\alpha_{ikk} = 0$ and $\alpha_{ikk'} = \frac{\lambda_{i\cdot k'}\tau^2_{kk'}}{\lambda_{i\cdot k'}\tau^2_{k'k'}+\sigma^2_{k'}}$ in (3.2). The credibility estimator of θ_{ik}, based only on the available $F_{i\cdot k'}$, becomes

$$\hat{\theta}_{ikk'} = \theta_{0k} + \frac{\lambda_{i\cdot k'}\tau^2_{kk'}}{\lambda_{i\cdot k'}\tau^2_{k'k'} + \sigma^2_{k'}} \left(F_{i\cdot k'} - \theta_{0k'}\right). \tag{3.3}$$

Note that, in order to be able to evaluate (3.3), estimates of θ_{0k}, $\tau^2_{kk'}$, $\tau^2_{k'k'}$, $\sigma^2_{k'}$ and $\theta_{0k'}$ need to be obtained from a collateral data set consisting of customers with both products k and k' active and whose characteristics are as close to the characteristics of the customers, for which an estimate of the risk profile θ_{ik} is sought after.

Considering that $F_{i\cdot k'} = \frac{\sum_{j=1}^{J_i} N_{ijk'}}{\sum_{j=1}^{J_i} \lambda_{ijk'}}$, a customer i with $F_{i\cdot k'} < 1$ has reported fewer insurance claims than a priori expected (for the existing product k') and has therefore been a more profitable customer than a customer i' with $F_{i'\cdot k'} > 1$. Hence, from the insurance company's point of view, a customer i is preferred over a customer i' if $\hat{\theta}_{ikk'} < \hat{\theta}_{i'kk'}$ and the expected most profitable set Φ^* of size ϕ^* to cross-sale a product k to is the first ϕ^* customers when ordered by increasing $\hat{\theta}_{ikk'}$ as

$$\hat{\theta}_{(1)kk'} \leq \hat{\theta}_{(2)kk'} \leq \ldots \leq \hat{\theta}_{(\phi^*)kk'} \leq \ldots \leq \hat{\theta}_{(I)kk'}. \tag{3.4}$$

There are two ways to select the cross-sale set Φ^*. Either by setting ϕ^* to a predefined number of customers and using (3.4) to define Φ^* or by setting an upper limit θ_L for which all customers with $\hat{\theta}_{ikk'} \leq \theta_L$ are in Φ^*. A similar remark is made in Knott et al. (2002) regarding the probability of a successful cross-sale. We define the expected average profitability for a cross-sale set Φ as

$$\bar{\theta}_k(\Phi) = \frac{1}{\phi}\sum_{i\in\Phi}\hat{\theta}_{ikk'}. \tag{3.5}$$

We also define the corresponding observed value $\bar{F}_k(\Phi)$ as

$$\bar{F}_k(\Phi) = \frac{\sum\limits_{i\in\Phi}\alpha_{ikk}F_{i\cdot k}}{\sum\limits_{i\in\Phi}\alpha_{ikk}}, \tag{3.6}$$

where the weighting with α_{ikk} is needed for an unbiased comparison between $\bar{\theta}_k(\Phi)$ and $\bar{F}_k(\Phi)$. Please note that the notation Φ represents an arbitrary subset of customers and that Φ^* an optimal subset of arbitrary size.

3.4 Data study

We have a large data set available from the personal lines of business of a Danish insurance company consisting of number of claims n_{ijl} (assumed to be realisations of a Poisson distributed random variable $N_{ijl} \sim Po\left(\hat{\lambda}_{ijl}\Theta_{il}\right)$) and an estimate of the a priori expected number of claims $\hat{\lambda}_{ijl}$, for 3 different insurance products, $l \in \{1,2,3\}$ where $l = 1$ represents motor insurance, $l = 2$ represents building insurance and $l = 3$ represents content insurance . The estimate $\hat{\lambda}_{ijl}$ is received via a Poisson regression based on a collateral data set from the same company, which is not available to us. In the data set we have available, there are 95668 unique customers who all have active products $\{1,2,3\}$ during the J_i years of engagement with the company. The number J_i is individual, between 1 and 5, and each record is unique for customer i in time period j making the total number of records in the data set 306196. Figure 3.1 presents histograms of the a priori expected number of claims $\hat{\lambda}_{ijl}$ and observed number of claims n_{ijl}, for $l \in \{1,2,3\}$, in the data set.

Notice the large number of the records with 0 number of claims (lower row of graphs), which is in line with what can be expected from personal lines insurance business where claims are infrequent. This is also reflected in the rather low values of the a priori expected number of claims for each record (upper row of graphs).

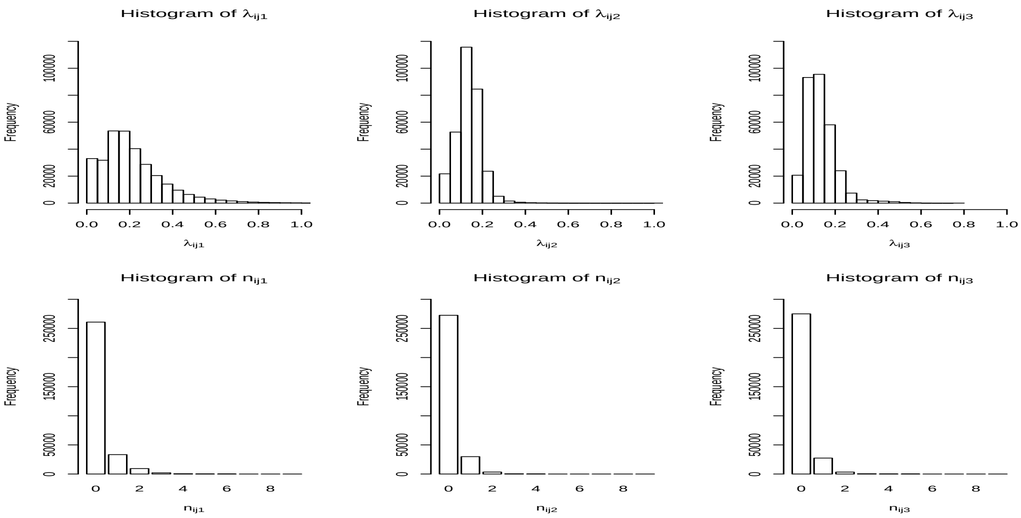

Figure 3.1: Histograms of the a priori expected number of claims $\hat{\lambda}_{ijl}$ and observed number of claims n_{ijl} , for the data set of Danish insurance customers.

We randomly divide the data set into one estimation data set (75% of the 95668 customers), for estimation of the model parameters as described in Bühlmann & Gisler (2005) pp. 185-186 or Englund et al. (2009), and one validation data set (the remaining 25% of the customers) . The estimates of the model parameters, obtained from the estimation data set, are found in Table 1.

Table 1
Estimates of the model parameters,
obtained from the estimation data set.

l	$\hat{\sigma}_l$	$\hat{\tau}_{l1}$	$\hat{\tau}_{l2}$	$\hat{\tau}_{l3}$	$\hat{\theta}_{0l}$
1	1.377	0.322	0.151	0.247	0.946
2	0.986	0.151	0.402	0.489	0.919
3	0.915	0.247	0.489	0.609	0.892

For the validation data set, we define the cross-sale product k to be any of $\{1, 2, 3\}$ and define the existing product k' to be any of the other products in $\{1, 2, 3\}$ with $k \neq k'$. Thereafter $\hat{\theta}_{ikk'}$ is estimated using (3.3), for every customer i in the validation data set, using the estimated model parameters in Table 1 and by using that $F_{i \cdot k'}$ has taken the observed individual values $F_{i \cdot k'} = \frac{\sum_{j=1}^{J_i} n_{ijk'}}{\sum_{j=1}^{J_i} \hat{\lambda}_{ijk'}}$, for every customer i in the validation data set, with respect to the existing product k'. Hence, we imagine only knowing about the a priori expected number of claims $\hat{\lambda}_{i\cdot k'}$ and the observed number of claims $n_{i\cdot k'}$, with respect to the existing product k', in the validation data set. This is a very realistic situation for an insurance company aiming at cross-selling the i-th customer another insurance product k, by estimating specific model parameters, using a collateral data set, and thereafter evaluating the model with available customer specific information in order to assess (in this case) the customer's individual risk profile.

Since we have the a priori expected number of claims $\hat{\lambda}_{i\cdot k}$ and the observed number of claims $n_{i\cdot k}$ available, for the customers in the validation data set, with respect to the cross-sale product k, we are able to evaluate if $\hat{\theta}_{ikk'}$ is a good estimator of the risk profile θ_{ik}. It should be noted that since we have claims information, $\hat{\lambda}_{i\cdot k}$ and $n_{i\cdot k}$ with respect to the cross-sale product k, for every customer i in the validation data set, our imaginary cross-sale campaign has resulted in every customer accepting the cross-sale offer. This is of course unlikely in practice, where normally as few as 1 of 10 approached customers accept a cross-sale offer, but this is assumed in order not to obstruct the study. A more realistic study would be to incorporate a (possibly generalised linear) model for the cross-sale probability and analyse observed data from a cross-sale campaign, but this is outside the scope of the paper.

As stated in Section 3.3, we aim at cross-selling a product k to an expected profitable subset Φ^* from a larger group of customers. We estimate θ_{ik} with $\hat{\theta}_{ikk'}$ using (3.3), for the customers in the validation data set, order these by increasing $\hat{\theta}_{ikk'}$ (see (3.4)) and divide them into a number of equally sized sets Φ_m, with $m = 1, \ldots, M$. We set $M = 10$ which gives

$$\begin{aligned}
\Phi_1 :\ & \hat{\theta}_{(1)kk'} \leq \hat{\theta}_{(2)kk'} \leq \ldots \leq \hat{\theta}_{(\phi_1)kk'} \\
\Phi_2 :\ & \hat{\theta}_{(\phi_1+1)kk'} \leq \hat{\theta}_{(\phi_1+2)kk'} \leq \ldots \leq \hat{\theta}_{(2\phi_1)kk'} \\
& \vdots \\
\Phi_{10} :\ & \hat{\theta}_{(9\phi_1+1)kk'} \leq \hat{\theta}_{(9\phi_1+2)kk'} \leq \ldots \leq \hat{\theta}_{(I)kk'}.
\end{aligned}$$

The size of each set Φ_m is $\phi_m = 2{,}382$ for $m = 1, \ldots, 10$ and $I = 23820$ is here the number of customers in the validation data set. It is obvious that $\bar{\theta}_k(\Phi_1) \leq \bar{\theta}_k(\Phi_2) \ldots \leq \bar{\theta}_k(\Phi_{10})$, see (3.5).

Figures 3.2-3.4 show $\bar{\theta}_k(\Phi_m)$ and $\bar{F}_k(\Phi_m)$ for the 10 sets $\Phi_1, \ldots, \Phi_{10}$ for $k = 1$, $k = 2$ and $k = 3$, respectively. The risk profile $\hat{\theta}_{ikk'}$ is estimated using the model parameter estimates obtained from the estimation data set (see Table 1) and the available customer specific information about $\hat{\lambda}_{i\cdot k'}$ and $n_{i\cdot k'}$ for customer i in the validation data set, with respect to one of the other products $k' \in \{1,2,3\}$ with $k \neq k'$. From figures 3.2-3.4, it should be noted that the observed average profitability $\bar{F}_k(\Phi_m)$ follows the expected average profitability $\bar{\theta}_k(\Phi_m)$ nicely, which suggest that the estimator in (3.3) produces estimates of the risk profile θ_{ik} close to the actual values.

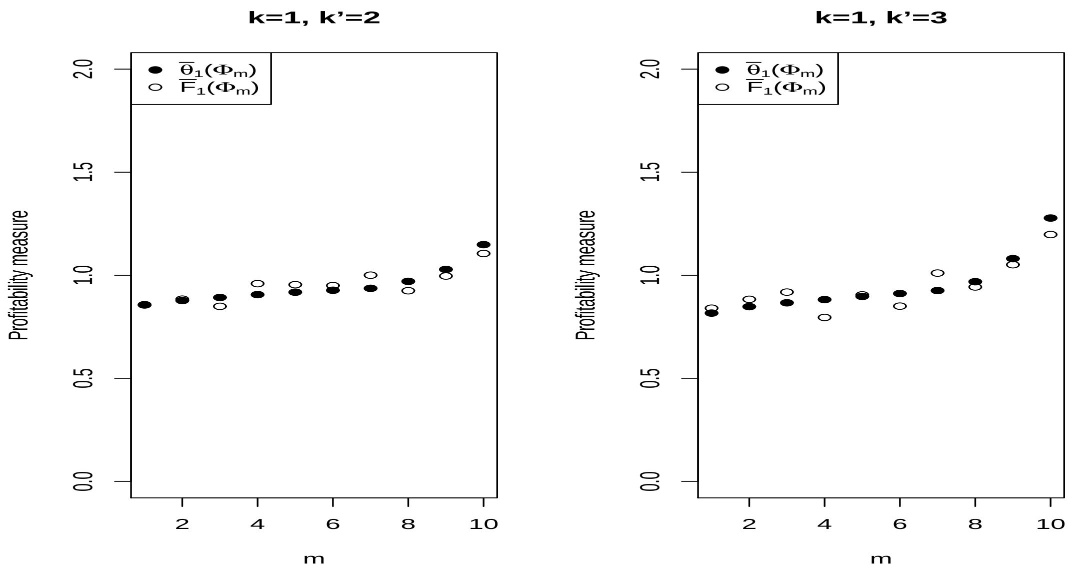

Figure 3.2: Expected (filled dots) and observed (circles) average profitability for cross-selling of product $k = 1$,
based on data from either product $k' = 2$ or product $k' = 3$.

As can be seen, from figures 3.2-3.4, different product combinations (of the cross-sale product k and the existing product k') produces slightly different shapes of the expected average profitability as a function of the different subsets Φ_m, $m = 1, \ldots 10$. Comparing e.g. left and right hand sub-plot of figure 3.4, there is a larger spread between $\bar{\theta}_3(\Phi_1)$ and $\bar{\theta}_3(\Phi_{10})$ for the product combination $k = 3$ and $k' = 2$ than for product combination $k = 3$ and $k' = 1$. This indicates that the estimator $\hat{\theta}_{i32}$ (based on available data from product $k' = 2$) differentiates between expected profitable and expected unprofitable subsets Φ_m in a more effective way than the estimator $\hat{\theta}_{i31}$ (based on available data from product $k' = 1$). Hence, available claims information for product $k' = 2$ should be preferred over corresponding information from product $k' = 1$, when

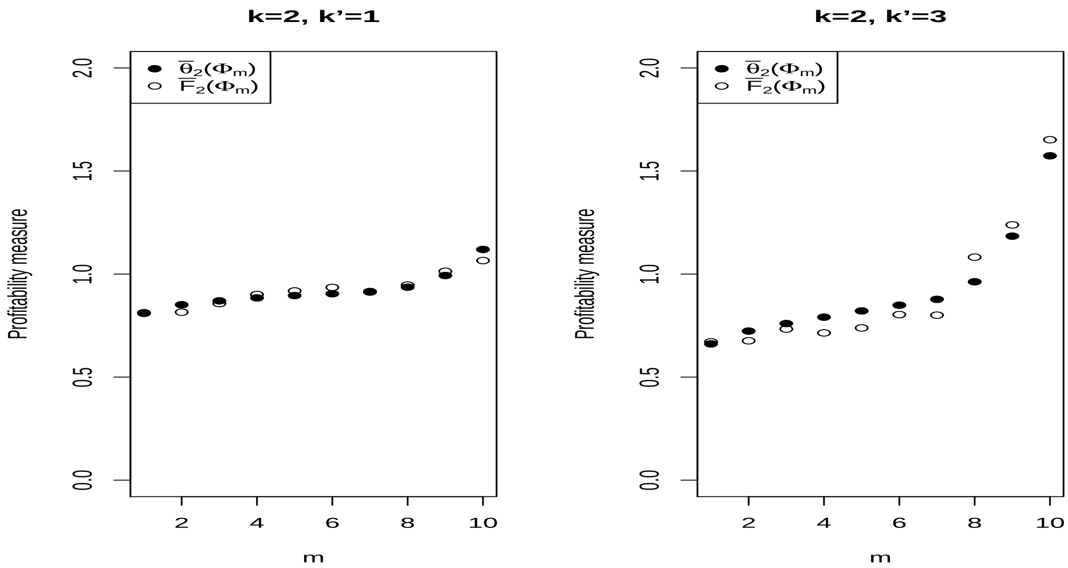

Figure 3.3: Expected (filled dots) and observed (circles) average profitability for cross-selling of product $k = 2$,
based on data from either product $k' = 1$ or product $k' = 3$.

selecting customers to cross-sale product $k = 3$ to. Similar comparisons can be made with respect to figure 3.2 and figure 3.3.

A realistic situation, for an insurance company aiming at cross-selling a product k to its existing customers with a product k', is to define a maximum number of customers to approach. The reason being limited resources for interacting with the customers, e.g. limited number of employees in the call centre. We replicate this situation by assuming that the maximum number of customers, which the insurance company has resources to approach, is $\phi_m = 2382$ (i.e. the size of one of the subsets Φ_m) and the company should select the expected most profitable customers from its portfolio of existing customers. We assume that the portfolio of existing customers is the validation data set where individual claims information from the cross-sale product k is imagined unavailable. Since $\bar{\theta}_k(\Phi_1) \leq \bar{\theta}_k(\Phi_2) \ldots \leq \bar{\theta}_k(\Phi_{10})$, the expected most profitable set of customers to approach is Φ_1. The expected average profitability $\bar{\theta}_k(\Phi_1)$ as well as the observed average profitability $\bar{F}_k(\Phi_1)$, for all combinations of k and k', is shown in Table 2. We assume that all of the 2382 imagined contacted customers accepted the offer of purchasing the cross-sale product k.

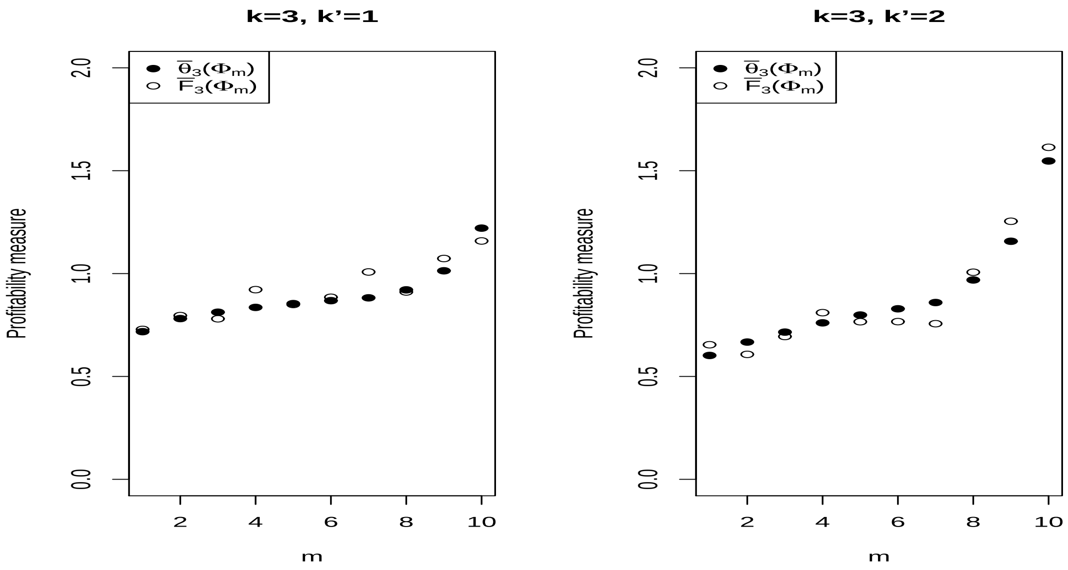

Figure 3.4: Expected (filled dots) and observed (circles) average profitability for cross-selling of product $k = 3$,
based on data from either product $k' = 1$ or product $k' = 2$.

Table 2
Expected $\bar{\theta}_k(\Phi_1)$ and observed $\bar{F}_k(\Phi_1)$ average profitability in the expected most profitable cross-sale set Φ_1. Also the corresponding minimum $\hat{\theta}_{(1)kk'}$ and maximum $\hat{\theta}_{(\phi_1)kk'}$ values of estimated risk profiles are shown.

k	k'	$\hat{\theta}_{(1)kk'}$	$\hat{\theta}_{(\phi_1)kk'}$	$\bar{\theta}_k(\Phi_1)$	$\bar{F}_k(\Phi_1)$
1	2	0.79	0.87	0.86	0.85
1	3	0.72	0.84	0.82	0.84
2	1	0.70	0.84	0.82	0.81
2	3	0.48	0.70	0.66	0.67
3	1	0.54	0.76	0.73	0.73
3	2	0.39	0.64	0.60	0.65

From Table 2, the observed average profitability $\bar{F}_k(\Phi_1)$ for the product combination ($k = 2$, $k' = 3$) and ($k = 3$, $k' = 2$) deserves special attention. For $k = 2$ and $k' = 3$ an average profitability of $\bar{F}_k(\Phi_1) = 0.67$ is observed, interpreted as this set consists of customers with on average 33% less observed claims than a priori expected. The corresponding situation for $k = 3$ and $k' = 2$ results in a set of customers with on average 36% less observed claims than a priori expected. This indicates not only that profitable selections are available but also that the correlation in claim occurrence is relatively high between the building product ($k = 2$) and the content product ($k = 3$), i.e. customers with reported number of claims lower than a priori expected for one of the products, suggests that a similar pattern can be expected with respect to the other. The smallest effect is shown for product $k = 1$, the set Φ_1 consists of customers with on average between 15% and 16% less observed claims than expected.

3.5 Concluding remarks

This paper presents a method for identifying an expected profitable set of customers Φ^*, to cross-sell to them an insurance product k, by estimating a customer specific latent risk profile θ_{ik} using the customer specific available data for another insurance product k'. For the purpose, we consider a multivariate credibility estimator found in Englund et al. (2009) and investigate the effect of assuming that one (of two) insurance products is inactive (without available claims information) when estimating the latent risk profile θ_{ik}. We also recognise that in order to estimate θ_{ik}, estimates of certain model parameters have to be obtained from collateral data consisting of customers with both products k and k' active, and whose characteristics are close to the characteristics of the customers for which an estimate of θ_{ik} is sought after.

In Section 3.4 we have tested the proposed cross-sale method with a large data set from a Danish insurance company consisting of personal lines customers with 3 active insurance products (at the time of data collection). The data set is randomly divided into two data sets, where estimates of the model parameters are obtained from the estimation data set and a customer specific latent risk profile θ_{ik} is estimated for every customer in the other validation data set. The estimate of the latent risk profile $\hat{\theta}_{ikk'}$, for the cross-sale product $k \in \{1, 2, 3\}$, is obtained with the model parameters from the estimation data set and the available, customer specific, information about $\hat{\lambda}_{i \cdot k'}$ and $n_{i \cdot k'}$, in the validation data set, with respect to the existing product $k' \in \{1, 2, 3\}$, with $k \neq k'$.

The observed average profitability $\bar{F}_k(\Phi)$ for a set Φ of customers is close to the expected average profitability $\bar{\theta}_k(\Phi)$, with only few exceptions, as seen in figures 3.2, 3.3 and 3.4, which suggest that the estimators $\hat{\theta}_{ikk'}$ ($k \in \{1, 2, 3\}$ and $k' \in$

$\{1, 2, 3\}$, with $k \neq k'$) would be useful in practice. However, the validation is performed using only a single data set and the method might give other results for other data sets, especially if the correlation in claim occurrence (between insurance products) is low. For the analysed Danish insurance data set, there are combinations of cross-sale product k and existing product k' which perform better than others, especially the combinations ($k = 2$, $k' = 3$) and ($k = 3$, $k' = 2$) produce very profitable cross-sale selections with an observed average profitability as low as 0.64, which is interpreted as 36% less reported claims than a priori expected. This indicates a strong correlation between building claims ($k = 2$) and content claims ($k = 3$) and it is argued that an insurance company would be interested in directing cross-sale efforts towards customers with high profitability for one of these products but lacking the other one. Even though the effect is smaller when cross-selling motor insurance ($k = 1$), figure 3.2 and Table 2 show that the proposed cross-sale method is able to identify a set of customers with 16% less claims than expected, which for a large insurance company translates into a considerable profit increase. The insurance company might also consider offering discounted premiums, on the cross-sale product k, to the customers in a set Φ^*, to increase sales volume. In this case $1 - \hat{\theta}_{ikk'}$, for the customers $i \in \Phi^*$, can be used as a limit for how large the discount, for a specific customer i, is allowed to be.

References

Bailey, A.L. (1950). Credibility Procedures, LaPlace's Generalization of Bayes' Rule, and the Combination of Collateral Knowledge with Observed Data, *Proceedings of the Casualty Actuarial Society*, 37, 7-23.

Brockett, P.L., Golden, L.L., Guillén, M., Nielsen, J.P., Parner, J., Perez-Marin, A.M. (2008). Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection?, *The Journal of Risk and Insurance*, 75, 713-737.

Bühlmann, H. (1967). Experience rating and credibility, *Astin Bulletin*, 4, 199-207.

Bühlmann, H. & Gisler, A. (2005). *A Course in Credibility Theory and its Applications* (Berlin, Germany: Springer Verlag).

Bühlmann, H. & Straub, E. (1970). Glaubwürdigkeit für Schadensätze, *Bulletin of Swiss Association of Actuaries*, 70, 111-133.

Englund, M., Guillén, M., Gustafsson, J., Nielsen, L.H. & Nielsen, J.P. (2008). Multivariate latent risk: A credibility approach, *Astin Bulletin*, 38, 137-146.

Englund, M., Gustafsson, J., Nielsen, J.P. & Thuring, F. (2009). Multidimensional Credibility with Time Effects - An Application to Commercial Business Lines, *The Journal of Risk and Insurance*, 76, 443-453.

Frees, E.W. (2003). Multivariate credibility for aggregate loss models, *North American Actuarial Journal*, 1, 13-37.

Gangopadhyay, A. & Gau, W.-C. (2007). Bayesian nonparametric approach to credibility modelling, *Annals of Actuarial Science*, 2, 91-114.

Jewell, W.S. (1974). Credibility means are exact Bayesian for exponential families, *ASTIN Bulletin*, 8, 77-90.

Jewel, W.S. (1989). A General Framework for Credibility Prediction of Multidimensional First and Second Moments, *Insurance: Mathematics and Economics*, 8, 127-136.

Kamakura, W.A., Ramaswami, S., & Srivastava R. (1991). Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services, *International Journal of Research in Marketing*, 8, 329-349.

Kamakura, W.A., Wedel, M., de Rosa, F., & Mazzon, J.A. (2003). Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction, *International Journal of Research in Marketing*, 20, 45-65.

Knott A., Hayes, A., & Neslin, S.A. (2002). Next-product-to-buy models for cross-selling applications, *Journal of Interactive Marketing*, 16, 59-75.

Li, S., Sun, B., & Wilcox, R.T. (2005). Cross-selling sequentially ordered products: an application to consumer banking services, *Journal of Marketing Research*, 42, 233-239.

Mowbray, A.H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium?, *Proceedings of the Casualty Actuarial Society*, 1, 24-30.

Thuring, F., 2012, A credibility method for profitable cross-selling of insurance products, *Annals of Actuarial Science*, 6(1), 65-75.

Venter, G. (1985). Structured Credibility in Applications - Hierarchical, Multidimensional, and Multivariate Models, *Actuarial Research Clearing House*, 2, 267-308.

Whitney, A.W. (1918). The theory of experience rating, *Proceedings of the Casualty Actuarial Society*, 4, 274-292.

4

Selecting prospects for cross-selling financial products using multivariate credibility

Abstract

Insurance policies or credit instruments are financial products that involve a long-term relationship between the customer and the company. For many companies a possible way to expand its business is to sell more products to preferred customers in its portfolio. Data on the customers' past behaviour is stored in the company's data base and these data can be used to assess whether or not more products should be offered to a specific customer. In particular, data on past claiming history, for insurance products, or past information on defaulting, for banking products, can be useful for determining how the client is expected to behave in other financial products. This study implements a method for using historical information of each individual customer, and the portfolio as a whole, to select a target group of customers to whom it would be interesting to offer more products. This research can help to improve marketing to existing customers and to earn higher profits for the company.

4.1 Introduction

This chapter is a revised version of the paper Thuring et al. (2012).
Cross-selling means approaching the present customers of a company and encouraging them to increase their engagement with the company by purchasing one or many additional products. It is one of the main tools for managers to strengthen the customer relationship (Kamakura et al., 1991). In the financial sector, customers have a long-term relationship with their service provider and data on their characteristics, transactions, demographics and behaviour is stored in the company's data base, see Seng and Chen (2010) and Liao et al. (2011). This information can be used to select preferred customers and cross-sell them products they do not yet possess.

We present a method that describes how to model past behaviour in multiple financial products in order to estimate a customer specific risk profile for a certain product not yet owned by him or her, see e.g. Bae and Kim (2010) and Guillén et al. (2012) for other examples of modelling customer behaviour. Thereafter, the risk profile estimate is used to select which customers, from the company's portfolio, to approach and attempt to make a cross-sale. Knowledge about past customer behaviour in one financial services product is known to explain the performance in another related product, of the same customer (see e.g. Englund et al., 2009 or Thuring, 2012). Our objective is to show a case study of this method and to explain how such a system can be implemented in practice. The general procedure is described in Figure 4.1 where we see, from data analysis to customer selection, how a financial services company can select a target group of customers in order to cross-sell them a certain product. Initially, data on customers with several products are analysed and a model is specified. The model predicts an individual score for each customer with respect to a financial services product he/she does not own. In our paper the score is called risk profile because it predicts the customer behaviour, for a particular not owned product, given the individual information about the behaviour in other owned and related products. The company can then select a target group for a marketing campaign based on the predicted risk profiles, offer a specific product to this group and thereafter the success of the cross-sale campaign can be analysed to refine the model, see also Malthouse (2010).

In an insurance company the method presented in Section 4.3 can be used to detect customers likely to report few insurance claims, with respect to a not yet owned

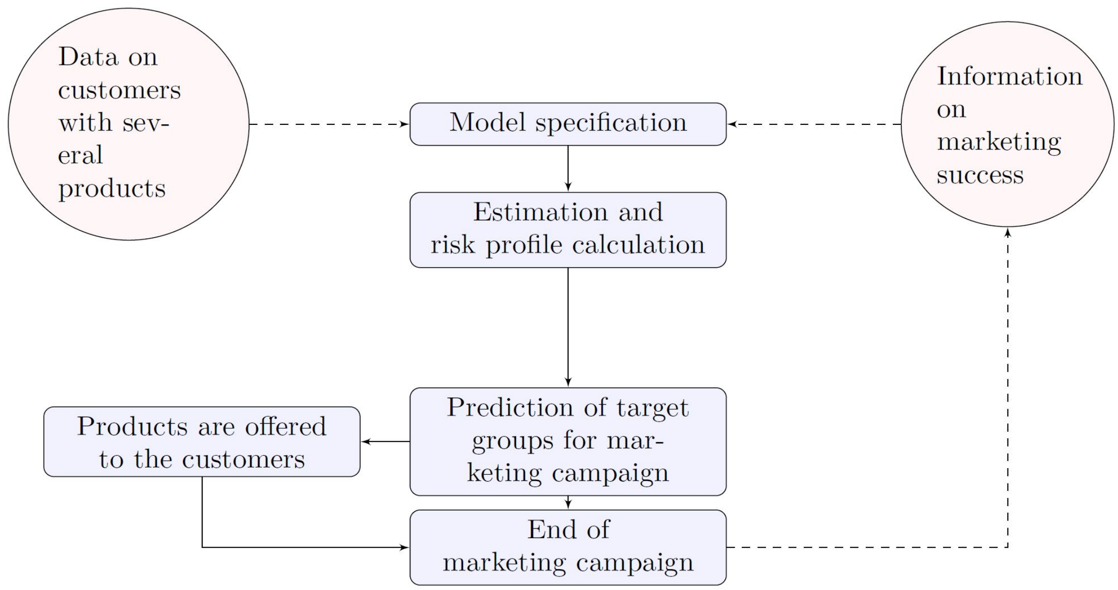

Figure 4.1: Workflow for cross-selling in the financial sector

insurance coverage, and cross-sell them that specific coverage at possibly a discounted premium level. Insurance companies normally have models for the expected (yearly) claim frequency, given certain characteristics of the customer and the insured object, which have been estimated based on collateral data on historic claims reported by past and present customers of the company, see Denuit et al. (2007) for details on claims frequency models. When predicting the claim frequency of a specific customer, such models do not usually take into consideration the individual claims experience of that customer, but predict the claims frequency based on a risk categorization which is a function of the characteristics with respect to the customer and the object. Since there can be customers with more or less risk adverse (individual) behaviour, there are cases for which the claim frequency model over-estimates or under-estimates the claim occurrence. If, for a certain customer, the claim frequency model over-estimates the claim occurrence the customer is reporting ”fewer claims than expected”, on the other hand if the claim frequency model under-estimates the claim occurrence the customer reports ”more claims than expected”. By knowing about the individual behaviour (more or less claims than expected) in one or many of a customer's existing coverages, a similar behaviour can be expected for another coverage, not yet owned by the customer. For instance, someone who has a motor insurance policy coverage and who claimed less than expected is probably also going to be claiming less than expected in other coverages such as house insurance. This phenomenon can be explained by the attitude that individuals have towards risk (see Slovic et al., 2004 and Harrison et al., 2007). People that are very much risk adverse drive carefully and also maintain their houses and belongings in good conditions. As a result, there is a correlation between the number of claims that they report to their insurance company in two different insurance coverages. On the other hand, some individuals have a completely different attitude towards risk, they are more aggressive when driving and are therefore expected to be careless about their properties too. So, when cross-selling house coverage to individuals who already have motor insurance with the company, it would be wise to take into consideration the observed number of car claims (for the specific customer) in comparison to the expected number of car claims. Note that the reverse is also true, the number of past house insurance claims can help to predict future car insurance claims.

A similar argument can be made for the banking sector. Customers who have not defaulted in the past on their loans and/or have a flawless credit card payment history, are the ones also expected to be profitable for other credit instruments. As for insurance companies, banks and other credit institutions have models and assessments for the likelihood of a customer not being able to repay credit card loans or mortgages and the concept of "fewer incident than expected" and "more incidents than expected" is applicable here as well. In the proceeding, we will refer to all events leading to a customer induced loss for a financial services company (insurance claims, loan defaults, non-repayment of credit card loans, etc.) as incidents.

The rest of the paper is organized as follows. In Section 4.2 we present the background of cross-selling and marketing of financial products. We show that selecting customers, based on behaviour in other related products, is an issue that has received limited attention in existing works. Section 4.2 also provides a short overview of credibility theory, which we use to estimate the individual risk profile. In Section 4.3 we briefly show how the risk profile can be obtained, in the cross-selling case, and Section 4.4 presents a real case study on customers from the database of a Swedish insurance company. The results illustrate how the methods can be used in practice, they show that implementation is straightforward and can lead to substantial profit improvement compared to a strategy, for cross-selling, where customers are selected randomly. Finally, Section 4.5 concludes.

4.2 Background

We first review recent cross-sale studies and thereafter the concept of credibility theory, which is the technique used for evaluating cross-sell prospects in this paper.

4.2.1 Cross sale models

Understanding and using cross-selling techniques is crucially important for a company because as the customers acquire more products from the same provider, the switching cost, associated with leaving for a competitor, increases (Kamakura et al., 2003). Therefore, cross-selling is considered a strong driver for lowering the customer churn, increasing the number of loyal customers and obtaining higher customer lifetime value (Akura and Srinivasan, 2005). In addition to this, considering product features allows significant contributions for managers striving for valuable and strong relationship with their current customer base (Larivière and Van den Poel, 2004). Another important, but not as obvious, benefit from cross-selling is that companies can learn more about the customers' preferences and buying behaviour (Kamakura et al., 2003) and cumulate various types of data to their data warehouse e.g. demographic information (Ahn et al., 2011). Such information can be used as explanatory variables to predict certain behaviours of the customers such as customer retention and profitability outcomes (Larivière and Van den Pol, 2005).

Other studies focus on modeling the probability of a successful cross-sale attempt. In an early study by Kamakura et al. (1991) probabilistic predictions are made on whether or not a customer would purchase a particular product/service based on their ownership of other products/services. In Knott et al. (2002), different models are applied to predict which product a customer is expected to buy next and the approach is further developed in Li et al. (2005), where also the appropriate time to approach a specific customer is studied.

Even though many studies have been made on cross-selling as a method for increasing a company's revenue, only few discuss potential heterogeneity in the profitability of the cross-sale prospects. As pointed out in Larivière and Van den Pol (2005), financial products are not the typical grocery products such as milk, coffee or cookies, but products that are bought and owned for a specific period in time. In addition to this, financial products are associated with uncertain costs which are determined at some (uncertain) time after the product is sold. Therefore it is not guaranteed that a successful cross-sale attempt, to a specific customer, will generate profit to the company. Instead if the cross-sold product generates claims (for an insurance company) or a loan default (for a lending bank) the financial services product actually generates a loss to the company, in most cases far greater than the income at the point of sale (insurance premium or interest payment). Englund et al. (2008) suggest that their multivariate credibility estimator could be used for evaluating cross-sale prospects by taking into account only information from the other insurance products of these specific prospects. The resulting estimate of the risk profile can be used to identify the expected profitable customers (having less than expected number of claims or loan defaults) and hence increase the company's total profit from cross-selling.

4.2.2 Credibility theory

In actuarial science, credibility theory is a technique widely used to price different insurance coverage such as health, life and property insurance (Frees, 2003). In general, the idea is to weight data, associated with an individual policyholder (or group of policyholders), with data from a collective of policyholders using a credibility weight α,

$$\text{individual estimate} = \alpha \times \text{individual data} \;+\; (1-\alpha) \times \text{collective data}\ .$$

A historical review of credibility theory starts with the papers by Mowbray (1914) and Whitney (1918) in which the credibility weight is determined ad hoc, focusing on practical applications, and not yet founded on concrete mathematical grounds. In Bühlmann (1967) (and in the more general Bühlmann and Straub, 1970, where the Bühlmann-Straub credibility model is presented) this was changed by viewing the determination of α as an optimisation problem where only the first and second order moments of the data is needed for the optimal estimator (Norberg, 2004). The generalisation of the credibility estimator to higher dimensions was introduced in Jewel (1973) and later in a multivariate hierarchical framework by Venter (1985). In Jewel (1989) the specific problem of multivariate predictions of first and second order are investigated, while a comprehensive reference to (multivariate) credibility in general is Bühlmann and Gisler (2005). A specific interpretation of the Bühlmann-Straub credibility model is found in

Englund et al. (2008) and Englund et al. (2009) where the dimensions, in the multidimensional credibility model, are interpreted as different insurance coverages, between which the claim occurrence can be more or less correlated.

4.3 Methodology

We use multivariate credibility theory to estimate a customer specific latent risk profile and thereafter evaluate if a specific additional product, of a specific customer, is expected to contribute positively to the profit of the company, if that product is cross-sold to the customer. The profit is measured as the customer specific deviation between the a priori expected number of incidents (insurance claims, loan defaults, etc) and the corresponding observed number. In the next paragraphs we present the methodology briefly and give reference to previous related work on the model and estimation technique.

4.3.1 Estimation of the risk profile

We use the multivariate credibility model of Englund et al. (2008), see also Bühlmann and Gisler (2005, p. 178) for the multivariate Bühlmann-Straub credibility model. Individuals $i = 1, \dots, I$ are customers to a financial services company and have been so during time periods $j = 1, \dots, J_i$. During these time periods, every customer has had $l = 1, \dots, K$ different financial products. We alter between k, k' and l as index for financial products in general. For each customer i in time period j and product l, we have an a priori expected number of incidents $\lambda_{ijl} = e_{ijl} g_l(\mathcal{Y}_{ijl})$, which depends on the risk exposure $0 \leq e_{ijl} \leq 1$, a regression function g_l and of a set of explanatory variables $\mathcal{Y}_{ijl}$ characterising the customer and the insured object. This can be viewed as a categorisation of the customer and the insured object into one of a large (but finite) number of risk categories. The function g_l is common for all customers i and time periods j and can be estimated, using a generalised linear model, based on collateral data of the company. We assume that e_{ijl} can take values between $[0, 1]$, where $e_{ijl} = 0$ means that the l-th product is not active (not owned) for customer i in time period j and correspondingly, $e_{ijl} = 1$ means that the product l of customer i is active (owned) during the entire time period j. We assume that N_{ijl} is a random variable describing the actual number of incidents for customer i in time periods j and product l. The observation of N_{ijl} is n_{ijl}.

Consider another random variable Θ_{il}, independent and identically distributed, which represents hidden characteristics such as risk aversion, attitude, etc. that are not captured by the explanatory variables. Θ_{il} random variables are often called the random effects. Let the pairs
$(N_{1jl}, \Theta_{1l}), (N_{2jl}, \Theta_{2l}), \ldots, (N_{Ijl}, \Theta_{Il})$ be independent. We assume $\mathbb{E}[N_{ijl}] = \lambda_{ijl}\theta_{0l}$ where $\mathbb{E}[\Theta_{il}] = \theta_{0l}$ and $Cov[\Theta_{ik}, \Theta_{ik'}] = \tau^2_{kk'}$ for $k = 1, \ldots, K$ and $k' = 1, \ldots, K$. Further we assume that N_{ijl} is Poisson distributed, conditioned on $\Theta_{il} = \theta_{il}$, with conditional expectation $\mathbb{E}[N_{ijl} \mid \Theta_{il} = \theta_{il}] = \lambda_{ijl}\theta_{il}$. The risk profile θ_{il} describes the risk that is not captured by the model for the a priori expected number of claims, of customer i and product l, and, as mentioned above, is sometimes called random effect.

We define F_{ijl} as the deviation between the actual number of incidents N_{ijl} and the a priori expected number of incidents λ_{ijl},

$$F_{ijl} = \frac{N_{ijl}}{\lambda_{ijl}} \text{ and } F_{i\cdot l} = \frac{N_{i\cdot l}}{\lambda_{i\cdot l}} = \frac{\sum_{j=1}^{J_i} N_{ijl}}{\sum_{j=1}^{J_i} \lambda_{ijl}}.$$

Other definitions, of the deviation between the expected and observed risk, are possible see e.g. Guillén et al. (2011). We assume that $Var[F_{ijl} \mid \Theta_{il}] = \frac{\sigma_l^2(\Theta_{il})}{\lambda_{ijl}}$ and that $Cov\left[F_{ijk}, F_{ijk'} \mid \Theta_{ik}, \Theta_{ik'}\right] = 0$, for $k \neq k'$.

The homogeneous multivariate credibility estimator (4.1) is the best linear unbiased estimator of $\theta_i = [\theta_{i1}, \ldots, \theta_{iK}]'$ (see Englund et al., 2009 and Bühlmann and Gisler, 2005, p. 181).

$$\hat{\theta}_i = \theta_0 + \alpha_i (F_{i\cdot} - \theta_0) \tag{4.1}$$

with $\theta_0 = [\theta_{01}, \ldots, \theta_{0K}]'$ and $F_{i\cdot} = [F_{i\cdot 1}, \ldots, F_{i\cdot K}]'$. The credibility weight $\alpha_i = T\Lambda_i(T\Lambda_i + S)^{-1}$ where T is a K by K matrix with elements $\tau^2_{kk'}$, $k = 1, \ldots, K$ and $k' = 1, \ldots, K$. The matrices Λ_i and S are diagonal matrices with, respectively, $\lambda_{i\cdot l}$, $l = 1, \ldots, K$ and σ_l^2, $l = 1, \ldots, K$ in the diagonal. The parameter $\sigma_l^2 = \mathbb{E}\left[\sigma_l^2(\Theta_{il})\right]$, where $\sigma_l^2(\Theta_{il})$ is the variance within an individual customer i, for a product l (for further details see Bühlmann and Gisler, 2005, p. 81). We also refer to Bühlmann and Gisler (2005, pp. 185-186) for parameter estimation procedures of the matrices S and T and the vector θ_0.

Performing the matrix multiplication in (4.1) and considering element k of $\hat{\theta}_i$ we get

$$\hat{\theta}_{ik} = \theta_{0k} + \sum_{k'=1}^{K} \alpha_{ikk'} \left(F_{i \cdot k'} - \theta_{0k'} \right)$$

where $\alpha_{ikk'}$ is element kk' of the matrix α_i. This can be rewritten as

$$\hat{\theta}_{ik} = \theta_{0k} + \alpha_{ikk} \left(F_{i \cdot k} - \theta_{0k} \right) + \sum_{k' \neq k} \alpha_{ikk'} \left(F_{i \cdot k'} - \theta_{0k'} \right). \tag{4.2}$$

We now assume that if product k is not active (not owned) by customer i, the risk exposure $e_{ijk} = 0$ for all j and consequently $\lambda_{ijk} = \lambda_{i \cdot k} = 0$. It is possible to show that $\lambda_{i \cdot k} = 0$ implies that $\alpha_{ikk} = 0$ and (4.2) becomes

$$\hat{\theta}_{ik} = \theta_{0k} + \sum_{k' \neq k} \alpha_{ikk'} \left(F_{i \cdot k'} - \theta_{0k'} \right), \tag{4.3}$$

where $\alpha_{ikk'}$ is element kk' of α_i when taken into consideration that $\lambda_{i \cdot k} = 0$ in Λ_i.

Equation (4.3) shows that even though a customer i does not have an active product k, it is possible to obtain an estimate of his/her specific risk profile θ_{ik} (with respect to product k) by using data of $F_{i \cdot k'} = \frac{N_{i \cdot k'}}{\lambda_{i \cdot k'}}$ with respect to the other (owned) products $k' \in \{1, \ldots, k-1, k+1, \ldots, K\}$. From a company's perspective, customers with a low risk profile are preferred and therefore the estimate of θ_{ik} can be used to assess which customers to cross-sell product k to.

4.4 Empirical study

In this section we describe the data set collected to test the cross-sale selection methodology and our experiments with this data. We require a data set describing customers who own more than one financial services product.

We conduct the experiment by neglecting the data with respect to one of the products and therefore imagine that this product is not owned by the customers. Instead the data for the other products is used to investigate if we are able to identify customers with fewer (or more) than expected number of incidents with respect to the discarded product.

4.4.1 Application data

The data sample is collected from the data base of a large Swedish insurance company writing business in both personal and commercial lines, however our sample consists solely of personal lines customers. The sample consist of a set of individuals who have been customers to the company between 1999 and 2004 and who, during this time period, have owned all of the $K = 3$ main insurance coverages provided: motor, building and content insurance. The customers have not owned the coverages for equally long time so the policy duration spans between $J_i = 3$ and $J_i = 6$ years.

We have collected data from $I = 3395$ customers and for each customer i we estimate the a priori expected number of insurance claims $\hat{\lambda}_{ijl} = e_{ijl}\hat{g}_l(\mathcal{Y}_{ijl})$ (where $\hat{g}_l$ is estimated using a collateral dataset from the same company) and collect the number of claims n_{ijl} for each year $j = 1, \ldots, J_i$ and for each of the three coverages $l = 1$ (motor), $l = 2$ (building) and $l = 3$ (content). The a priori expected number of insurance claims $\hat{\lambda}_{ijl}$ has been assessed with the claim frequency model $\hat{g}_l$, in force at the time, using the characteristics of each customer and insured object. We present the mean and standard deviation of our data in Table 1, where it can be seen that the mean of the a priori expected number of claims $\hat{\lambda}_{ijl}$ is close to the mean of the observed number of claims n_{ijl} with the exception for product $l = 2$ (building coverage). Note that the standard deviation of the a priori expected number of claims is lower than the standard deviation of the observed number of claims, which is the result of the random effects and justifies credibility estimation.

Table 1. Descriptive statistics for the Swedish insurer data from 1999 - 2004.

		Mean	Std. dev.
Motor	Expected	0.084	0.053
	Observed	0.083	0.295
Building	Expected	0.064	0.033
	Observed	0.046	0.220
Content	Expected	0.051	0.028
	Observed	0.052	0.237

In Table 2 we present the estimates of the credibility parameter matrices, S and T and the vector θ_0, when using the estimation procedures of Bühlmann and Gisler (2005, pp. 185-186). We have relatively limited amount of analysis data available, $I = 3395$

customers with between $J_i = 3$ and $J_i = 6$ years per customer, and we therefore use the same data set for estimation of the credibility parameters as for our cross-sale experiment, hence we are performing an in-sample validation.

Table 2. Estimates of the credibility parameters.

l	$\hat{\sigma}_l$	$\hat{\tau}_{l1}$	$\hat{\tau}_{l2}$	$\hat{\tau}_{l3}$	$\hat{\theta}_{0l}$
1	1.119	0.122	0.140	0.238	1.006
2	0.813	0.140	0.172	0.282	0.722
3	1.064	0.238	0.282	0.470	1.005

4.4.2 Experiment design and results

Our aim is to replicate the situation where the customers of a financial services company have a set of products but lacking one of the products offered by the company. We assume that the company is interested in selecting customers expected to have fewer than expected number of incidents. The company can achieve this by estimating the risk profile θ_{ik} for each customer i (with respect to the not owned product k) and select those with low risk profile. With our data set we imagine not knowing about the data for one of the products k and thereafter estimate the risk profile θ_{ik} with data from the other products $1, \ldots, k-1, k+1, \ldots, K$. Thereafter we order the data set by increasing $\hat{\theta}_{ik}$ and partition it into a certain number M of subsets Φ_m (of size ϕ_m) with $m = 1, \ldots, M$. The estimate of the risk profile $\hat{\theta}_{ik}$ is

$$\hat{\theta}_{ik} = \hat{\theta}_{0k} + \sum_{k' \neq k} \hat{\alpha}_{ikk'} \left(\hat{F}_{i \cdot k'} - \hat{\theta}_{0k'} \right), \text{ where } \hat{F}_{i \cdot k'} = \frac{n_{i \cdot k'}}{\hat{\lambda}_{i \cdot k'}} = \frac{\sum_{j=1}^{J_i} n_{ijk'}}{\sum_{j=1}^{J_i} \hat{\lambda}_{ijk'}}. \tag{4.4}$$

The partitioning into subsets Φ_m is needed for presenting the results in an understandable way, we used different values of M and finally concluded that $M = 5$ is an appropriate number of subsets. In this way, Φ_1 contains 20% of the customers associated with the lowest $\hat{\theta}_{ik}$, Φ_2 contains the next 20%, etc.. The number $\phi_m = 679$, for $m = 1, \ldots, 5$. Since the data sample is ordered by increasing $\hat{\theta}_{ik}$ before the partitioning into subsets Φ_m, we expect to capture customers with fewest incidents, compared to the a priori expected number, in subset Φ_1 and the customers with the most incidents, in comparison to the a priori expected number, in subset Φ_5. This can be validated by analysing the observed number of claims $n_{i \cdot k}$ in comparison to the a priori expected number $\hat{\lambda}_{i \cdot k}$ for the customers in the different subsets Φ_m, with respect to the previously imagined not owned product k. For each subset Φ_m, we are interested in the deviation Δ of the observed number of claims in comparison to the a priori expected number expressed as a percentage as follows,

$$\Delta\left(\Phi_m\right) = 100 \left(\frac{\sum\limits_{i \in \Phi_m} n_{i \cdot k}}{\sum\limits_{i \in \Phi_m} \hat{\lambda}_{i \cdot k}} - 1 \right) \ , \ m \in \{1, 2, 3, 4, 5\}. \tag{4.5}$$

Figure 4.2 describes our experiment with the data, for the situation where we are interested in identifying subsets Φ_m for product 2, using data from products 1 and 3. We use the notation $\hat{\theta}_{i,213}$ meaning that the risk profile θ_{i2} is estimated using data from products 1 and 3.

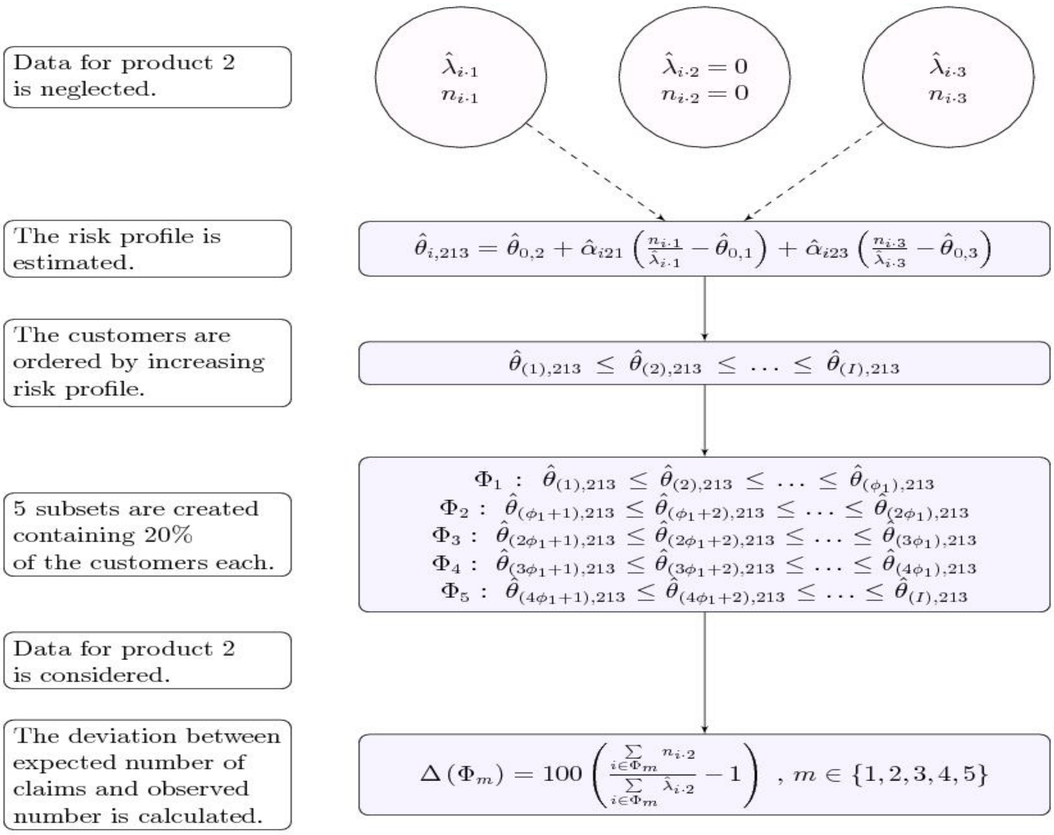

Figure 4.2: The design of the experiment for the particular case of creating subsets Φ_1 to Φ_5 based on the estimated risk profile for product 2 using information from products 1 and 3.

It is not uncommon that some customers of a financial services company only have one of the many products offered by the company. The presented methodology works in this specific case as well by setting $e_{ijk} = 0$ for the all products k which the customers does not own. I.e. for our data sample, we can also estimate the risk profile θ_{ik} by using information from only one of the two remaining products in the data set. For instance, for the estimate of the risk profile of product $k = 1$, θ_{i1}, we use the notation $\hat{\theta}_{i,12}$ if only data from product 2 is used in the estimation, and correspondingly for the other products.

The evaluation criteria (4.5), applied to investigate the deviation between the observed number of claims $n_{i\cdot k}$ and the estimated a priori expected number $\hat{\lambda}_{i\cdot k}$, in the 5 subsets Φ_m, is presented in Figures 4.3 to 4.5. In Figure 4.3, we imagine that product $k = 1$ (car coverage) is not owned by the customers and we use data from either product $k' = 2$ (building coverage) or product $k' = 3$ (content coverage) or data from both building and content coverage to estimate the risk profile θ_{i1}, with respect to product 1. Thereafter, for each of the three different estimators, we order the data set by increasing value of the risk profile estimate and partition the data into the subsets Φ_m with $m = 1, \ldots, 5$ for calculation of $\Delta(\Phi_m)$, see equation (4.5).

As seen in Figure 4.3, the credibility estimator $\hat{\theta}_{i,12}$, which uses data from product 2, does only slightly differentiate the customers with respect to claiming ($n_{i\cdot 1}$) in comparison to the a priori expected claiming ($\hat{\lambda}_{i\cdot 1}$) (left sub-figure of Figure 4.3). However, when ordering the data with respect to $\hat{\theta}_{i,13}$, which uses information from product 3, subset Φ_1 contains customers with on average 6% lower claims frequency than expected and subset Φ_5 contains customers with 22% more claims than expected, see center sub-figure of Figure 4.3. When using data from both product 2 and 3 the result is improved slightly and Φ_1 contains customers with 8% less claims than expected and Φ_5 contains customers with 26% more claims than expected.

In Figure 4.4 we imagine that product 2 (building coverage) is not owned by the customers. We see that almost all subsets Φ_m contain customers with fewer claims than expected because (according to Table 1) the average value of $\hat{\lambda}_{i\cdot 2}$ is far greater than the average value of $n_{i\cdot 2}$ since almost all customers have reported fewer claims than a priori expected. Still, the credibility estimators $\hat{\theta}_{i,23}$ (center sub-figure) and $\hat{\theta}_{i,213}$ (right

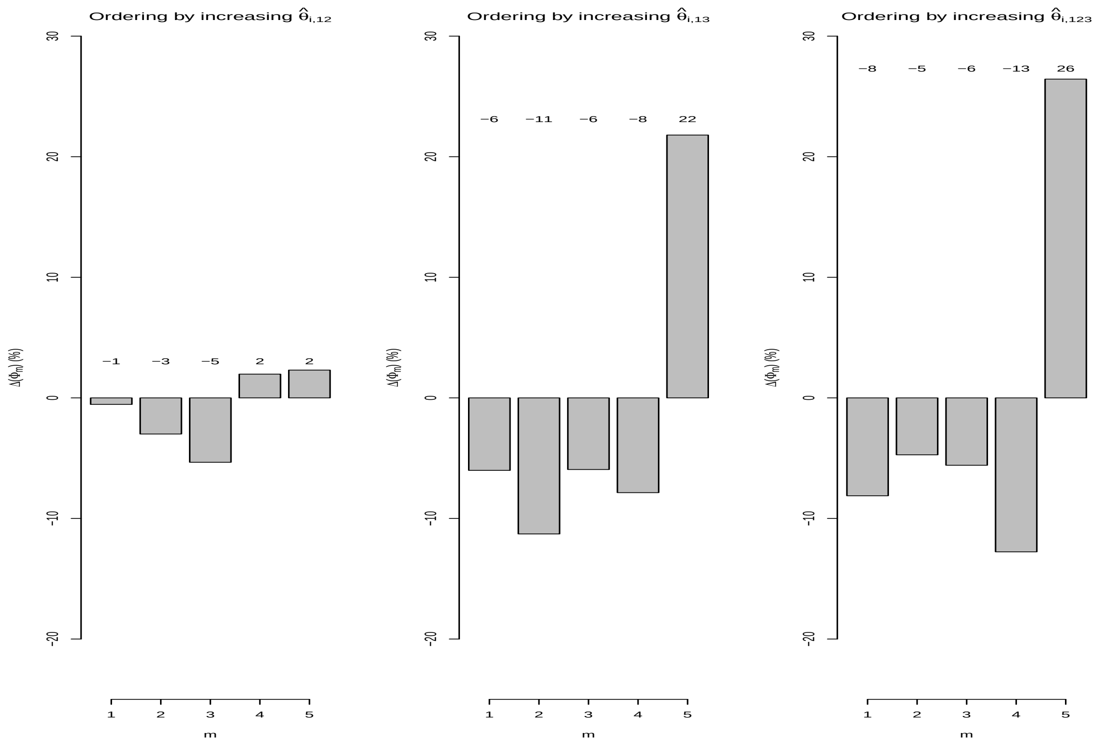

Figure 4.3: Average deviation between observed number of claims and a priori expected number with respect to product 1 (car coverage). The subsets Φ_m are created using only information from building coverage (left sub-figure), using only information from content coverage (center sub-figure) or using information from both building and content coverages (right sub-figure)

sub-figure) is able to differentiate between subsets containing customers with less than expected claiming and more than expected claiming.

In Figure 4.5, we imagine that product 3 (content coverage) is not owned by the customers. We see that all credibility estimators ($\hat{\theta}_{i,31}$, $\hat{\theta}_{i,32}$, $\hat{\theta}_{i,312}$) are identifying the customers in subset Φ_5 as having much more claims than expected. Especially the estimator $\hat{\theta}_{i,312}$ (right sub-figure) is able to identify, in the subset Φ_5, customers who have on average 64% more claims than a priori expected while also identifying the customers in the subset Φ_1 with on average 10% less claims than a priori expected.

In Figures 4.3 to 4.5 it would be expected and preferred that the deviation of the observed number of claims in comparison to the a priori expected number, $\Delta(\Phi_m)$, would be lowest for $m = 1$. However, this is not the case for many of the estimators

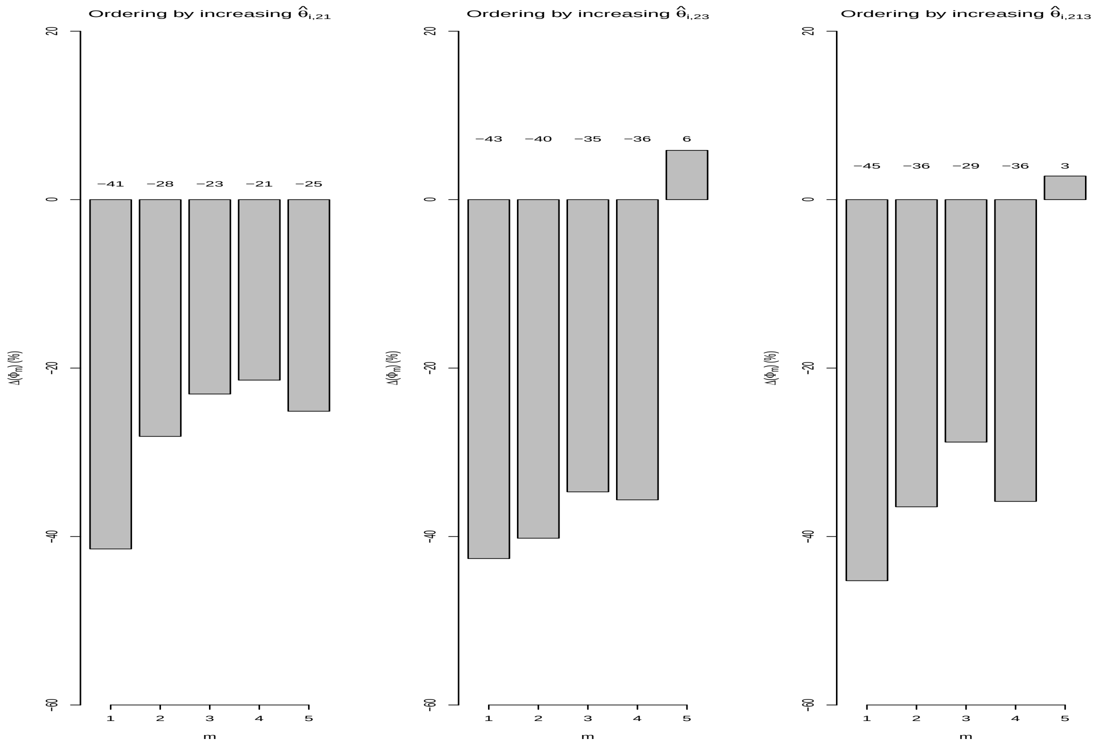

Figure 4.4: Average deviation between observed number of claims and a priori expected number with respect to product 2 (building coverage). The subsets Φ_m are created using only information from car coverage (left sub-figure), using only information from content coverage (center sub-figure) or using information from both car and content coverages (right sub-figure)

and especially for cross-selling product $k = 1$ (car) in Figure 4.3 the lowest $\Delta(\Phi_m)$ is recorded for $m = 3$, $m = 2$ and $m = 4$ for the credibility estimators $\hat{\theta}_{i,12}$, $\hat{\theta}_{i,13}$ and $\hat{\theta}_{i,123}$, respectively. A similar note can be made with regards to Figure 4.5. We draw the conclusion that for the collected data sample it is more efficient to identify a small group of customers to avoid to cross-sale to (Φ_5) than a small group of customers to target (Φ_1). Consequently, we find that by avoiding the 20% of the customers associated with the highest risk profile estimates $\hat{\theta}_{ik}$ (Φ_5) and targeting the remaining 80% the company would increase its profit significantly. In Table 3 we compare $\Delta(\Phi_5)$ to $\Delta\left(\cup_{m=1}^{4}\Phi_m\right) = \Delta(\Phi_1 \cup \Phi_2 \cup \Phi_3 \cup \Phi_4)$ where $\Phi_m \cup \Phi_{m+1}$ denotes the union of Φ_m and Φ_{m+1}.

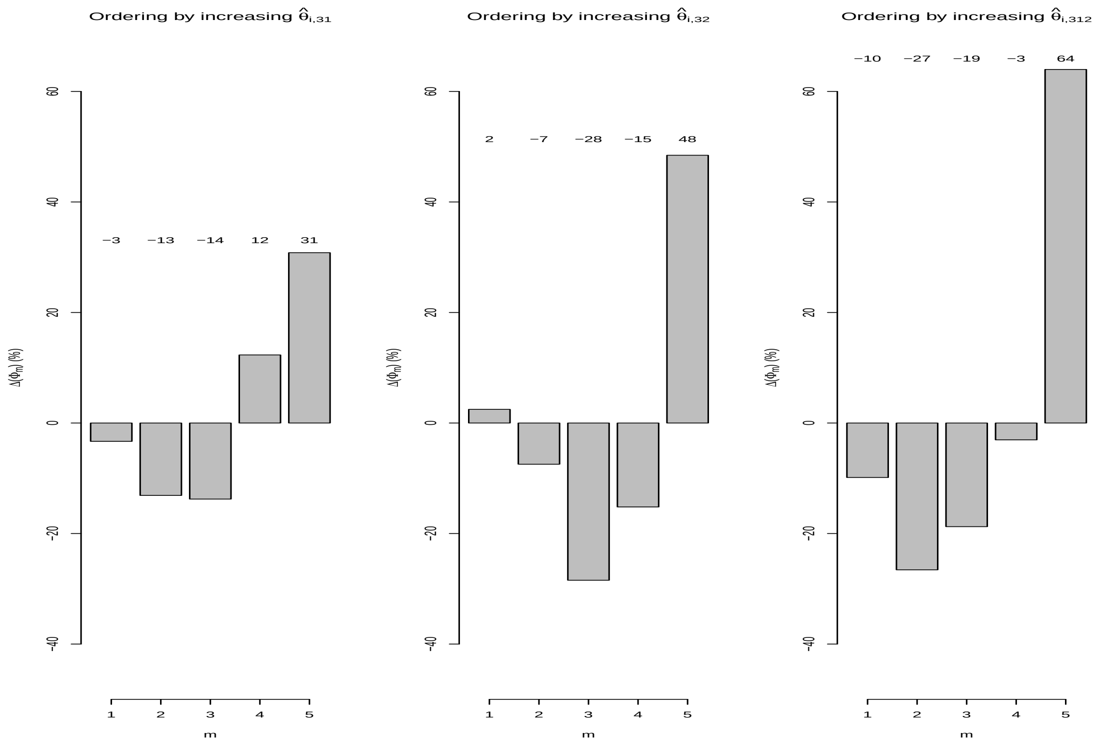

Figure 4.5: Average deviation between observed number of claims and a priori expected number with respect to product 3 (content coverage). The subsets Φ_m are created using only information from car coverage (left sub-figure), using only information from building coverage (center sub-figure) or using information from both car and building coverages (right sub-figure)

Table 3. Percentage deviation between observed and expected number of claims. Note that a positive value indicates that the subset of customers is associated with more claims than a priori expected.

Car			Building			Content		
Order	$\Delta\left(\cup_{m=1}^{4}\Phi_m\right)$	$\Delta(\Phi_5)$	Order	$\Delta\left(\cup_{m=1}^{4}\Phi_m\right)$	$\Delta(\Phi_5)$	Order	$\Delta\left(\cup_{m=1}^{4}\Phi_m\right)$	$\Delta(\Phi_5)$
Random	0%	-4%	Random	-29%	-30%	Random	4%	3%
$\hat{\theta}_{i,12}$	-2%	2%	$\hat{\theta}_{i,21}$	-30%	-25%	$\hat{\theta}_{i,31}$	-4%	31%
$\hat{\theta}_{i,13}$	-8%	22%	$\hat{\theta}_{i,23}$	-40%	6%	$\hat{\theta}_{i,32}$	-9%	48%
$\hat{\theta}_{i,123}$	-8%	26%	$\hat{\theta}_{i,213}$	-38%	3%	$\hat{\theta}_{i,312}$	-13%	64%

Table 3 shows that by selecting the 80% ($\cup_{m=1}^{4}\Phi_m$) most favorable customers, with respect to the estimate of the risk profile θ_{ik}, the company is able to avoid customers associated with up to 64% more claims than a priori expected (content coverage, product 3). In the table we have also included results produced when the data sample has been randomly ordered and partitioned into 80% of the data and 20% of the data. The random order does not differentiate between subsets of customers with respect to percentage deviation between observed and expected number of claims. We see a similar pattern for product 1 (car) where the 80% most favorable customers are associated with 8% less claims than a priori expected while the remaining 20% are associated with 26% more claims than a priori expected. The performance of the credibility estimators in Product 2 (building) is difficult to interpret because almost all customers are associated with lower observed claim occurrence than a priori expected. However, even for this particular situation a subset Φ_5 can be identified consisting of customers with on average 6% more claims than a priori expected. Note that this is not received for the credibility estimator which uses all available information ($\hat{\theta}_{i,213}$) but the estimator which only uses data from the content product $k' = 3$, $\hat{\theta}_{i,23}$.

4.5 Discussion

This study investigates identification of customers to whom additional products should be offered, by estimating a customer specific risk profile with the use of behavioural data from other products of the specific customers. We use a standard multivariate credibility model applied to a portfolio of customers, of a financial services company, owning several financial products from the company. The model allows us to take into consideration the possible (positive) correlation in customer behaviour between different financial products and estimate the customer specific risk profiles, for a specific product not owned by the customer, without having observed any customer specific information with respect to that particular product. Instead, data on customer behaviour, with respect to the other (owned) products, is the only necessity for estimating the risk profile.

The methodology uses only two observables: the a priori expected number of incidents and the observed number of incidents. We assume that the financial services company has a model for the a priori expected number of incidents or is able to assess a value specific for each customer or category of customers. When estimating such models it is unusual to incorporate information about the number of incidents related to a specific customer. Instead the company finds patterns which can be used to categorise the customers, with respect to the expected occurrence of incidents, based on customers' characteristics. It is not uncommon that customers are associated with more or less number of incidents, than suggested by the categorisation, based on their attitude towards risk. In our methodology we use that the attitude towards risk seems to be similar across different financial products. I.e. if a customer is associated with more or less number of incidents, than a priori expected in some products, it is likely that this pattern will also emerge in other related products.

With the presented credibility estimators we are able to assign, to each customer, a specific estimate of his/her risk profile based on data which the company has available. We use the estimate, of each customer's risk profile, to identify subsets from the data containing customers associated with more or less incidents than a priori expected. In this way the company receives knowledge about which customers to target for cross-selling and which to avoid.

In our empirical study we analyse our methodology on real data from a large Swedish insurance company, consisting of personal lines customers with three different insurance coverages. We find that there are subsets of the data sample with large heterogeneity with respect to claiming in comparison to expected claiming. Furthermore, we find that these subsets are identifiable by using an appropriate credibility estimator of the risk profiles. The appropriateness of a specific credibility estimator is dependent of the considered product, but in most cases an estimator which uses all available information is preferable. We find that it is easier to identify the 20% of the data containing customers to avoid than the 20% of the data containing customers to target. In fact, by targeting all customers but the worst 20%, the company could expect a subset of customer associated with less claims than a priori expected indifferent of which product is considered. The remaining 20% of the data sample consist of customers with up to 64% more claims than a priori expected.

References

Ahn, H., Ahn, J.J., Oh, K.J. and Kim, D.H. (2011). Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques. *Expert Systems with Applications*, 38(5), 5005-5012.

Akura, M.T. and Srinivasan, K. (2005). Research Note: Customer Intimacy and Cross-Selling Strategy. *Management Science*, 51(6), 1007-1012.

Bae, J., K. and Kim, J. (2010). Integration of heterogeneous models to predict consumer behavior. *Expert Systems with Applications*, 37(3), 1821-1826.

Bühlmann, H. (1967). Experience rating and credibility, *Astin Bulletin*, 4, 199-207.

Bühlmann, H. and Gisler, A. (2005). *A Course in Credibility Theory and its Applications* (Berlin, Germany: Springer Verlag).

Bühlmann, H. and Straub, E. (1970). Glaubwürdigkeit für Schadensätze, *Bulletin of Swiss Association of Actuaries*, 70, 111-133.

Denuit, M., Marechal, X., Pitrebois, S. and Walhin, J.F. (2007) *Actuarial modelling of claim counts: risk classification, credibility and bonus-malus systems* John Wiley and Sons, Ltd. New York.

Englund, M., Guillén, M., Gustafsson, J., Nielsen, L.H. and Nielsen, J.P. (2008). Multivariate latent risk: A credibility approach, *Astin Bulletin*, 38, 137-146.

Englund, M., Gustafsson, J., Nielsen, J.P. and Thuring, F. (2009). Multidimensional credibility with time effects - an application to commercial business lines, *The Journal of Risk and Insurance*, 76(2), 443-453.

Frees, E.W. (2003). Multivariate credibility for aggregate loss models. *North American Actuarial Journal*, 7(1), 13-37.

Guillén, M., Nielsen, J.P. Scheike, T.H. and Pérez-Marín, A.M. (2012). Time-varying effects in the analysis of customer loyalty: A case study in insurance, *Expert Systems with Applications*, 39(3), 3551-3558.

Guillén, M., Pérez-Marín, A.M. and Alcañiz, M. (2011). A logistic regression approach to estimating customer profit loss due to lapses in insurance. *Insurance Markets and Companies: Analyses and Actuarial Computations*, 2, 42-54.

Harrison, G.W., Lau, M.I. and Rutstrm, E.E. (2007). Estimating risk attitudes in Denmark: A field experiment. *Scandinavian Journal of Economics*, 109(2), 341-368.

Jewel, W.S. (1973). Multidimensional credibility. *Actuarial Research Clearing House* 4.

Jewel, W.S. (1989). A General Framework for Credibility Prediction of Multidimensional First and Second Moments, *Insurance: Mathematics and Economics*, 8, 127-136.

Kamakura, W.A., Ramaswami, S., and Srivastava R. (1991). Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing*, 8, 329-349.

Kamakura, W.A., Wedel, M., de Rosa, F., and Mazzon, J.A. (2003). Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing*, 20, 45-65.

Knott A., Hayes, A., and Neslin, S.A. (2002). Next-product-to-buy models for cross-selling applications. *Journal of Interactive Marketing*, 16(3), 59-75.

Li, S., Sun, B., and Wilcox, R.T. (2005). Cross-selling sequentially ordered products: an application to consumer banking services. *Journal of Marketing Research*, 42, 233-239.

Liao, S-H., Chen, Y-J. and Hsieh, H-H. (2011). Mining customer knowledge for direct selling and marketing. *Expert Systems with Applications*, 38(5), 6059-6069.

Larivière, B. and Van den Poel, D. (2004). Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. *Expert Systems with Applications*, 27(2), 277-285.

Larivière, B. and Van den Poel, D. (2005). Predicting customer retention and profitability by using random forest and regression techniques. *Expert Systems with Applications*, 29(2), 472-484.

Malthouse, E.C. (2010). Accounting for the long-term effects of a marketing contact. *Expert Systems with Applications*, 37(7), 4935-4940.

Mowbray, A.H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium? *Proceedings of the Casualty Actuarial Society*, 1, 24-30.

Norberg, R. (2004). Credibility theory. Encyclopedia of Actuarial Science. John Wiley & Sons. 398-406.

Seng, J-L., and Chen T.C. (2010). An analytical approach to select data mining for business decision. *Expert Systems with Applications*, 37(9), 8042-8057.

Slovic, P., Finucane, M., Peters, E., and MacGregor, D.G. (2004). Risk as Analysis and risk as feelings: Some thoughts about affect, reason, risk and rationality. *Risk Analysis*, 24(2), 311-322.

Thuring, F., 2012, A credibility method for profitable cross-selling of insurance products, *Annals of Actuarial Science*, 6(1), 65-75.

Thuring, F., Nielsen, J.P., Guillén, M., Bolancé, C., 2012, Selecting prospects for cross-selling financial products using multivariate credibility, *Expert systems with application*, 39(10), 8809-8816.

Venter, G. (1985). Structured Credibility in Applications - Hierarchical, Multidimensional, and Multivariate Models, *Actuarial Research Clearing House*, 2, 267-308.

Whitney, A.W. (1918). The theory of experience rating *Proceedings of the Casualty Actuarial Society*, 4, 274-292.

5

Optimal customer selection for cross-selling of financial services products

Abstract

A new methodology, for optimal customer selection in cross-selling of financial services products, such as mortgage loans and non life insurance contracts, is presented. The optimal cross-sales selection of prospects is such that the expected profit is maximized, while at the same time the risk of suffering future losses is minimized. Expected profit maximization and mean-variance optimization are considered as alternative optimality criteria. In order to solve these optimality problems a stochastic model of the profit, expected to emerge from a single cross-sales prospect and from a selection of prospects, is developed. The related probability distributions of the profit are derived, both for small and large portfolio sizes and in the latter case, asymptotic normality is established. The proposed, profit optimization methodology is thoroughly tested, based on a real data set from a large Swedish insurance company and is shown to achieve considerable profit gains, compared to traditional cross-selling methods, which use only the estimated sales probabilities.

5.1 Introduction

This chapter is a revised version of the paper Kaishev et al. (2013).
This paper addresses the challenge of optimally selecting a subset of customers, for cross-selling products to, where the profit of a given cross-sale is unknown and customer specific. Imagine a financial services company with a significant data base and a traditional long relationship with each customer, once they purchase their products. This is indeed the situation for most financial services products. In that situation the cross sale challenge becomes to use your data base in general and your specific knowledge of your individual cross-sale target to estimate, for the specific customer, the probability of a cross sale, the cost of a cross sale attempt, the average discounted future profit and the uncertainty of the profit of the entire cross sale attempt for that individual. Once reliable estimates for the stochastics of the cross sale process have been established, one can optimize the cross sale profit according to a variety of criteria including return and risk. In this paper, we first consider the simple question of optimizing the average profit, but we also consider one version of adjusting for risk when optimizing cross sale profits. Our extensive case study is taken from non-life insurance, where our sales probability model is provided to us by the company that also provided us with the data. When estimating our cross sale profit, we combine classical regression techniques and state-of-the-art actuarial latent risk technology enabling us to combine the overall cross sectional information in our data with experience information on a specific customer. Our technique generalises to other situations, one could apply classical regression alone leaving out the latent risk part or vice versa, one could work only with the latent risks. While our approach has been developed with an eye to the financial service industry, with its abundant data bases, our approach would be useful also in other businesses.

Profitability in the general context of direct marketing has been researched by a number of authors, such as Bult and Wansbeek (1995), Venkatesan and Kumar (2004) and Gönül and Hofstede (2006). The early paper by Bult and Wansbeek (1995) addresses the problem of finding an optimal selection of target customers from a mailing list but does not consider cross-sales. The optimal selection is based on the customer response (sale or no sale) to a direct marketing offer of books, periodicals and music to households by a retailer in the Netherlands. Given sale, it is assumed that the marginal, i.e. per customer, return (profit) is deterministic. Venkatesan and Kumar (2004) consider customer selection based on their customer life time value. While this customer life time value clearly is a stochastic variable, Venkatesan and Kumar (2004) concentrates on average profit values closely related to the average profit approach of this paper. The customer specific information of Venkatesan and Kumar (2004) comes from a classical regression technique. The approach of Venkatesan and Kumar (2004) is useful both when considering first sales and cross sales. Were they to consider cross sale only, as we do in this paper, then specific individual customer information would be available and could be used to further optimize the customer selection. Gönül and Hofstede (2006) consider a broader set of optimisation objectives such as profit maximisation, customer retention and utility maximisation. They find that optimising their objective function over multiple periods leads to higher expected profits and higher expected utility. They apply their methodology to the problem of setting optimal sales catalogue mailing strategies. Their optimal solutions indicate that fewer catalogues should be mailed than is the current practice in order to maximise the expected profit and expected utility. In their set-up both profit margin and the campaign costs are modelled deterministically resulting in an approach closely related to the optimal average profit approach of this paper. They do not specifically consider cross sales and the added specific customer data available in this case. In contrast to Bult and Wansbeek (1995), Venkatesan and Kumar (2004) and Gönül and Hofstede (2006), our approach allows us to exploit the extra customer specific information available in a cross sale context. In our concrete example, we use recently developed actuarial technology based on multivariate credibility theory to assess the individual specifics in case of a cross sale, but we also point out that other approaches could be possible. Another novel feature of our profit optimisation approach is that one of our optimisation criteria balances the contradictory goals of maximising profit and minimising risk. We illustrate, based on a real data set, how our optimisation methodology works by applying it to the context of cross-selling of financial services products and in particular, insurance policies. So, the proposed methodology is thoroughly tested with real data from an insurance company and it is demonstrated that significant profit gains can be achieved by applying it in practice.

There is a considerable marketing literature on cross-selling and we refer the interested reader to papers by Kamakura et al. (1991), Knott et al. (2002), Kamakura et al. (2003), Kamakura et al. (2004), Li et al. (2005), Kamakura (2007), and Li et al. (2010). Cross-selling through call center's has recently been addressed also by Gurvich et al. (2009) who study the operational control problem of decision making, staffing, call routing and cross-selling to possibly different classes of customers. These authors consider segmenting the (caller) population of sales prospects in order to decide to whom and at what price to cross-sell so as to increase the expected profitability of a call center's dynamic cross-selling campaign. Increased profitability is achieved by customizing the (product) price, offered to each segment (type of customers) while keeping the product specification common to all segments, and by reducing the volume (cost) of cross-selling attempts unlikely to be profitable. As an illustration of their approach, the authors consider certificates of deposit (CD) which guarantee a fixed interest rate over a fixed time interval, a product offered by banks to different customers. In this paper we consider profitability of cross-selling and propose a stochastic model of the profit . Although our main example is cross-selling of a financial product, stochastic profits (including stochastic costs) is of course also relevant in a broader context of direct marketing. For example, sellers who use electronic sales channels usually offer free delivery, the costs of which are not known before the order is placed and therefore are of stochastic nature. In general, in direct marketing, a data base of customers from other campaigns may be available and recorded profits of these customers may vary considerably. For example, one could imagine that some type of customers only take the company's "Welcome offer" and nothing else. The profit then will be small, or even negative, on those customers. On the other hand, other customers may take the welcome offer and also buy other products. It is possible to extract information from the data base on "who is who", in terms of profit and cost, and it is possible to take advantage of that in selecting the customers that maximise the total expected profit.

While our overall model is indeed general in nature, it seems particularly relevant when cross selling financial service products. Financial services offered by banks and insurance companies, such as mortgage contracts and other types of loans, household, car and motorcycle insurance policies, and other types of personal lines insurance products, differ in several ways from other conventional retail products and services which other firms (call centers) attempt to cross-sell. There is a policy duration specified at the date of sale of a financial product and also the cost associated with a specific customer is stochastic and becomes known to the organization at some random time after the sales date. For example, the cost generated by an insurance policy is mainly determined by the claim amount which depends on the occurrence and severity of the related insured event. In a mortgage setting, a holder of a mortgage contract may default on his/her loan repayment at some random moment within the duration of the contract, which may lead to a loss for the lending bank or its insurance company, of unknown (random) size.

Our stochastic model of profit involves three random quantities, a binary random variable, modelling the event of cross-selling, a random variable modelling the price of the offered product and another random variable, modelling the cost associated with a specific customer for the cross-sale product. In the appendix, we study the distributional properties of this profit model and propose formal criteria for optimizing not only the profit but also the risk of suffering future losses, faced by the financial services organization in a cross-sales campaign. In this way, the contradictory goals of maximizing profit while at the same time minimizing the risk of losses are achieved already at the marketing stage. The proposed novel, profit optimization methodology allows us to find the size and the composition of an optimal selection of cross sales prospects, from a large portfolio of existing customers, so that an appropriate profit/risk optimization criterion is maximized. We further address the estimation of the profit model parameters, among which, the individual risk profile parameter, the claim frequency and severity and the sales probability. The methodology is validated on a real, insurance data example. The results confirm that substantial profit gains can be achieved by applying it in cross-selling of financial services products.

Our paper is organized as follows. In Section 5.2 we propose a stochastic model for the profit associated with cross-selling an additional product to an existing customer. Section 5.3 elaborates on two established methods for capturing customer heterogeneity and how they are combined in this paper. In Section 5.4 we relate our profit optimization methodology to the existing marketing literature cases mentioned in the introduction and we discuss how these existing marketing cases could be generalised to the varying profit set-up of this paper. Thereafter, in Section 5.5, we study an example of cross-selling insurance policies to existing customers of an insurance company. Concluding remarks are found in Section 5.6 followed by an appendix with details on results from the insurance example.

5.2 Optimal selection of cross-sale prospects

Our contribution of this paper is to consider marketing campaigns where the profit of the customer is stochastic. Our particular interest is that some prior knowledge is available on this stochasticity and we want to take advantage of this prior knowledge. So, in the paper, knowledge on profit is focused on, on top of the probability of sales model - the latter is not our center attention. In Section 5.4 we give a wide array of possible situations where a profit formula might be of interest.

5.2.1 Modelling the stochastic cross-sales profit

It is natural to model the (stochastic) profit (loss), H_{ik}, associated with cross-selling an additional product, indexed k, to the i-th existing customer as

$$H_{ik} = \mathbf{1}_{\{A_{ik}\}} \left(\Pi_{ik} - S_{ik}\right) - \omega_{ik}, \tag{5.1}$$

where $\mathbf{1}_{\{A_{ik}\}}$ is the indicator random variable, A_{ik} is the event of cross selling to the i-th customer the k-th product with cross-sale probability p_{ik}, at the stochastic price Π_{ik}, and $\omega_{ik} > 0$ is the (deterministic) customer-specific cost of a cross sale attempt. The random variable S_{ik} is the stochastic cost related to the i-th customer and k-th product. The cost ω_{ik} is usually related to organizing the cross-sale campaign through call centers or otherwise. The motivation behind representation (5.1) is straightforward, given sale occurs, the profit is equal to the price charged to the customer minus his/her stochastic cost, less the cost ω_{ik}, incurred by the company for approaching the i-th cross-sale prospect. Alternatively, if no sale occurs, a loss of ω_{ik} is accounted for by the company. At this point we do not assume independence of the incidence of a cross-sale and the stochastic profit and we do not assume independence between different customers. In our main example given in Section 5, we follow the classical approach of actuarial pricing and cross selling and assume such independence.

We denote by $\mu_{ik} = \mathbb{E}\left[H_{ik}\right]$ the mean of the stochastic variable H_{ik} and by $v_{ik} = \mathrm{Var}\left[H_{ik}\right]$ the variance of the same. The mean of the profit can take both positive and negative values and it is obvious that the company should try to cross-sale to customers with a positive profit. So, one alternative to select customers who should be targeted is to select those associated with $\mu_{ik} > 0$. An obvious way of doing so is to order the customers in a non-increasing order of the expected profit. The cut-off point is then the point at which the cumulative sums, $\sum_{i=1}^{l} \mu_{ik}$, $l = 1, \ldots, I$, do not increase any more.

Another alternative criterion for selecting customers takes into account both the expected profit and its variance since it is desirable not only to maximise the profit (interpreted as a performance measure) but also to minimise its variance (interpreted as a risk measure). One way of combining these two performance and risk measures is to consider the mean-variance selection criterion, $MV_{ik} = \mu_{ik} - \xi v_{ik}$, where $\xi > 0$ (see Section 5.2). Note that any correlation between $\mathbf{1}_{\{A_{ik}\}}$ and S_{ik} will only affect selections with the mean-variance criteria.

In summary we have two separate criteria for selecting customers to approach for cross-selling a policy k; all customers associated with a positive expected profit μ_{ik} (called the EP-criteria) or all customers associated with a positive mean-variance value MV_{ik} (called the MV-criteria).

5.3 Modeling customer heterogeneity

The overall approach suggested in this paper requires customer specific knowledge leading to a more accurate optimization of profit. In this section, we point out two established methods for capturing such customer heterogeneity. The choice of a multivariate model depends on the nature of the available customer information. If only descriptive information such as age, geography and sex is available, the first idea that comes to mind would be to set up a multivariate generalised linear model to describe customer heterogeneity. As mentioned below, this type of approach is well known in the marketing literature. However, if also some historical information is available on the individual behavior of a given customer, then this could be modelled through an individual latent variable. While this type of approach has a long and celebrated history in the academics and practice of actuarial science, it seems less focused on in marketing applications. The two multivariate modelling approaches - and their combination - are briefly described below.

5.3.1 Multiple regression analysis

The key issue in multiple regression analysis (specifically in marketing) is to estimate a set of weights corresponding to a set a characteristics, sometime called antecedents, of the customers. When estimated, the weights are used to produce a weighted sum of the corresponding set of characteristics, of other similar customers, in order to estimate e.g. a probability, a price, or any other customer metric of interest. The resulting metric is received by applying a so called link function to the weighted sum of customer characteristics.

There are many examples of modeling customer heterogeneity using multiple regression analysis and one straightforward, and very related to our paper, is Knott et al. (2002). This study is on so called next-product-to-buy models for institutions with a large customers database, aiming at selecting the most appropriate customers to approach and the most appropriate product to offer them. The authors compare different regression (and other modeling) techniques on data from a retail bank interested in increasing sales of a particular loan product.

Another example of multiple regression analysis in marketing is Malthouse (1999) where the specific problem of modeling mail order responses is considered. The author seeks a simple but predictive model using either multiple regression with variable subset selection or so called ridge regression. As mentioned, it is common for direct marketers to be more interested in overall model performance (measured with e.g. gains charts) than unbiased parameter estimates which is why the ridge regression is considered in this particular case.

5.3.2 Latent variable models

No matter how much cross sectional data we might have available, there is likely to remain some unobserved heterogeneity of specific customers. Two households with the same number of children, living on the same street and with all other observable characteristics being equal might have completely different profitability for a particular product we wish to cross sell. The unobservable mountain climbing habit of one of the fathers or the unobservable alcohol habits of one of the mothers could for example play a role for the profitability of many type of products. One dimensional unobservable variables have a long history in theoretical as well as practical non-life insurance pricing, where it some times is called experience rating. Latent variables are also considered in the marketing context, see for example Rossi and Allenby (2003) or Kamakura et al. (1991). Other applications of latent variables can be found in the related research field of moral hazard and adverse selection where these effects typically are modelled as latent variables, see Akerlof (1970) and Rothschild and Stiglitz (1976) for a theoretical discussion on these issues and e.g. Cohen (2005) for more practical study. In our practical concrete example from non-life insurance below, we have introduced a multivariate latent variable modelling all relevant products at the same time. When optimizing our cross sale profit, we then exploit the general information on how an individual's latent variable from one product correlates with that very same individuals latent variable from the product we wish to cross sell.

5.3.3 Combining multiple regression analysis with latent variable models

For our model, for the stochastic profit H_{ik} (5.1), we propose that the two stochastic variables $\mathbf{1}_{\{A_{ik}\}}$ and S_{ik} can be modeled with multiple regression analysis and latent variable techniques, respectively. Furthermore we propose using credibility theory which includes experience of customers beyond covariate (antecedents) information. Consequently, when implementing this model for cross-sale selections, the company makes use of its data base more effectively by using one source of data for the multiple regression analysis and another source of data for latent variable techniques. The latter data source is often neglected, since the literature on latents variables in cross-selling is limited, however we will show, in Section 5.5.2, how this data can be useful and improve the overall profit from cross-selling.

5.4 Examples of modeling profit in direct marketing

In this section we relate our above profit optimization methodology to the existing marketing literature cases mentioned in the introduction and we give some insight into how these existing marketing cases could be generalised to the varying profit set-up of this paper. All the three marketing cases treated in the introduction have a fixed profit given sale, we point out that a varying profit given sale could be considered in these cases and we point out that the methodology of this paper would be applicable in these three well known marketing cases if they would be generalised to the varying profit case. Varying profit modelling requires statistical estimation of the multivariate nature of our customer data base and we point out the type of data needed in each case to carry out either a generalised model estimation approach, the latent variable estimation approach or a combination of both in the cited works. In the next section we will treat in detail an example from the insurance industry, where sufficient data is available to combine the generalised model estimation approach and the latent variable estimation approach

5.4.1 Bult and Wansbeek (1995)

In the early paper by Bult and Wansbeek (1995) it is assumed that the returns (profit) of a positive reply is constant across households and based on an ordering of the customer data base, with respect to the estimated probability of a customer responding to a direct mail, the authors find an optimal selection consisting of customers with positive marginal profit. The varying profit for a given customer depends in this model only on the varying probability of a cross-sale. Given a sale, the profit is the same for all customers. If one was to follow our approach one could model the profit given a sale as a stochastic variable, where both the mean profit and its variance can vary among customers. This is relevant if the customer has a choice among a variety of products to buy at the cross-sale, in this example the choice of buying one or more books or records. One could also consider the probability of buying more books or records at a later point in time or the probability of canceling an order, etc.. All these events would affect the total profit from one particular customer (household) and would be helpful to target the most profitable customers if taken into account. If one would have data available to model the multivariate nature of how much a given customer would buy given a sale, one could implement the profit optimization method of this paper. Such data could be given by co-variates - e.g. age, sex, geographic details - where a generalised linear model might be useful, or one could imagine that information was present on the historical nature of this particular customers likeliness to buy during a cross-sale, in this latter case, the latent variable approach might work well. Or one could have both types of data available allowing one to combine the two methods of multivariate modelling. Therefore, the approach of Bult and Wansbeek (1995) could be sophisticated and more profit could be made if extra relevant data would be available.

5.4.2 Venkatesan and Kumar (2004)

The second study, related to our work, is by Venkatesan and Kumar (2004) on selecting customers based on their customer lifetime value. The model they are presenting considers estimated profits from every possible purchase of computer hardware a customer will make during the engagement. Venkatesan and Kumar (2004) have useful co-variate information of their customers and model the lifetime value through a generalised linear model approach. However, as the customer data base of the computer hardware company grow, it seems plausible that historical information could be gathered on the nature of the loyalty of each customer, such that a latent variable measuring loyalty could supplement the approach given in Venkatesan and Kumar (2004) leading to even more specific marginal profit calculations.

5.4.3 Gönül and Hofstede (2006)

The third example of Gönül and Hofstede (2006) considers direct marketing and optimal catalog mailing decisions. The authors model order incidence and order volume separately to later combine them into a utility based profit optimisation where the (constant) cost of sending a catalog and the (constant) profit margin is included. Based on the level of risk aversion of the company managers, optimal mailing strategies are selected. As in the example of Bult and Wansbeek (1995), the profit from a single customer can be considered variable by assuming that different customers might require different treatment and e.g. might demand facilities for canceling orders or returning already received items. The probability of a specific customers requiring such facilities could be modelled with data on historic customer behaviour from related products or orders. The specific cost of sending a catalog can also be considered as variable, as we allow for in our model by incorporating an index i of the cost of a cross-sale contact ω_{ik}. Introducing variability in the catalog mailing cost and the profit is mentioned as an interesting topic for further research by Gönül and Hofstede (2006). We consider the more flexible profit optimization model of this paper to be a natural place to start for such further research.

5.5 An example from the insurance industry

In the specific case of cross-selling insurance policies, the stochastic variable S_{ik} is normally called the aggregate claim amount resulting from customer i in insurance coverage k which is composed of the number of insurance claims N_{ik} and their corresponding severities $X_{ik1}, \ldots, X_{ikN_{ik}}$ as the following sum

$$S_{ik} = \sum_{n=1}^{N_{ik}} X_{ikn}.$$

We follow classical actuarial approaches to insurance modelling, see among many others Klugman et al. (1998) and assume independence between customers. That is of course not fully correct. The insurance policies of different policyholders might be affected by the same external circumstances such as weather conditions or economic conditions. Such correlation could affect our preferences when we apply our mean-variance optimization, but it will not affect our main example optimizing the average profit. Further discussion about these, and other, common assumption in actuarial science are found in Beard et al. (1984, p. 33), Jong and Heller (2008, p. 81) and Ohlsson and Johansson (2010, p.18). Furthermore we assume that conditioned on the latent random risk variable Θ_{ik}, the aggregate claim amount and the indicator random variable $\mathbf{1}_{\{A_{ik}\}}$ are independent and that Θ_{ik} is independent of $\mathbf{1}_{\{A_{ik}\}}$. The second assumption could be challenged by the fact that customers associated with a low risk variable could be less inclined to purchase the offered product due to experience rating at the company currently providing that particular product. Assume from now on that N_{ik} is conditionally Poisson distributed given a latent random variable. We do not make any assumptions on the distribution of the latent variable, however, should it be gamma distributed, then this implies a negative binomial distribution of our counts N_{ik}. In Section 5.5.2 we test this conditional Poisson assumption in more than one way and we provide a graph indicating that our counts indeed very needly follow the appropriate negative binomial distribution. The expectation of N_{ik}, conditioned on the latent random risk variable Θ_{ik}, is $\mathbb{E}\left[N_{ik} \mid \Theta_{ik} = \theta_{ik}\right] = \lambda_{ik}\theta_{ik}$ and X_{ik} has expectation $\mathbb{E}\left[X_{ik}\right] = m_{ik}$, we do not make any distributional assumption about X_{ik}. We call λ_{ik} the a priori expected number of insurance claims and assume that the insurance company has a method for estimating it. By assuming independence between N_{ik} and X_{ik} the expectation of S_{ik} (conditioned on Θ_{ik}) becomes

$$\mathbb{E}\left[S_{ik} \mid \Theta_{ik} = \theta_{ik}\right] = \mathbb{E}\left[N_{ik} \mid \Theta_{ik} = \theta_{ik}\right] \mathbb{E}\left[X_{ik}\right] = \lambda_{ik}\theta_{ik}m_{ik}.$$

In our example, we assume that the price (premium) π_{ik} is deterministic. Premium setting in insurance is a highly complex task including estimating the expected claims frequency and severity as well as cost loadings for administration, sales commission, discounts, re-insurance, etc. Additionally, with the recent introduction of dynamic pricing, the premium will in some cases also depend on customer demand, market and competitor situation and customer life time value. The scope of this example does not allow for any further details on premium setting. Under these assumptions we can express the conditional mean (μ_{ik}) and variance (v_{ik}) of the profit H_{ik} as

$$\mu_{ik} = \mathbb{E}\left[H_{ik}|\Theta_{ik} = \theta_{ik}\right] = p_{ik}\left(\pi_{ik} - \theta_{ik}\lambda_{ik}m_{ik}\right) - \omega_{ik} \tag{5.2}$$

$$v_{ik} = \mathrm{Var}\left[H_{ik}|\Theta_{ik} = \theta_{ik}\right] = \left(p_{ik} - p_{ik}^2\right)\left(\pi_{ik} - \theta_{ik}\lambda_{ik}m_{ik}\right)^2 + p_{ik}{m_{ik}}^2\theta_{ik}\lambda_{ik} \tag{5.3}$$

where we have assumed the claim severities being non-stochastic $X_i = m_i$ as this is the situation in the data application of Section 5.5.2. For further details, see the Appendix. Note that μ_{ik} and v_{ik} now depend on the latent random risk variable Θ_{ik} wherefore the correct notation is $\mu_{ik}\left(\Theta_{ik}\right)$ and $v_{ik}\left(\Theta_{ik}\right)$ but is omitted to ease the notation.

5.5.1 Model parameter estimation

We only briefly mention how the parameters in equation (5.2) and (5.3) can be obtained. The parameter p_{ik} is the customer specific probability of a successful cross-sale attempt (the customer purchases the offered policy). The sales probability is estimated using a regression model $\hat{p}_{ik} = \hat{f}_{p,k}\left(\mathcal{Y}_{p,ik}\right)$, where $\hat{f}_{p,k}$ is an appropriate regression function, estimated based on collateral data from the insurance company, collected from past cross-sale campaigns, and $\mathcal{Y}_{p,ik}$ is a set of customer specific covariates of the approached customer. Examples of such research and applications are the papers by Knott et al. (2002) and Li et al. (2005).

The a priori expected number of claims $\hat{\lambda}_{ik}$ and the a priori expected claim severity $\hat{m}_{ik}$ are estimated in conceptually the same way as the cross-sale probability $\hat{p}_{ik}$. The data used for the estimation of the regression functions $\hat{f}_{\lambda,k}$ and $\hat{f}_{m,k}$ is data on reported insurance claims from past and present customers of the company, for further details on how this is done, we refer to, e.g., Klugman et al. (1998). Once $\hat{f}_{\lambda,k}$ and $\hat{f}_{m,k}$ are estimated, the expected number of insurance claims and the expected severity can be estimated, for any customer, by only taking into consideration the sets of appropriate covariates $\mathcal{Y}_{\lambda,ik}$ and $\mathcal{Y}_{m,ik}$ for the specific customer i and policy k as $\hat{\lambda}_{ik} = e_{ik}\hat{f}_{\lambda,k}\left(\mathcal{Y}_{\lambda,ik}\right)$ and $\hat{m}_{ik} = \hat{f}_{m,k}\left(\mathcal{Y}_{m,ik}\right)$. The factor $0 \leq e_{ik} \leq 1$ measures the risk exposure and is equal to 0 if the customer i does not own a specific policy k. Note that the sets $\mathcal{Y}_{p,ik}$, $\mathcal{Y}_{\lambda,ik}$ and $\mathcal{Y}_{m,ik}$ are normally not identical since different covariates might be needed to explain the behaviour of the different stochastic variables $\mathbf{1}_{\{A_{ik}\}}$, N_{ik} and X_{ik}, respectively.

An estimate of the cost of a cross-sale attempt, ω_{ik} needs to be obtained from the company by analysing cost distributions, profit margins and overheads for the specific policy k, however the scope of this study does not allow us to discuss this in detail.

The risk profile parameter θ_{ik} can be seen as a factor for changing the a priori expected number of claims λ_{ik} since the conditional expectation of the number of insurance claims N_{ik} is $\mathbb{E}\left[N_{ik} \mid \Theta_{ik} = \theta_{ik}\right] = \lambda_{ik}\theta_{ik}$. Normally, the set of covariates $\mathcal{Y}_{\lambda,ik}$, needed for the regression function $\hat{f}_{\lambda,k}$, for the a priori expected number of claims $\hat{\lambda}_{ik}$, does not include information about past claiming of the specific customer i. Instead, $\mathcal{Y}_{\lambda,ik}$ usually contains covariates such as policy holder age, occupation, type of household, etc.. By assuming that an estimate of the risk profile $\hat{\theta}_{ik}$ can be expressed as a function of customer specific claim information we might obtain a better estimate of the number of insurance claims N_{ik} from the i-th customer. However, a specific problem related to cross-selling is that, obviously, no customer specific information is available, with respect to the cross-sold product, prior to approaching that specific customer. We solve this problem by estimating θ_{ik} with claim information of an existing policy k', of the specific customer, see Thuring (2012). Hence, we express $\hat{\theta}_{ik}$ as a function of the reported number of claims $n_{ik'}$ (with respect to an existing policy k') as well as the estimate of the a priori expected number of claims $\hat{\lambda}_{ik'}$, also with respect to the existing policy k', as $\hat{\theta}_{ik} = \hat{f}_{\theta,k}\left(n_{ik'}, \hat{\lambda}_{ik'}\right)$. We use multivariate credibility theory to estimate the function $f_{\theta,k}$ which results in the following

$$\hat{\theta}_{ik} = \hat{f}_{\theta,k}\left(n_{ik'}, \hat{\lambda}_{ik'}\right) = \hat{\theta}_{0k} + \frac{\hat{\lambda}_{ik'}\hat{\tau}^2_{kk'}}{\hat{\lambda}_{ik'}\hat{\tau}^2_{k'k'} + \hat{\sigma}^2_{k'}}\left(\frac{n_{ik'}}{\hat{\lambda}_{ik'}} - \hat{\theta}_{0k'}\right). \tag{5.4}$$

The model parameters $\hat{\theta}_{0k}$, $\hat{\tau}^2_{kk'}$, $\hat{\tau}^2_{k'k'}$, $\hat{\sigma}^2_{k'}$ and $\hat{\theta}_{0k'}$ need to be estimated based on a collateral data set consisting of claim information for customers owning both policy k and k'. We refer to the Appendix for details on the multivariate credibility estimation of $\hat{\theta}_{ik}$.

5.5.2 Real data application

We have a unique data set available, consisting of $I = 4463$ insurance customers who were targeted for a cross-sale campaign. The campaign was executed by approaching these specific customers, who at that time owned a household insurance coverage, and offering them to purchase a car insurance coverage. We acknowledge the risk of endogeneity related to using this kind of data, however we assume (as part of our model) that the latent random risk variable is independent of the indicator random variable for the event of cross selling. A formal test using the Fisher z-transform indicates that the assumption is valid. In the following we will refer to household coverage as coverage $k' = 1$ and car insurance coverage as coverage $k = 2$. Not every customer accepted the cross-sale offer, of the 4463 contacted household policyholders, 177 purchased the car insurance coverage, i.e. $\sum_{i=1}^{I} \mathbf{1}_{\{A_{i2}\}} = 177$. For these customers, the insurance company recorded the number of claims reported after the sale, with respect to the cross-sold policy (car insurance). With this data set available, we are able to estimate the customer specific expected profit $\hat{\mu}_{i2}$ (for the cross-sold coverage 2) and evaluate how closely related it is to the observed value h_{i2}, with h_{i2} being a realisation of the stochastic profit H_{i2} from representation (5.1). As a result of approaching all the 4463 customers, covered by the cross-sale campaign, the company recorded a total observed profit of $\sum_{i=1}^{I} h_{i2} = \$7,917$. It is interesting to analyse if the company could have executed the campaign with higher total profit by approaching fewer customers, taking the EP-criteria or MV-criteria into account.

In the expressions for the expected value of the profit (5.2) and its variance (5.3), we allow for customer specific values of all the included parameters, see Section 5.2. Unfortunately, the available data, from the cross-sale campaign, is not complete with respect to customer specific information about the premium (price) π_{ik}, the a priori expected number of insurance claims λ_{ik} or the observed claim severity x_{i2}, with x_{i2} being the realisation of the stochastic claim severity X_{i2} (note that index $k = 2$ refers to the cross-sale car insurance policy). Instead we use customer generic estimates $\hat{\pi}_2$, instead of $\hat{\pi}_{i2}$, $\hat{\lambda}_2$, instead of $\hat{\lambda}_{i2}$ and $\hat{m}_2$, instead of x_{i2} and m_{i2}. Also the cost of a cross-sale attempt is assumed to be a constant estimate ($\hat{\omega}_{i2} = \hat{\omega}_2$). The observed profits are customer dependent through the indicator variable $\mathbf{1}_{\{A_{i2}\}}$ and the customer dependent observed number of claims n_{i2} (which is a realisation of the stochastic variable N_{i2}).

Note that the estimated cross-sale probability $\hat{p}_{i2}$ and the estimate of the risk profile $\hat{\theta}_{i2}$ are customer specific. We estimate the model parameters $\hat{\theta}_{0k}$, $\hat{\tau}^2_{kk'}$, $\hat{\tau}^2_{k'k'}$, $\hat{\sigma}^2_{k'}$ and $\hat{\theta}_{0k'}$ (see (5.4)) based on a collateral data set from the insurance company consisting of claim information for customers owning both a household insurance policy and a car insurance policy. We use the closed form expressions of the parameter estimates found in Bühlmann and Gisler (2005, pp. 185-186). The resulting estimates are found in Table 3.

Table 3
Estimates of the model parameters for estimating the customer specific risk profile $\hat{\theta}_{i2}$.

l	$\hat{\sigma}^2_l$	$\hat{\tau}^2_{l1}$	$\hat{\tau}^2_{l2}$	$\hat{\theta}_{0l}$
1	1.755	0.081	0.130	1.12
2	1.349	0.130	0.211	0.91

In Table 4 we present summary statistics of the campaign data set of household customers approached for cross-selling car insurance.

Table 4
Descriptive statistics of the campaign data set, note that $k' = 1$ represents household insurance coverage and that $k = 2$ represents car insurance coverage.

	Constant	Min	Max	Mean
$\hat{\lambda}_{i1}$	-	0.0083	3.92	0.64
n_{i1}	-	0	20	1.17
$\hat{\theta}_{01}$	1.12	-	-	-
$\mathbf{1}_{\{A_{i2}\}}$	-	0	1	0.040
$\hat{p}_{i2}$	-	0.0040	0.13	0.069
$\hat{\theta}_{i2}$	-	0.71	2.05	0.95
n_{i2}	-	0	4	0.26
$\hat{\theta}_{02}$	0.91	-	-	-
$\hat{\lambda}_2$	0.375	-	-	-
$\hat{m}_2$ (\$)	$2,025$	-	-	-
$\hat{\pi}_2$ (\$)	949	-	-	-
$\hat{\omega}_2$ (\$)	15	-	-	-
$\hat{\mu}_{i2}$ (\$)	-	-54	25	1.03
$\hat{v}_{i2}$	-	$5.8 \cdot 10^3$	$3.0 \cdot 10^5$	$1.0 \cdot 10^5$
h_{i2} (\$)	-	$-7,166$	934	1.77

From Table 4 it can be seen that the expected number of household claims $\hat{\lambda}_{i1}$ has a very large spread and that one particular customer is associated with as much as $n_{i1} = 20$ household claims. Comparing the mean of $\hat{\lambda}_{i1}$ to the mean of n_{i1} shows that the customers have reported, on average, more claims than was expected which is also reflected in the estimate $\hat{\theta}_{01} > 1$. The mean value of $\mathbf{l}_{\{A_{i2}\}}$ is smaller than the mean value of $\hat{p}_{i2}$ meaning that the company expected to cross-sale car insurance coverage to more customer than was realised. The constant values of the common parameters representing the expected claim frequency $\hat{\lambda}_2$, the expected claim severity $\hat{m}_2$, the premium $\hat{\pi}_2$ and the cost of cross-selling $\hat{\omega}_2$, with respect to the car insurance coverage, are also given in Table 4. The values of these parameters are received from the insurance company and should be appropriate estimates for our particular situation. The estimate $\hat{\theta}_{02}$ is less than 1 meaning that customers are reporting fewer car insurance claims, on average, than the model, for the a priori number of car insurance claims, predicts. Note also that the estimate of the customer specific risk profile $\hat{\theta}_{i2}$ ranges between 0.71 and 2.05 meaning that it alters the conditional expectation of the number of claims N_{i2}, by between almost a 30% reduction to more than doubling it, keeping in mind the assumption that the conditional expectation of N_{i2} is $\mathbb{E}\left[N_{i2} \mid \Theta_{i2} = \theta_{i2}\right] = \lambda_2 \theta_{i2}$. It can be seen that the estimated expected profit $\hat{\mu}_{i2}$ can take both positive and negative values and that the realised profit h_{i2} has a large range; one customer is associated with a huge loss of $\$-7,166$ while at the other extreme the company made a profit of \$934 from one single customer.

We find that 2647 of the 4463 customers have a positive value of $\hat{\mu}_{i2}$. To illustrate how profit emerges from different customer selections we order the campaign data set, by non-increasing expected profit $\hat{\mu}_{i2}$, and compare cumulative sums for the expected profit $\sum_{i=1}^{l} \hat{\mu}_{i2}$ (referred to as the expected total profit) to cumulative sums of the observed profit $\sum_{i=1}^{l} h_{i2}$ (referred to as the observed total profit), for $l = 1, \ldots, 4463$. In Figure 5.1, we give the expected total profit as a function of the selection size l, note that the customers are ordered by non-increasing $\hat{\mu}_{ik}$ prior to cumulative summation and plotting. This is the total profit which would have been expected to emerge if the company had applied our proposed EP-criteria methodology. In Figure 5.1, we also present the observed total profit as a function of the same selection size l.

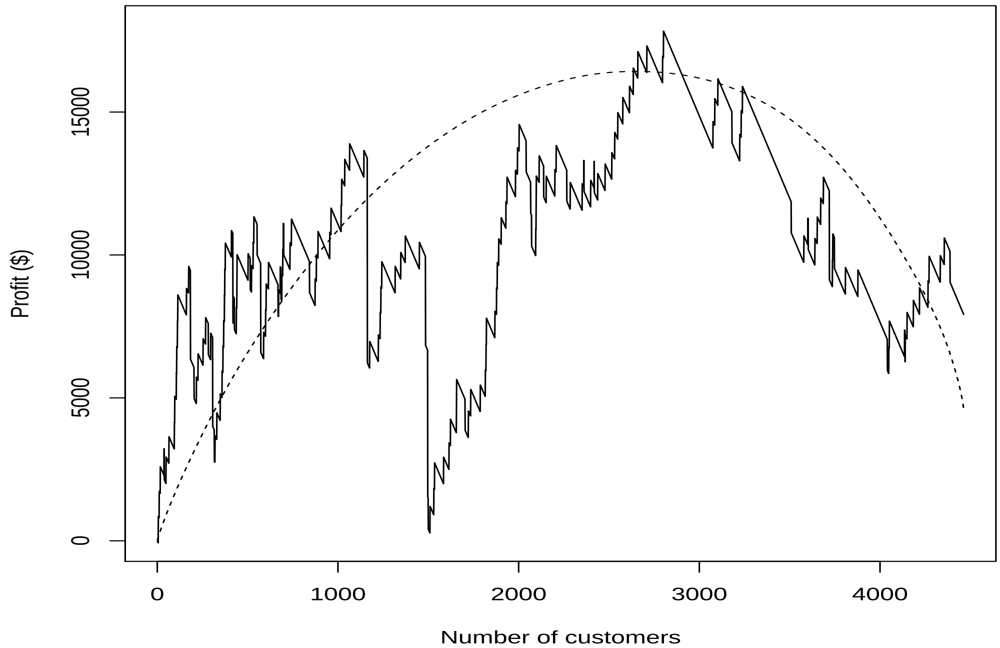

Figure 5.1: The expected total profit (dotted line) and the observed total profit (solid line), as cumulative sums, emerging from approaching an increasing number of customers l, with $l = 1, \ldots, 4463$. The customers are ordered by non-increasing expected profit $\hat{\mu}_{i2}$ prior to cumulative summation and plotting.

The sharp drop in the observed profit at approximately $l = 1500$ is due to three specific customers, for whom the estimate of the expected profit $\hat{\mu}_{i2}$ is reasonably high, whereas the observed profit is very low, due to 6 reported claims worth \$12, 150 in total. As can be seen, comparing the observed and the expected profit in Figure 5.1, the company would have made a profit of \$16, 424, by approaching only the prospects with a positive $\hat{\mu}_{i2}$. This is more than double the profit which the company made by approaching all of the 4463 customers (\$7, 917).

It is also interesting to compare the value of the total observed profit, \$16, 424, emerging from approaching customer with positive $\hat{\mu}_{i2}$, to the observed profit when approaching the 2647 customers associated with the largest estimates of the sales probability $\hat{p}_{i2}$. It is common to select prospects taking only the estimated sales probability $\hat{p}_{ik}$ into account and we find that these 2647 customers are associated with a total profit of \$7, 060. This is significantly less than the profit of \$16, 424 obtained when using the proposed EP-criterion.

For the second selection criteria, we select customer with positive mean-variance value MV_{i2} and show the resulting graph in Figure 5.2, where the customers are ordered by non-increasing MV_{i2} prior to plotting. The curve obviously depends of the value of ξ and we have tested a number of different values where $\xi = 5 \cdot 10^{-5}$ finally was chosen. It should be noted that the optimum is found at 1319, i.e. 1319 customers are associated with a positive mean-variance value (MV_{i2}).

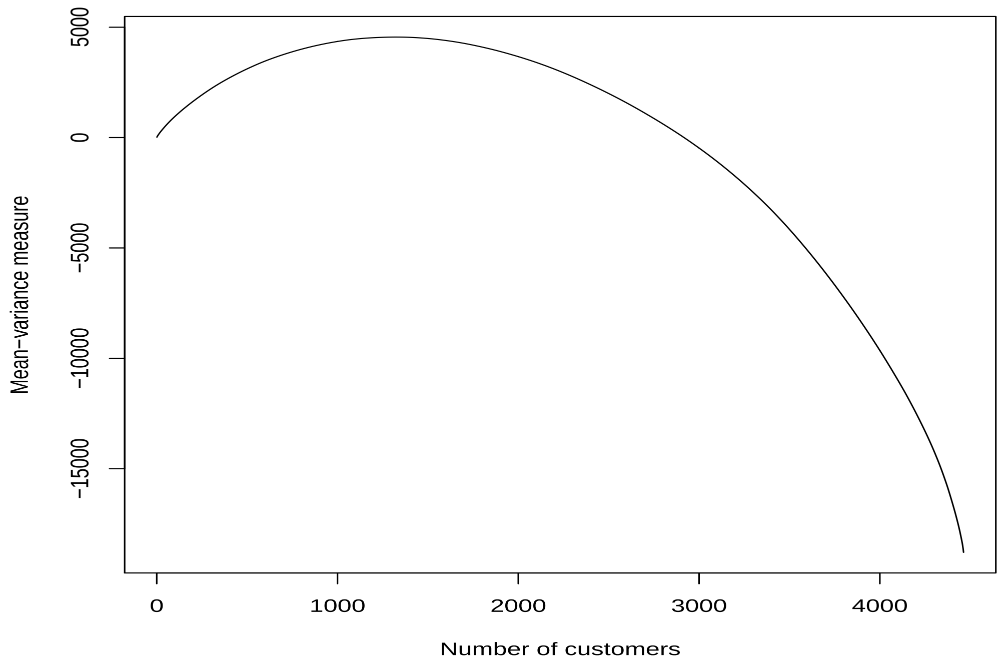

Figure 5.2: Mean-variance, as cumulative sums, emerging from approaching an increasing number of customers l, with $l = 1, \ldots, 4463$. The customers are ordered by non-increasing mean variance values MV_{i2} prior to cumulative summation and plotting.

We compare the two criteria (EP and MV) with respect to the expected total profit, the variance of the expected total profit and the observed total profit. As can be expected, looking at Table 5, under the EP-criterion the optimal selection size is higher and the expected profit is higher, whereas the MV-criterion has lower expected profit, but also lower profit variance. Of course, the total observed profit is lower for the MV-criterion, since it takes into account the profit variance.

Table 5

Summary of the results for the EP- and MV-criteria.

Criteria	Number of customers	Expected total profit	Variance of total profit	Observed total profit
EP	2647	$16,424	$3.0 \cdot 10^8$	$16,362
MV	1319	$12,787	$1.6 \cdot 10^8$	$3,882

5.6 Concluding remarks

In this paper, we have introduced a new flexible approach to optimal cross selling. We solve the optimization problem of maximizing both an optimal mean criteria and a mean-variance criterion. Our profit/risk performance optimization approach has, to the best of our knowledge, not been previously considered in the context of cross-sales marketing.

For the purpose of solving the proposed optimization problems, we have developed a stochastic model of the profit, emerging from a successful cross-sale to an individual prospect and a group of prospects. The model is expressed in terms of certain random variables, characterizing the occurrence of sale, the price and the cost. When trying our methodology out on real data (we consider a large insurance data set) we get practical and convincing answers suggesting potential cross sale strategies. Further dynamics of the model could be considered, e.g. allowing for the probability of cross-sale p_{ik} to be dependent of the price Π_{ik}, in (5.1). Such extensions would introduce the concept of dynamic pricing in the cross-sale selection methodology. While this is outside the scope of this paper it is currently our focus for further research and we have started an extended data collection exercise in collaboration with our non-life insurance contact that eventually will enable us to introduce dynamic pricing to our flexible cross-sale model. Notice, that dynamic pricing will introduce a less linear and more complex optimization algorithm, probably of a recursive nature. It will be part of our future research to provide stable algorithms for this new challenging optimization.

In Section 5.5.2, we have validated the proposed methodology based on a real data set from a large insurance company. As our validation results demonstrate, the proposed methodology is capable of providing appropriate optimal selections of customers, so that the expected profit/mean-variance criterion is maximized. This is confirmed in the data study, where the observed profit is volatile but follows the expected (see Section 5.5.2). In conclusion, we confirm that the proposed profit optimization methodology has been successfully validated and, as demonstrated, is practically applicable for the purpose of profit efficient cross-selling of financial services products.

Appendix

Derivation of the expected profit μ_{ik} and variance v_{ik}

For simplicity we omit the index k. The proof of (5.2) is straightforward and is omitted. We denote by η_i the variance of X_i. For the variance v_i of H_i, noting that the r.v.s $\mathbf{1}_{\{A_i\}}$ and S_i are assumed independent, we have

$$v_i = \mathrm{Var}\left[H_i \mid \Theta_i = \theta_i\right] = \mathrm{Var}\left[\mathbf{1}_{\{A_i\}}\left(\pi_i - S_i\right) \mid \Theta_i = \theta_i\right] =$$

$$\mathrm{Var}\left[\mathbf{1}_{\{A_i\}}\left(\pi_i - \sum_{n=1}^{N_i} X_{in}\right) \mid \Theta_i = \theta_i\right] =$$

$$= \mathbb{E}\left[\left(\mathbf{1}_{\{A_i\}}\left(\pi_i - \sum_{n=1}^{N_i} X_{in}\right)\right)^2 \mid \Theta_i = \theta_i\right] - \left(\mathbb{E}\left[\mathbf{1}_{\{A_i\}}\left(\pi_i - \sum_{n=1}^{N_i} X_{in}\right) \mid \Theta_i = \theta_i\right]\right)^2 =$$

$$= \mathbb{E}\left[\mathbf{1}^2_{\{A_i\}}\left(\pi_i^2 - 2\pi_i\sum_{n=1}^{N_i} X_{in} + \left(\sum_{n=1}^{N_i} X_{in}\right)^2\right) \mid \Theta_i = \theta_i\right] -$$

$$\left(\mathbb{E}\left[\mathbf{1}_{\{A_i\}}\left(\pi_i - \sum_{n=1}^{N_i} X_{in}\right) \mid \Theta_i = \theta_i\right]\right)^2 =$$

$$= p_i\left(\pi_i^2 - 2\pi_i\lambda_i\theta_i m_i + \lambda_i\theta_i\eta_i + \lambda_i\theta_i m_i^2 + \lambda_i^2\theta_i^2 m_i^2\right) - p_i^2\left(\pi_i - \theta_i\lambda_i m_i\right)^2$$

which simplifies to 5.3 by assuming that X_i is non-stochastic. In the derivation we have used that

$$\mathbb{E}\left[\left(\sum_{n=1}^{N_i} X_{in}\right)^2 \mid \Theta_i = \theta_i\right] = \mathrm{Var}\left[S_i \mid \Theta_i = \theta_i\right] + \mathbb{E}\left[S_i \mid \Theta_i = \theta_i\right]^2 =$$

$$= \mathbb{E}\left[N_i \mid \Theta_i\right]\mathrm{Var}\left[X_i \mid \Theta_i\right] + \mathrm{Var}\left[N_i \mid \Theta_i\right]\mathbb{E}\left[X_i \mid \Theta_i\right]^2 + \mathbb{E}\left[N_i \mid \Theta_i\right]^2\mathbb{E}\left[X_i \mid \Theta_i\right]^2 =$$

$$\lambda_i\theta_i\eta_i + \lambda_i\theta_i m_i^2 + \lambda_i^2\theta_i^2 m_i^2$$

where we have used the shorter notation for conditioning on Θ_i and that the variance of the aggregate loss S_i, assuming independence between N_i and X_i, is $\mathrm{Var}\left[S_i\right] = \mathbb{E}\left[N_i\right]\mathrm{Var}\left[X_i\right] + \mathrm{Var}\left[N_i\right]\mathbb{E}\left[X_i\right]^2$, see e.g. Mikosch (2009) p. 73.

Derivation of the cumulative distribution function of H_i

Formulas (5.2) and (5.3) are useful in establishing the mean and variance of the total profit. In order to gain further insight into the way profit emerges as a result of cross-selling of an additional policy to the i-th policyholder, in the following proposition, we give the cumulative distribution function of H_i, conditional on $\Theta_i = \theta_i$.

Proposition 2 Given $\Theta_i = \theta_i$, the cumulative distribution function, $F_{H_i}(x)$, is

$$F_{H_i}(x) = P\left(H_i \leq x \mid \Theta_i = \theta_i\right) = \begin{cases} 1 & \text{if } x \geq \pi_i - \omega_i \\ 1 - p_i \sum_{j=0}^{[[\tilde{x}]]} e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^j}{j!} & \text{if } -\omega_i \leq x < \pi_i - \omega_i \\ \left(1 - \sum_{j=0}^{[[\tilde{x}]]} e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^j}{j!}\right) p_i & \text{if } x < -\omega_i \end{cases} \tag{5.5}$$

where $\tilde{x} = \frac{\pi_i - \omega_i - x}{m_i}$ and $[[\tilde{x}]] = \begin{cases} [\tilde{x}] & \text{if } \tilde{x} \text{ is non} - \text{integer} \\ \tilde{x} - 1 & \text{if } \tilde{x} \text{ is integer} \end{cases}$ and $[\tilde{x}]$ is the integer part of $\tilde{x}$.

Proof We have

$$\begin{aligned} P\left(H_i \leq x\right) = P\left(\mathbf{1}_{\{A_i\}}\left(\pi_i - N_i m_i\right) - \omega_i \leq x\right) = \\ = P\left(\mathbf{1}_{\{A_i\}}\left(\pi_i - N_i m_i\right) \leq x + \omega_i | \mathbf{1}_{\{A_i\}} = 1\right) p_i + \\ P\left(\mathbf{1}_{\{A_i\}}\left(\pi_i - N_i m_i\right) \leq x + \omega_i | \mathbf{1}_{\{A_i\}} = 0\right)\left(1 - p_i\right) = \\ = P\left(N_i \geq \tilde{x}\right) p_i + P(0 \leq x + \omega_i)\left(1 - p_i\right) = \\ = \left(1 - P\left(N_i < \tilde{x}\right)\right) p_i + P(0 \leq x + \omega_i)\left(1 - p_i\right) \end{aligned} \tag{5.6}$$

where we have used the independence of the r.v.s $\mathbf{1}_{\{A_i\}}$ and N_i. Representation (5.5) follows from (5.6), recalling that, conditional on $\Theta_i = \theta_i$, $N_i \sim \text{Poisson}\left(\theta_i \lambda_i\right)$. $\square$

Let us note that, if π_i is not a multiple of m_i, i.e. $\pi_i \neq r m_i$, for r, positive integer, the set of values, the random variable, H_i can take is:

$$\begin{aligned} \text{Im} H_i = \{x_j = \pi_i - \omega_i - j m_i, j = 0, 1, \ldots, j^*, \\ x_{j^*+1} = -\omega_i, x_j = \pi_i - \omega_i - (j-1) m_i, j = j^* + 2, j^* + 3, \ldots\} \end{aligned} \tag{5.7}$$

where j^* is such that, $\pi_i - j^* m_i > 0$ and $\pi_i - \left(j^* + 1\right) m_i < 0$. If π_i is a multiple of m_i, i.e. $\pi_i = j^* m_i$, where j^* is a suitable positive integer, then

$$\mathrm{Im} H_i = \{x_j = \pi_i - \omega_i - j m_i, j = 0, 1, \ldots, j^* - 1, j^* + 1, j^* + 2, \ldots, x_{j^*} = -\omega_i\}. \tag{5.8}$$

Derivation of the probability mass function of H_i

From Proposition 2, it is straightforward to derive the conditional p.m.f. $P(H_i = x_j \mid \Theta_i = \theta_i)$, $j = 1, 2, \ldots$.

Proposition 3 Given $\Theta_i = \theta_i$, and

1. Assuming that $\mathrm{Im} H_i$ is as in (5.7), the probability mass function of H_i is

$$P(H_i = x_j \mid \Theta_i = \theta_i) = \begin{cases} p_i e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^j}{j!} & \text{for } j = 0, 1, \ldots, j^* \\ 1 - p_i & \text{for } j = j^* + 1 \\ p_i e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^{j-1}}{(j-1)!} & \text{for } j = j^* + 2, j^* + 3, \ldots \end{cases} \tag{5.9}$$

2. Assuming that $\mathrm{Im} H_i$ is as in (5.8), the probability mass function of H_i is

$$P(H_i = x_j \mid \Theta_i = \theta_i) = \begin{cases} p_i e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^j}{j!} & \text{for } j = 0, 1, \ldots, j^* - 1 \\ 1 - p_i + p_i e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^j}{j!} & \text{for } j = j^* \\ p_i e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^j}{j!} & \text{for } j = j^* + 1, j^* + 2, \ldots \end{cases} \tag{5.10}$$

Proof Formulas (5.9) and (5.10) follow directly from (5.5) noting that, for assumption 1. (formula (5.9)), by the definition of j^* in (5.7), we have that $j^* < \frac{\pi_i}{m_i} < j^* + 1$, hence $\left[\left[\frac{\pi_i}{m_i}\right]\right] = j^*$, and for assumption 2. (formula (5.10)) by the definition of j^* in (5.8) we have that $\frac{\pi_i}{m_i} = j^*$, hence $\left[\left[\frac{\pi_i}{m_i}\right]\right] = j^* - 1$. $\square$

Distributional properties of the total profit $\mathcal{H}_s(l)$

The c.d.f., $F_{H_i}(x)$ and the p.m.f., $P(H_i = x_j \mid \Theta_i = \theta_i)$, given in Propositions 2, and 3 embeds the entire information about the behaviour of the profit, H_i emerging from the i-th prospect. Therefore (5.5), (5.9) and (5.10) are useful in addressing some further questions, related to the profitable marketing of financial services products. One such important question which we will address in this section is to provide confidence bounds for the total profit from a cross-sales campaign.

We are now in a position to consider the total profit, $\mathcal{H}_s(l)$, related to a subset, $s(l) \subset \mathcal{P}$ of size l, which is

$$\mathcal{H}_s(l) = \sum_{i=1}^{l} H_i = \sum_{i=1}^{l} \left(\mathbf{1}_{\{A_i\}} \left(\pi_i - N_i m_i \right) - \omega_i \right). \tag{5.11}$$

Given $\Theta = \theta$, the total expected profit, $\mathbb{E}\left[\mathcal{H}_s(l) \mid \Theta = \theta\right]$, related to a subset, $s(l) \subset \mathcal{P}$ of size l, is

$$\mathbb{E}\left[\mathcal{H}_s(l) \mid \Theta = \theta\right] = \sum_{i=1}^{l} \mathbb{E}\left[H_i | \Theta_i = \theta_i\right] = \sum_{i=1}^{l} \left(p_i \left(\pi_i - \theta_i \lambda_i m_i \right) - \omega_i \right), \tag{5.12}$$

and the conditional variance, $\mathrm{Var}_s(l)$, of the total profit, $\mathcal{H}_s(l)$ from a subset, $s(l) \subset \mathcal{P}$ of size l, given $\Theta = \theta$ is

$$\mathrm{Var}_s(l) = \sum_{i=1}^{l} \mathrm{Var}\left[H_i \mid \Theta_i = \theta_i\right] = \sum_{i=1}^{l} \left(\mathrm{Var}\left[\mathbf{1}_{\{A_i\}}\right] \left(\pi_i - \theta_i \lambda_i m_i \right)^2 + p_i {m_i}^2 \theta_i \lambda_i \right). \tag{5.13}$$

Clearly, one way in which the company may deal with the contradictory goals of maximizing its expected profit while minimizing the related risk is to maximize the total (expected) cross-sales profit and minimize its variance by combining the two quantities in a common mean-variance criterion.

Given the distribution of H_i, conditional on $\Theta = \theta$, the conditional distribution of $\mathcal{H}_s(l)$ is obtained as the following convolution

Proposition 4 Given $\Theta = \theta$, the p.m.f. of $\mathcal{H}_s(l)$ is

$$\begin{aligned} P\left(\mathcal{H}_s(l) = h \mid \Theta = \theta\right) = \sum_{x_1 \in \mathrm{ImH}_1} \dots \sum_{x_{l-1} \in \mathrm{ImH}_{l-1}} P\left(H_1 = x_1 \mid \Theta_1 = \theta_1\right) \times \dots \\ \dots \times P\left(H_{l-1} = x_{l-1} \mid \Theta_{l-1} = \theta_{l-1}\right) P\left(H_l = h - x_1 - \dots - x_{l-1} \mid \Theta_l = \theta_l\right), \end{aligned} \tag{5.14}$$

where $h \in D$, $D = \{x_1 + \ldots + x_l : (x_1, \ldots, x_l) \in \{\mathrm{Im}H_1 \times \ldots \times \mathrm{Im}H_l\}\}$.
Based on (5.14), for the cdf $F_{\mathcal{H}_s(l)}(x) = P\left(\mathcal{H}_s(l) \le x \mid \Theta = \theta\right)$ we have

Proposition 5 Given $\Theta = \theta$, the c.d.f. of $\mathcal{H}_s(l)$ is

$$F_{\mathcal{H}_s(l)}(x) = P\left(\mathcal{H}_s(l) \leq x \mid \Theta = \theta\right) =$$

$$\sum_{h \in D, h \leq x} \sum_{x_1 \in \mathrm{ImH}_1} \cdots \sum_{x_{l-1} \in \mathrm{ImH}_{l-1}} P\left(H_1 = x_1 \mid \Theta_1 = \theta_1\right) \times \ldots \tag{5.15}$$

$$\ldots \times P\left(H_{l-1} = x_{l-1} \mid \Theta_{l-1} = \theta_{l-1}\right) P\left(H_l = h - x_1 - \ldots - x_{l-1} \mid \Theta_l = \theta_l\right),$$

where $x \in \mathbb{R}$ and D is defined as in Proposition 4.

Proposition 5 can be used in order to produce confidence intervals for the total profit, $\mathcal{H}_s(l)$, of the form

$$P\left(Q_{\frac{\alpha}{2}} \leq \mathcal{H}_s(l) \leq Q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha, \tag{5.16}$$

where $Q_{\frac{\alpha}{2}}$ and $Q_{1-\frac{\alpha}{2}}$ are the corresponding $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of the distribution $F_{\mathcal{H}_s(l)}$. The latter quantiles, $Q_{\frac{\alpha}{2}} = F^{-1}_{\mathcal{H}_s(l)}\left(\frac{\alpha}{2}\right)$ and $Q_{\frac{\alpha}{2}} = F^{-1}_{\mathcal{H}_s(l)}\left(1 - \frac{\alpha}{2}\right)$, where $F^{-1}_{\mathcal{H}_s(l)}(\cdot)$ is the inverse of $F_{\mathcal{H}_s(l)}$. Computing, $P\left(\mathcal{H}_s(l) = h\right)$, $F_{\mathcal{H}_s(l)}(x)$ and $F^{-1}_{\mathcal{H}_s(l)}(\cdot)$ using (5.14) and (5.15) is, facilitated by the reasonably simple form of $F_{H_i}(x)$ and $P\left(H_i = x_j \mid \Theta_i = \theta_i\right)$, $j = 1, 2, \ldots$ which stems from the assumption that N_i has a conditional Poisson distribution. Therefore, confidence intervals of the form (5.16) can be easily computed for small, up to moderate portfolio sizes, I. For large values of I, which is often the case in practice, representations (5.14) and (5.15) may become cumbersome to evaluate and it is important to consider asymptotic approximations of the distribution of $\mathcal{H}_s(l)$. We show that, under some conditions on the model parameters, θ_i, λ_i and m_i, the distribution of the appropriately normalized total profit, $\mathcal{H}_s(l)$, converges to a standard normal distribution, as the size, l goes to infinity. This result can be used in order to provide approximate confidence regions for the total profit, for large portfolio sizes l.

In what follows, it will be convenient to use the simpler notation, C_l B_l^2, for the mean $\mathbb{E}\left[\mathcal{H}_s(l) | \Theta = \theta\right]$ and the variance, $\mathrm{Var}_s(l)$, respectively. We will also assume that the real positive parameters, λ_i, θ_i, and m_i, $i = 1, 2, \ldots$ are such that the Lindeberg condition

$$\frac{1}{B_l^2} \sum_{k=1}^{l} \sum_{\{j : |x_j - \mathbb{E}(H_k)| > \varepsilon B_l\}} P\left(H_j = x_j\right)\left(x_j - \mathbb{E}\left(H_k\right)\right)^2 \underset{l \to \infty}{\longrightarrow} 0 \tag{5.17}$$

holds. Let us note that there exists a set of values for the parameters, λ_i, θ_i and m_i

$i = 1, 2, \ldots$, such that, H_i, $i = 1, 2, \ldots$ form a sequence of independent identically distributed random variables, in which case (5.17) holds, i.e., the set of values for which condition (5.17) is fulfilled is not empty. Since in general, H_i, $i = 1, 2, \ldots$ are independent, non-identically distributed random variables, with c.d.f.s, $F_{H_i}(x)$, $i = 1, 2, \ldots$, following the Lindeberg-Feller central limit theorem one can state

Proposition 6 Given that, λ_i, θ_i, and m_i, are such that the Lindeberg condition (5.17) holds, the distribution functions of the normalized total profit,$\left(\mathcal{H}_s(l) - C_l\right)/B_l$ tend to a standard normal cdf, as l tends to infinity.

Proposition 6 allows for the construction of approximate confidence regions, of the form (5.16), for the total profit random variable, $\mathcal{H}_s(l)$, when l is sufficiently large, given that (5.17) holds. For a given confidence level, α, we have that

$$P\left(q_{\frac{\alpha}{2}} \leq \left(\mathcal{H}_s(l) - C_l\right)/B_l \leq q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha, \tag{5.18}$$

where $q_{\frac{\alpha}{2}}$, and $q_{1-\frac{\alpha}{2}}$ are the corresponding quantiles of the standard normal distribution. From (5.18), for $\alpha = 0.05$ we have that, $P\left(C_l - 1.96B_l \leq \mathcal{H}_s(l) \leq C_l + 1.96B_l\right) = 0.95$. We acknowledge that these approximate confidence regions are in fact too optimistic since we are using an estimate of the latent risk profile θ_i, and not an observed value, to condition on. Since $\hat{\theta}_i$ is an estimator it too has a variance that would broaden the approximate confidence regions if taken into consideration. However, this issue is neglected in the derivation above.

Estimation of the latent risk profile θ_{ik}

In this section we re-introduce the product index k. In order to estimate θ_{ik}, one could apply an estimator motivated by the classical credibility theory and in particular by the Bühlmann-Straub credibility model (see Bühlmann (1967) and Bühlmann and Straub (1970)). A similar estimator, but in the context of insurance pricing, has been applied by Englund et al. (2008) and Englund et al. (2009). We assume that $\Theta_{il}, \ldots, \Theta_{Il}$ are i.i.d. random variables with $E\left[\Theta_{il}\right] = \theta_{0l}$, $i = 1, \ldots, I$ and $Cov\left[\Theta_{il}, \Theta_{ir}\right] = \tau^2_{lr}$, $l, r \in \{k', k\}$. We further assume that the conditional covariance structure of the random variables $F_{ijl} = \frac{N_{ijl}}{\lambda_{ijl}}$, $l \in \{k', k\}$ is given by

$$Cov\left[F_{ijl}, F_{ijr} \mid \Theta_{il} = \theta_{il}, \Theta_{ir} = \theta_{ir}\right] = \begin{cases} \frac{\sigma_l^2(\theta_{il})}{\lambda_{ijl}} & \text{if } l = r \\ 0 & \text{if } l \neq r \end{cases},$$

and $\sigma_l^2(\theta_{il})$ is the variance within a specific customer i for $l \in \{k', k\}$. We use standard credibility notation and define $\lambda_{i\cdot l} = \sum\limits_{j=1}^{J_i} \lambda_{ijl}$, $n_{i\cdot l} = \sum\limits_{j=1}^{J_i} n_{ijl}$ and $F_{i\cdot l} = \frac{n_{i\cdot l}}{\lambda_{i\cdot l}}$. Under these assumptions, it is possible to generalize the univariate Bühlmann-Straub homogeneous estimator of the standardized frequency θ_{ik} (see corollary 4.10 of Bühlmann and Gisler (2005), p. 102) to our two dimensional setting as

$$\hat{\theta}_i = \theta_0 + \alpha_i \left(F_{i\cdot} - \theta_0\right) \tag{5.19}$$

with $\hat{\theta}_i = \left[\hat{\theta}_{i1}\hat{\theta}_{i2}\right]'$, $\theta_0 = [\theta_{01}\theta_{02}]'$ and $F_{i\cdot} = [F_{i\cdot 1}F_{i\cdot 2}]'$. The credibility weight $\alpha_i = T\Lambda_i(T\Lambda_i + S)^{-1}$ where T is a 2 by 2 matrix with elements $\tau_{kk'}^2$, $k = 1, 2$ and $k' = 1, 2$. The matrices Λ_i and S are diagonal matrices with, respectively, $\lambda_{i\cdot l}$, $l = 1, 2$ and σ_l^2, $l = 1, 2$ in the diagonal and $\lambda_{i\cdot l} = \sum_{j=1}^{J_i} \lambda_{ijl}$. The parameter $\sigma_l^2 = E\left[\sigma_l^2(\theta_{il})\right]$, where $\sigma_l^2(\theta_{il})$ is the variance within an individual customer i, for a product l (for further details see Bühlmann and Gisler, 2005, p. 81). We also refer to Bühlmann and Gisler (2005, pp. 185-186) for parameter estimation procedures of the matrices S and T and the vector θ_0.

Performing the matrix multiplication in (5.19) and considering element 2 of $\hat{\theta}_i$ we get

$$\hat{\theta}_{i2} = \theta_{02} + \alpha_{i22}\left(F_{i\cdot 2} - \theta_{02}\right) + \alpha_{i21}\left(F_{i\cdot 1} - \theta_{01}\right). \tag{5.20}$$

where $\alpha_{ikk'}$ is element kk' of the matrix α_i.

We now assume that if product 2 is not active (not owned) by customer i, the risk exposure $e_{ij2} = 0$ for all j and consequently $\hat{\lambda}_{ij2} = \hat{\lambda}_{i\cdot 2} = 0$. It is possible to show that $\hat{\lambda}_{i\cdot 2} = 0$ implies that $\hat{\alpha}_{i22} = 0$ and (5.20) becomes

$$\hat{\theta}_{i2} = \hat{\theta}_{02} + \hat{\alpha}_{i21}\left(\hat{F}_{i\cdot 1} - \hat{\theta}_{01}\right),$$

where $\hat{\alpha}_{i21} = \frac{\hat{\lambda}_{i\cdot 1}\hat{\tau}_{21}^2}{\hat{\lambda}_{i\cdot 1}\hat{\tau}_{11}^2 + \hat{\sigma}_1^2}$. This shows that even though a customer i does not have an active product 2, it is possible to obtain an estimate of his/her specific risk profile

$\hat{\theta}_{i2}$ (with respect to product 2) by using data of $\hat{F}_{i\cdot 1} = \frac{n_{i\cdot 1}}{\hat{\lambda}_{i\cdot 1}}$ with respect to the other (owned) product 1.

References

Akerlof, G. A., 1970, The market for 'lemons': Quality uncertainty and the market mechanism, *The Quarterly Journal of Economics*, 84(3), 488-500.

Beard, R. E., Pentikinen, T. and Pesonen, E., 1984, *Risk theory*, Chapman & Hall, London.

Bult, J.R. and Wansbeek, T., 1995, Optimal selection for direct mail, *Marketing Science*, 14(4), 378-394.

Bühlmann, H., 1967, Experience rating and credibility, *Astin Bulletin*, 4(3), 199-207.

Bühlmann, H. and Straub, E., 1970, Glaubwürdigkeit für Schadensätze, *Bulletin of Swiss Association of Actuaries*, 70(1), 111-133.

Bühlmann, H. and Gisler, A., 2005, *A Course in Credibility Theory and its Applications* (Berlin, Germany: Springer Verlag).

Cohen, A., 2005, Asymmetric information and learning: Evidence from the automobile insurance market. *The Review of Economics and Statistics* 87(2), 197-207.

Englund, M., Guillén, M., Gustafsson, J., Nielsen, L.H. and Nielsen, J.P., 2008, Multivariate latent risk: A credibility approach, *Astin Bulletin*, 38(1), 137-146.

Englund, M., Gustafsson, J., Nielsen, J. P. and Thuring, F., 2009, Multidimensional Credibility with Time Effects - An Application to Commercial Business Lines, *The Journal of Risk and Insurance*, 76(2), 443-453.

Gurvich, I., Armony, M. and Maglars, C., 2009, Cross-selling in a call center with a heterogeneous customer population, *Operations Research*, 57(2), 299-313.

Gönül, F.F. and Hofstede, F.T., 2006, How to compute optimal catalog mailing decisions, *Marketing Science*, 25(1), 65-74.

Jong, P., Heller, G., Z., 2008, *Generalized linear models for insurance data*, Cambridge University Press.

Kaishev, V.K., Nielsen, J.P., Thuring, F., 2013, Optimal customer selection for cross-selling of financial services products, *Expert Systems with Applications*, 40(5), 1748-1757.

Kamakura, W.A., 2007, Cross-Selling: Offering the Right Product to the Right Customer at the Right Time in Lehzan Aksoy, Timothy Keiningham and David Bejou (Eds.) Profit Maximization through Customer Relationship Marketing , Haworth Press.

Kamakura, W. A., Kossar, B. and Wedel, M., 2004, Identifying innovators for the cross-selling of new products, *Management Science*, 50, 1120-1133.

Kamakura, W. A., Wedel, M., de Rosa, F. and Mazzon, J. A., 2003, Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction, *International Journal of Research in Marketing*, 20(1), 45-65.

Kamakura, W. A., Ramaswami, S. and Srivastava R., 1991, Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services, *International Journal of Research in Marketing*, 8, 329-349.

Klugman, S. A., Panjer, H. H. and Willmot, G. E., 1998, *Loss models: From data to decisions*, Wiley & Sons, New York

Knott A., Hayes, A. and Neslin, S.A., 2002, Next-product-to-buy models for cross-selling applications, *Journal of Interactive Marketing*, 16(3), 59-75.

Li, S., Sun, B. and Montgomery, A.L. 2010, Cross-Selling the Right Product to the Right Customer at the Right Time, *Journal of Marketing Research*, 48(4), 683-700.

Li, S., Sun, B. and Wilcox, R.T., 2005, Cross-selling sequentially ordered products: an application to consumer banking services, *Journal of Marketing Research*, 42, 233-239.

Malthouse, E. C., 1999, Ridge regression and direct marketing scoring models, *Journal of Interactive Marketing*, 13(4), 10-23.

Mikosch, T., 2009, *Non-Life Insurance Mathematics: An Introduction with the Poisson Process* (Berlin, Germany: Springer Verlag).

Ohlsson, E., Johansson, B., 2010, *Non-life insurance pricing with generalized linear models*, Springer Verlag, Berlin.

Rossi, P. E. and Allenby, G. M, 2003, Bayesian Statistics and Marketing, *Marketing Science*, 22(3), 304-328.

Rothschild, M. and Stiglitz J., 1976, Equilibrium in competitive insurance markets: An essay on the economics of imperfect information, *The Quarterly Journal of Economics* 90(4), 629-649.

Thuring, F., 2012, A credibility method for profitable cross-selling of insurance products, *Annals of Actuarial Science*, 6(1), 65-75.

Venkatesan, R. and Kumar, V., 2004, A customer lifetime value framework for customer selection and resources allocation strategy, *Journal of Marketing*, 68, 106-125.

Declaration

I herewith declare that I have produced this dissertation without assistance of any third parties other than the co-authors of the papers. Additionally, without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This dissertation has not previously been presented in identical or similar form to any other UK or foreign examination board.

The thesis work was conducted from October 2008 to December 2012 under the supervision of Prof. Vladimir Kaishev and Prof. Jens Perch Nielsen at Sir John Cass Business School, City University London.

Fredrik Thuring