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.

1 Introduction

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

1

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. We acknowledge cross-sale models as being part of companies' customer segmentation efforts and refer to studies such as Hwang et al. (2004), Jonker et al. (2004), Lee and Park (2005), Larivire and Van den Pol (2005) and Kim et al. (2006) which would be possible to extend towards a stochastic profit approach, as described in this paper.

In our concrete example, we use recently developed actuarial technology based on multivariate credibility theory to assess the individual specifics in case

2

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

3

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 2 we propose a stochastic model for the profit associated with cross-selling an additional product to an existing customer. Section 3 elaborates on two established methods for capturing customer heterogeneity and how they are combined in this paper. In Section 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, we study an example of cross-selling insurance policies to existing customers of an insurance company. Concluding remarks are found in Section 6 followed by an appendix with details on results from the insurance example.

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.

4

In Section 4 we give a wide array of possible situations where a profit formula might be of interest.

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

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

where $\mathbf{l}_{\{A_{ik}\}}$ 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 (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 $v_{ik} = \text{Var}\left[H_{ik}\right]$ 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 2). Note that any correlation between $\mathbf{l}_{\{A_{ik}\}}$ 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).

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.

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 is this particular case.

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.

3.3 Combining multiple regression analysis with latent variable models

For our model, for the stochastic profit H_{ik} (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 5.2, how this data can be useful and improve the overall profit from cross-selling.

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

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.

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. Time-varying effects of customer loyalty is considered in Guillén et al. (2012).

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   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).

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 5.2 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

$$\mu_{ik} = \mathbb{E}\left[H_{ik}|\Theta_{ik} = \theta_{ik}\right] = p_{ik}\left(\pi_{ik} - \theta_{ik}\lambda_{ik}m_{ik}\right) - \omega_{ik} \tag{2}$$

$$v_{ik} = \mathrm{Var}\left[H_{ik}|\Theta_{ik} = \theta_{ik}\right] = \left(p_{ik} - p_{ik}^2\right)\left(\pi_{ik} - \theta_{ik}\lambda_{ik}m_{ik}\right)^2 + p_{ik}m_{ik}^2\theta_{ik}\lambda_{ik} \tag{3}$$

For further details, see the Appendix.

5.1 Model parameter estimation

We only briefly mention how the parameters in equation (2) and (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 $\hat{\lambda}_{ik} = e_{ik}\hat{f}_{\lambda,k}(\mathcal{Y}_{\lambda,ik})$ and $\hat{m}_{ik} = \hat{f}_{m,k}(\mathcal{Y}_{m,ik})$. The factor $0 \le e_{ik} \le 1$ 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 $\mathbb{E}[N_{ik} \mid \Theta_{ik} = \theta_{ik}] = \lambda_{ik}\theta_{ik}$. 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) and Thuring et al. (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 $\hat{\theta}_{ik} = \hat{f}_{\theta,k}(n_{ik'}, \hat{\lambda}_{ik'})$. We use multivariate credibility theory to estimate the function $f_{\theta,k}$ which results in the following

$$\hat{\theta}_{ik} = \hat{f}_{\theta,k}\left(n_{ik'}, \hat{\lambda}_{ik'}\right) = \hat{\theta}_{0k} + \frac{\hat{\lambda}_{ik'}\hat{\tau}_{kk'}^2}{\hat{\lambda}_{ik'}\hat{\tau}_{k'k'}^2 + \hat{\sigma}_{k'}^2}\left(\frac{n_{ik'}}{\hat{\lambda}_{ik'}} - \hat{\theta}_{0k'}\right). \qquad (4)$$

The model parameters $\hat{\theta}_{0k}$, $\hat{\tau}_{kk'}^2$, $\hat{\tau}_{k'k'}^2$, $\hat{\sigma}_{k'}^2$ 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 $\hat{\theta}_{ik}$.

5.2 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 (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.

We focus, for a moment, on the conditional Poisson assumption of claim counts. As mentioned above, had the unobserved latent variable been gamma distributed then the resulting counts would be negative binomial distributed. To validate this assumptions, we therefore tested our counts towards the relevant negative binomial distribution. The test was rejected at a very low significance level. However, it turned out that this rejection is due to our enormous collateral data set of 200.000 policyholders, almost all parsimoneous models would be rejected faced by this number of observations. We stress that this data set is not the campaign data set for which we test the cross-sale selections, but a larger data set needed to estimate the credibility parameters, see Table 3. When we tested the negative binomial distribution on a wide variety of submodels, high risk, middle risk and low risk submodels, we saw that the negative binomial assumption was accepted for most data sets below 500 in number indicating that the negative binomial provides a good distributional assumption of our data. To get a notion of the accuracy of the fit of the negative binomial distribution to our data, see Figure 1 which shows an almost perfect fit. We did the same figure for our submodels and the negative binomial always provided a satisfactory model fit. Also consider the mean, variances and standard deviations given in Table 1 and Table 2. Notice that variances are close to - but higher - than the means for our selected portfolios. Therefore, our data is really quite close to being Poisson where the mean equals the variance. The reason for this seems to be that the variances for our selected mean frequencies (with and without the latent variable) are indeed very small. Therefore, the our mixed Poisson distribution has a moment structure of its first two moments close to the Poisson distribution.

12

Table 1 here.

Table 2 here.

Figure 1 here.

In the expressions for the expected value of the profit (2) and its variance (3), we allow for customer specific values of all the included parameters, see Section 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 (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 here.

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

Table 4 here.

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 $\mathbf{1}_{\{A_{i2}\}}$ 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

13

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 $\mathbb{E}[N_{i2} \mid \Theta_{i2} = \theta_{i2}] = \lambda_2 \theta_{i2}$. 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 2, 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 2, we also present the observed total profit as a function of the same selection size l.

Figure 2 here.

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 2, 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 3, 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 3 here.

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 here.

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 a 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 (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 5.2, 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 5.2). 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}

To simplify the notation in what follows we will omit the index k. The proof of (2) is straightforward and is omitted. For the variance v_i of H_i, noting that the r.v.s $\mathbf{l}_{\{A_i\}}$ and N_i are assumed independent, we have

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

$$= \mathbb{E}\left[\left(\mathbf{l}_{\{A_i\}}\left(\pi_i - N_i m_i\right)\right)^2 \mid \Theta_i = \theta_i\right] - \left(\mathbb{E}\left[\mathbf{l}_{\{A_i\}}\left(\pi_i - N_i m_i\right) \mid \Theta_i = \theta_i\right]\right)^2 =$$

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

which simplifies to (3), noting that $p_i - p_i^2 = \mathrm{Var}\left(\mathbf{l}_{\{A_i\}}\right)$.

Derivation of the cumulative distribution function of H_i

Formulas (2) and (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

$$F_{H_i}(x) = P\left(H_i \leq x \mid \Theta_i = \theta_i\right) = \begin{cases} 1 & \text{if } x \geq \pi_i - \omega_i \\[2ex] 1 - p_i \sum_{j=0}^{[[\tilde{x}]]} e^{-\theta_i\lambda_i} \frac{(\theta_i\lambda_i)^j}{j!} & \text{if } -\omega_i \leq x < \pi_i - \omega_i \\[2ex] \left(1 - \sum_{j=0}^{[[\tilde{x}]]} e^{-\theta_i\lambda_i} \frac{(\theta_i\lambda_i)^j}{j!}\right) p_i & \text{if } x < -\omega_i \end{cases}$$

(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

$$P\left(H_i \leq x\right) = P\left(\mathbf{l}_{\{A_i\}}\left(\pi_i - N_i m_i\right) - \omega_i \leq x\right) =$$

$$= P\left(\mathbf{l}_{\{A_i\}}\left(\pi_i - N_i m_i\right) \leq x + \omega_i \mid \mathbf{l}_{\{A_i\}} = 1\right) p_i +$$

$$P\left(\mathbf{l}_{\{A_i\}}\left(\pi_i - N_i m_i\right) \leq x + \omega_i \mid \mathbf{l}_{\{A_i\}} = 0\right)\left(1 - p_i\right) =$$

(6)

$$= P\left(N_i \geq \tilde{x}\right) p_i + P(0 \leq x + \omega_i)\left(1 - p_i\right) =$$

$$= \left(1 - P\left(N_i < \tilde{x}\right)\right) p_i + P(0 \leq x + \omega_i)\left(1 - p_i\right)$$

where we have used the independence of the r.v.s $\mathbf{1}_{\{A_i\}}$ and N_i. Representation (5) follows from (6), recalling that, conditional on $\Theta_i = \theta_i$, $N_i \sim \text{Poisson}\,(\theta_i \lambda_i)$.
$\square$

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:

$$
\begin{aligned}
\text{Im}H_i = \{ & x_j = \pi_i - \omega_i - j m_i, j = 0, 1, \ldots, j^*, \\
& x_{j^*+1} = -\omega_i, x_j = \pi_i - \omega_i - (j-1) m_i, j = j^* + 2, j^* + 3, \ldots \}
\end{aligned}
\tag{7}
$$

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

$$
\text{Im}H_i = \{ x_j = \pi_i - \omega_i - j m_i, j = 0, 1, \ldots, j^* - 1, j^* + 1, j^* + 2, \ldots, x_{j^*} = -\omega_i \} .
\tag{8}
$$

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 $\text{Im}H_i$ is as in (7), the probability mass function of H_i is

$$
P\,(H_i = x_j \mid \Theta_i = \theta_i) = \begin{cases} p_i e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^j}{j!} & \text{for } j = 0, 1, \ldots, j^* \\[2mm] 1 - p_i & \text{for } j = j^* + 1 \\[2mm] p_i e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^{j-1}}{(j-1)!} & \text{for } j = j^* + 2, j^* + 3, \ldots \end{cases}
\tag{9}
$$

2. Assuming that $\text{Im}H_i$ is as in (8), the probability mass function of H_i is

$$
P\,(H_i = x_j \mid \Theta_i = \theta_i) = \begin{cases} p_i e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^j}{j!} & \text{for } j = 0, 1, \ldots, j^* - 1 \\[2mm] 1 - p_i + p_i e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^j}{j!} & \text{for } j = j^* \\[2mm] p_i e^{-\theta_i \lambda_i} \frac{(\theta_i \lambda_i)^j}{j!} & \text{for } j = j^* + 1, j^* + 2, \ldots \end{cases}
\tag{10}
$$

Proof Formulas (9) and (10) follow directly from (5) noting that, for assumption 1. (formula (9)), by the definition of j^* in (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 (10)) by the definition of j^* in (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., $P\left(H_i = x_j \mid \Theta_i = \theta_i\right)$, 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), (9) and (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

$$\mathcal{H}_s(l) = \sum_{i=1}^{l} H_i = \sum_{i=1}^{l} \left(\mathbf{1}_{\{A_i\}} \left(\pi_i - N_i m_i\right) - \omega_i\right). \tag{11}$$

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

$$\mathbb{E}\left[\mathcal{H}_s(l) \mid \Theta = \theta\right] = \sum_{i=1}^{l} \mathbb{E}\left[H_i \mid \Theta_i = \theta_i\right] = \sum_{i=1}^{l} \left(p_i \left(\pi_i - \theta_i \lambda_i m_i\right) - \omega_i\right), \tag{12}$$

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

$$\mathrm{Var}_s(l) = \sum_{i=1}^{l} \mathrm{Var}\left[H_i \mid \Theta_i = \theta_i\right] = \sum_{i=1}^{l} \left(\mathrm{Var}\left[\mathbf{1}_{\{A_i\}}\right] \left(\pi_i - \theta_i \lambda_i m_i\right)^2 + p_i m_i^2 \theta_i \lambda_i\right). \tag{13}$$

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

$$P\left(\mathcal{H}_s(l) = h \mid \Theta = \theta\right) = \sum_{x_1 \in \mathrm{Im}H_1} \cdots \sum_{x_{l-1} \in \mathrm{Im}H_{l-1}} P\left(H_1 = x_1 \mid \Theta_1 = \theta_1\right) \times \ldots$$

$$\ldots \times P\left(H_{l-1} = x_{l-1} \mid \Theta_{l-1} = \theta_{l-1}\right) P\left(H_l = h - x_1 - \ldots - x_{l-1} \mid \Theta_l = \theta_l\right), \tag{14}$$

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 (14), for the cdf $F_{\mathcal{H}_s(l)}(x) = P\left(\mathcal{H}_s(l) \leq x \mid \Theta = \theta\right)$ we have

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

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

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

$$\ldots \times P\left(H_{l-1} = x_{l-1} \mid \Theta_{l-1} = \theta_{l-1}\right) P\left(H_l = h - x_1 - \ldots - x_{l-1} \mid \Theta_l = \theta_l\right),$$
$$(15)$$

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

$$P\left(Q_{\frac{\alpha}{2}} \leq \mathcal{H}_s(l) \leq Q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha, \tag{16}$$

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, $Q_{\frac{\alpha}{2}} = F_{\mathcal{H}_s(l)}^{-1}\left(\frac{\alpha}{2}\right)$ and $Q_{\frac{\alpha}{2}} = F_{\mathcal{H}_s(l)}^{-1}\left(1 - \frac{\alpha}{2}\right)$, where $F_{\mathcal{H}_s(l)}^{-1}(\cdot)$ 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 $F_{\mathcal{H}_s(l)}^{-1}(\cdot)$ using (14) and (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 (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 (14) and (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 $\mathbb{E}\left[\mathcal{H}_s(l) \mid \Theta = \theta\right]$ 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

$$\frac{1}{B_l^2} \sum_{k=1}^{l} \sum_{\{j : |x_j - \mathbb{E}(H_k)| > \varepsilon B_l\}} P\left(H_j = x_j\right)\left(x_j - \mathbb{E}\left(H_k\right)\right)^2 \underset{l \to \infty}{\longrightarrow} 0 \tag{17}$$

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 (17) holds, i.e., the set of values for which condition (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 (17) holds, the distribution functions of the normalized total profit,$(\mathcal{H}_s(l) - C_l)/B_l$ tend to a standard normal cdf, as l tends to infinity.

Proposition 6 allows for the construction of approximate confidence regions, of the form (16), for the total profit random variable, $\mathcal{H}_s(l)$, when l is sufficiently large, given that (17) holds. For a given confidence level, α, we have that

$$P\left(q_{\frac{\alpha}{2}} \leq (\mathcal{H}_s(l) - C_l)/B_l \leq q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha, \tag{18}$$

where $q_{\frac{\alpha}{2}}$, and $q_{1-\frac{\alpha}{2}}$ are the corresponding quantiles of the standard normal distribution. From (18), for $\alpha = 0.05$ we have that, $P\left(C_l - 1.96B_l \leq \mathcal{H}_s(l) \leq C_l + 1.96B_l\right) = 0.95$.

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 $E[\Theta_{il}] = \theta_{0l}$, $i = 1, \ldots, I$ and $Cov[\Theta_{il}, \Theta_{ir}] = \tau_{lr}^2$, $l, r \in \{k', k\}$. 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

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

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 $\lambda_{i \cdot l} = \sum_{j=1}^{J_i} \lambda_{ijl}$, $n_{i \cdot l} = \sum_{j=1}^{J_i} n_{ijl}$ 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

$$\hat{\theta}_i = \theta_0 + \alpha_i(F_{i \cdot} - \theta_0) \tag{19}$$

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 (19) and considering element 2 of $\hat{\theta}_i$ we get

$$\hat{\theta}_{i2} = \theta_{02} + \alpha_{i22} \left(F_{i\cdot2} - \theta_{02} \right) + \alpha_{i21} \left(F_{i\cdot1} - \theta_{01} \right). \qquad (20)$$

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 $\hat{\lambda}_{ij2} = \hat{\lambda}_{i\cdot2} = 0$. It is possible to show that $\hat{\lambda}_{i\cdot2} = 0$ implies that $\hat{\alpha}_{i22} = 0$ and (20) becomes

$$\hat{\theta}_{i2} = \hat{\theta}_{02} + \hat{\alpha}_{i21} \left(\hat{F}_{i\cdot1} - \hat{\theta}_{01} \right),$$

where $\hat{\alpha}_{i21} = \frac{\hat{\lambda}_{i\cdot1}\hat{\tau}_{21}^2}{\hat{\lambda}_{i\cdot1}\hat{\tau}_{11}^2 + \hat{\sigma}_1^2}$. 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 $\hat{F}_{i\cdot1} = \frac{n_{i\cdot1}}{\hat{\lambda}_{i\cdot1}}$ with respect to the other (owned) product 1. For a thorough description of multivariate credibility in cross-selling, see Thuring (2012) and Thuring et al. (2012).

References

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

Ahn, H., Ahn, J.J., Oh, K.J. & 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.

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

Bult, J.R. & 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. & Straub, E., 1970, Glaubwürdigkeit für Schadensätze, *Bulletin of Swiss Association of Actuaries*, 70(1), 111-133.

Bühlmann, H. & 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. & Nielsen, J.P., 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(2), 443-453.

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

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

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

Hwang, H., Jung, T. & Suh, E., 2004, An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry, *Expert Systems with Applications*, 26(2), 181-188.

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

Jonker, J.J., Piersma, N. & Van den Poel, D., 2004, Joint optimization of customer segmentation and marketing policy to maximize long-term profitability, *Expert Systems with Applications*, 27(2), 159-168.

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. & Wedel, M., 2004, "Identifying innovators for the cross-selling of new products," *Management Science*, 50(8), 1120-1133

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(1), 45-65.

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.

Kim, S.Y., Jung, T.S., Suh, E.H. & Hwang, H.S., 2006, Customer segmentation and strategy development based on customer lifetime value: A case study, *Expert Systems with Applications*, 31(1), 101-107.

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

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

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

Lee, J.H. & Park, S.C., 2005, Intelligent profitable customers segmentation system based on business intelligence tools, *Expert Systems with Applications*, 29(1), 145-152.

Li, S., Sun, B. & 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. & 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.

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

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

Rothschild, M. & 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.

Thuring, F., Nielsen, J.P., Guilln, M. & Bolanc, C., 2012, Selecting prospects for cross-selling financial products using multivariate credibility, *Expert Systems with Applications*, 39(10), 8809-8816.

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

Table 1

Mean, variance and standard deviation for household insurance data.

	Mean	Variance	Standard deviation
n_{i1}	0.138	0.166	0.407
$\hat{\lambda}_{i1}$	0.130	0.00553	0.0743
$\hat{\theta}_{i1}$	1.12	0.00169	0.0411
$\hat{\lambda}_{i1}\hat{\theta}_{i1}$	0.155	0.00692	0.0832

Table 2

Mean, variance and standard deviation for car insurance data.

	Mean	Variance	Standard deviation
n_{i2}	0.219	0.273	0.523
$\hat{\lambda}_{i2}$	0.220	0.0187	0.137
$\hat{\theta}_{i2}$	0.922	0.0220	0.148
$\hat{\lambda}_{i2}\hat{\theta}_{i2}$	0.202	0.0176	0.133

Table 3

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

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

1

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

Table 5

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

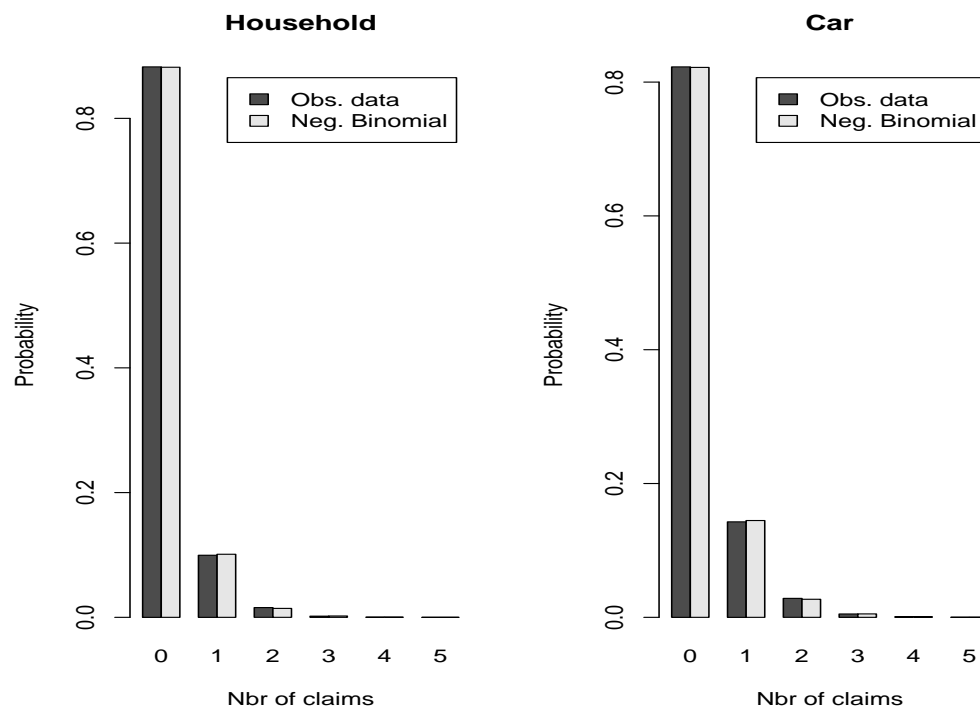| Criteria	Number of customers	Expected total profit	Variance of total profit	Observed total profit
EP	2647	$16,424	$3.0 \cdot 10^{8}$	$16,362
MV	1319	$12,787	$1.6 \cdot 10^{8}$	$3,882

Figures

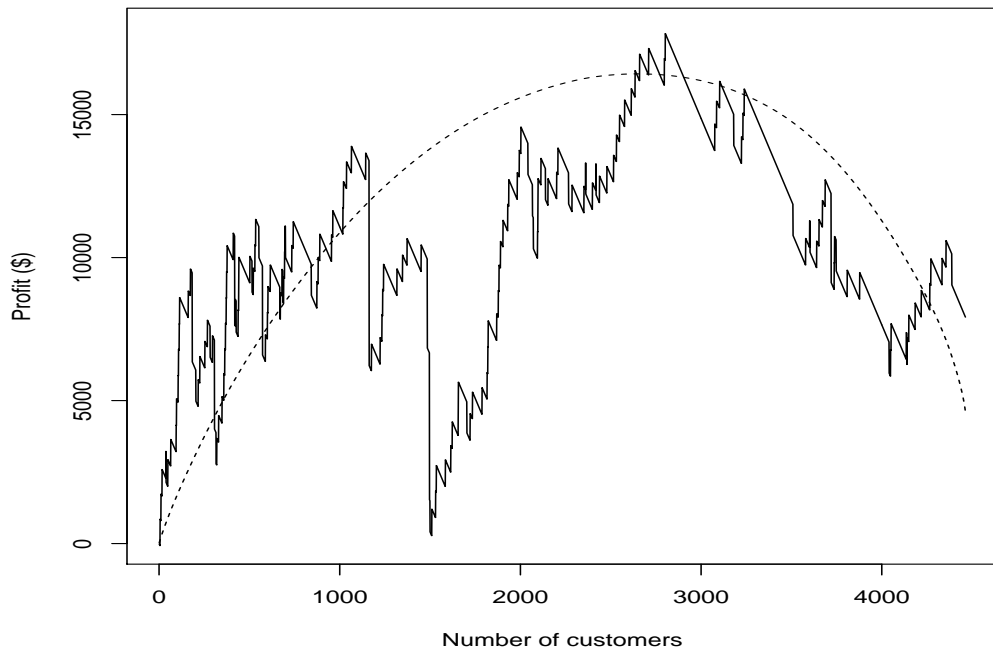Figure 1: Negative binomial fit to motor and household insurance data.

Figure 2: 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.
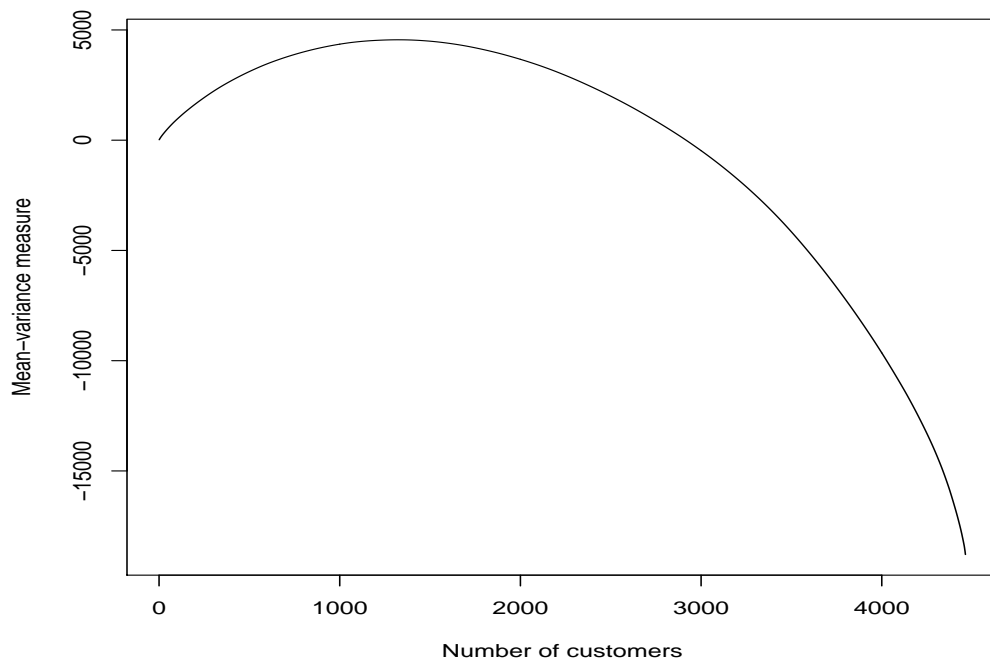
Figure 3: Mean-variance, as cumulative sums, emerging from approaching an increasing number of customers l, with $l = 1, \ldots, 4463$. The customers are ordered by non-increasing mean variance values MV_{i2} prior to cumulative summation and plotting.