

PREDICTING CUSTOMER POTENTIAL VALUE

AN APPLICATION IN THE INSURANCE INDUSTRY

PETER C. VERHOEF AND BAS DONKERS

ERIM REPORT SERIES <i>RESEARCH IN MANAGEMENT</i>	
ERIM Report Series reference number	ERS-2001-01-MKT
Publication	January 2001
Number of pages	35
Email address first author	verhoef@few.eur.nl
URL (electronic version)	http://www.eur.nl/WebDOC/doc/erim/erimrs20010110120024.pdf
Address	Erasmus Research Institute of Management (ERIM) Rotterdam School of Management / Faculteit Bedrijfskunde Erasmus Universiteit Rotterdam PoBox 1738 3000 DR Rotterdam, The Netherlands Phone: # 31-(0) 10-408 1182 Fax: # 31-(0) 10-408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES *RESEARCH IN MANAGEMENT*

BIBLIOGRAPHIC DATA AND CLASSIFICATIONS		
Abstract	For effective Customer Relationship Management (CRM), it is essential to have information on the potential value of customers. Based on the interplay between potential value and realized value, managers can devise customer specific strategies. In this article we introduce a model for predicting the potential value of a current customer. Furthermore, we discuss and apply different modeling strategies for predicting this potential value.	
Library of Congress Classification (LCC)	5001-6182	Business
	5410-5417.5	Marketing
	HF 5415.55	Relationship Marketing
Journal of Economic Literature (JEL)	M	Business Administration and Business Economics
	M 31	Marketing
	C 44	Statistical Decision Theory
	M 31	Marketing
European Business Schools Library Group (EBSLG)	85 A	Business General
	280 G	Managing the marketing function
	255 A	Decision theory (general)
	280 C	Marketing Science
Gemeenschappelijke Onderwerpsontsluiting (GOO)		
Classification GOO	85.00	Bedrijfskunde, Organiseatiekunde: algemeen
	85.40	Marketing
	85.03	Methoden en technieken, operations research
	85.40	Marketing
Keywords GOO	Bedrijfskunde / Bedrijfseconomie	
	Marketing / Besliskunde	
	Relatiemarketing, Klantgerichtheid, Verzekeringswezen	
Free keywords	Customer Relationship Management, Customer Potential, Marketing Models, Insurance Industry	
Other information		

PREDICTING CUSTOMER POTENTIAL VALUE

AN APPLICATION IN THE INSURANCE INDUSTRY

Peter C. Verhoef*

Bas Donkers†

Abstract

For effective Customer Relationship Management (CRM), it is essential to have information on the potential value of customers. Based on the interplay between potential value and realized value, managers can devise customer specific strategies. In this article we introduce a model for predicting the potential value of a current customer. Furthermore, we discuss and apply different modeling strategies for predicting this potential value.

Key-words: Customer Relationship Management, Customer Potential, Marketing Models, Insurance Industry

* Corresponding Author: Peter C. Verhoef, Erasmus University Rotterdam, Department of Marketing and Organization, Office H15-12, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, Phone +31 10 408 2809, Fax +31 10 408 9169, E-mail: Verhoef@few.eur.nl, <http://www.few.eur.nl/few/people/verhoef>

† Bas Donkers, Erasmus University Rotterdam, Department of Marketing and Organization, Office H15-12, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, Phone +31 10 408 2411, Fax +31 10 408 9169, E-mail: Donkers@few.eur.nl

PREDICTING CUSTOMER POTENTIAL VALUE

AN APPLICATION IN THE INSURANCE INDUSTRY

1. INTRODUCTION

Since general recognition of the marketing principle that keeping customers is more profitable than attracting new customers [2], many companies have adopted relationship marketing [7]. In relationship marketing, managers strive to develop and maintain successful customer relationships [16]. Only recently, companies realized that in order to develop such relationships a differentiated approach is needed [3, 22]. Instead of treating all customers equally, managers have come to understand that it is more effective to develop customer-specific strategies. As a result, companies are now adopting customer relationship management (CRM). CRM means that companies manage relationships with individual customers with the aid of (customer) databases and interactive and mass customization technologies [17]. The adoption of CRM has been enhanced by recent developments in Information and Communication Technology (e.g. Database Technology, E-commerce, and the Internet).

By using customer information contained in databases, companies can invest in the customers that are (potentially) valuable for the company, but also minimize their investments in non-valuable customers. Figures on the turnover of each customer or customer profitability are often used as segmentation variables to distinguish between valuable and non-valuable customers. In this way database analysts construct customer pyramids, as shown in [19, p. 187]. This type of segmentation can be valuable in a single service setting, but it can also be misleading for multi-service or multi-product providers. These providers are not only interested in the current value

of customers, but also attach importance to information on cross-selling opportunities. For example, although a customer may currently purchase only a small number of the services offered by the focal company, he might potentially be very valuable, as he may also purchase many other services. Therefore, we propose to use not only information on the current value of a customer, but also the potential value of a customer [4, 12, 13]. Potential value is defined as the profit or value delivered by a customer if this customer behaves ideally, i.e., the customer purchases all products or services he currently buys in the market at full prices at the focal company [12]. Combining information on a customer's potential value and a customer's current value provides the CRM-manager with an opportunity to extend the "customer pyramid" segmentation. A two-by-two segmentation, as displayed in figure 1, is proposed, which creates a better basis for customer specific strategies. For example, companies can decide to target investments on the customers with a low current value, but high potential value. We will discuss such a segmentation of the customer base in section 2.

To obtain information on the potential value of a customer, analysts need data on the customer's purchasing behavior at their own company, as well as at other companies in the market. Usually companies only have data on customers' purchasing behavior at their own company in their customer information file (CIF) [21]. Hence, models are needed to predict the potential value of a customer, based on the purchasing behavior in the CIF, and on any available socio-demographic data.

Zeithaml [22] states that a lot of work needs to be done on identifying the potential value of current customers. Numerous models have been developed to predict single transactions (e.g. [4]) and some work has been done to predict purchase patterns at the

focal supplier [20]. Kim and Kim [15] describe a model that estimates the upselling potential for a one-product or service provider, but apparently no models are available that predict the potential value of a customer in a multi-service context. An exception is the work of Kamakura, Ramaswamy and Srivastava [14], who describe a model that explains the financial maturity of customers. However, their approach depends critically on the hierarchy of investment objectives, which is not a general feature of multiple product or service industries.

Given the above literature overview on customer potential value models, the objective of our paper is to develop a framework that provides insight into the potential value of customers to CRM-managers in a multi-service industry. We will compare different modeling approaches to find the most informative ones. Specifically, we will compare a choice-based model using Univariate and Multivariate Probit, with a potential value model, based on a linear regression model.

By this paper we extend the CRM-literature in the following respects. First, in the scientific context our study is the first to focus on the modeling and prediction of the potential value of customers of a multi-service provider. Thus, we compare the performance of competing models that predict customer potential value. Second, in a managerial context we provide CRM-managers in multi-service industries with a framework which can be used to predict customer potential. This framework takes account of the data limitations a company usually has, by using socio-demographic information and transaction information from the customer database solely. The results can then be used as input for customer segmentation, which we will approach more conceptually in section 2.

The structure of this paper is as follows. In section 2, we start with a discussion on the potential value of customers and a segmentation based on it. Then we will provide our conceptual framework for customer potential. In section 3, we describe the methodology and the data requirements for the prediction of customer potential value. In section 4, we present an application of this methodology in the insurance industry. We also discuss the market segmentation and management implications for this application. Finally, we end with a conclusion, model limitations and directions for future model developments in section 5.

2. BACKGROUND AND MODEL

The first part of this section will be devoted to a discussion on customer potential value and a segmentation method for CRM that uses customer potential value. Next, we will describe the possible antecedents of a customer's potential value, and we will present our conceptual model.

Potential value

As already mentioned in the introduction, the potential value of a customer refers to the profitability of a customer if that customer buys all purchased products or services from the supplier [12]. Hence, customer value depends heavily on the number of purchases in the product or service category made by an individual customer [13]. The potential value is computed as the total profit margin on all purchases. From a managerial perspective a customer's potential profitability is very interesting, since customer specific optimal budgets for relational marketing efforts can be derived from it [3].

We note that from a CRM-perspective the potential value of a customer reflects not only the current potential, but also the future potential [12]. This is especially true for markets with unstable purchase patterns. Since often no information is available on future purchase patterns, the prediction of this ideal measure of customer potential is difficult. Therefore, we focus on the current potential value of a customer. In our empirical application in the insurance market, purchase patterns are rather stable, so current potential and future potential are strongly linked.

Customer Segmentation and Customer Potential

In CRM, managers develop specific strategies for different segments of their customer base. The customer pyramid is often used as a segmentation method. Using this pyramid, strategies mainly focus on moving promising customers to the top of the pyramid and optimizing revenues from less promising customers by, for example, increasing prices or reducing costs [19]. However, although these strategies are useful, using a customer's current value as segmentation variable solely might lead to sub-optimal strategies. We will illustrate this statement with two examples. First, a customer might belong in the low value segment of the customer pyramid. Hence, companies would strive to optimize revenues by reducing costs (that is: lower service levels and marketing expenditures) and increasing prices. However, when considering the potential value of the customer, this might indicate huge cross-selling opportunities, and so a manager should invest in this customer in order to take a larger share of this potential value. Second, again using a customer pyramid, CRM-managers might strive to move customers with a reasonable value into higher tiers of the pyramid. However, these customers might have reached their full potential and no cross-selling opportunities exist. Hence, investments in moving these customers into higher tiers would be wasted. Clearly, a more differentiated approach is needed,

which explicitly takes the potential value of a customer into account. Such a differentiation can be derived from a two-by-two segmentation matrix as displayed in figure 1. Using this matrix CRM-managers can formulate better segment specific strategies. Note that this segmentation method can be fine-tuned by distinguishing more groups on each axis. We will briefly discuss the strategies for each segment:

Segment I: Segment I can be regarded as unattractive. It has low potential value and low current value. Therefore, it is expected that future profitability is low. In order to maximize the profitability of this segment, strategies should focus on cost reductions and possibly on price increases (i.e. less promotions) instead of trying to increase the purchase level.

Segment II: Segment II has high potential value, but the company has not succeeded in taking a large share of this value. Therefore, companies should aim to get a larger part of the customer potential in this segment. Customers in this segment have many opportunities for up-selling activities. Of course, some customers might be more sensitive to such activities than others.

Segment III: Segment III has low potential value and high current value. We are concerned here with relatively loyal customers with low up-selling possibilities. As loyal customers are important for companies [18], companies should strive to keep these customers. However, up-selling efforts are not likely to be successful.

Segment IV: This segment is the most valuable segment. These customers are loyal and have a large potential value. Losing this group of customers would really harm the company. Management should strive to keep this group

of customers using all kinds of relational efforts. This group might, for example, get priority in the service delivery process.

<<Insert Figure 1 about here >>

Given the relevance of potential value in CRM, we will continue with a discussion on the antecedents of potential value and a detailed description of our conceptual model.

Antecedents of Potential Value

In consumer research, consumer needs and the available resources are important drivers of acquisition decisions for products and services [11]. An individual's needs are affected by factors such as household composition, gender, attitudes (e.g. risk attitude) and social class [6]. The extent to which these needs can be satisfied depends on the consumer's resources. Complete information on needs and resources is hardly ever available, but you could use socio-demographic information relating to tastes, needs, and resources. For example, from research in the financial services industry, it is well known that the family lifecycle is a determinant of the type of services acquired [1]. In addition, Kamakura, Ramaswami and Srivastava [14] report that demographic factors, such as income, age, and education, are important determinants in the acquisition of financial services.

To predict the purchasing of different products or services, data on the purchasing of other products or services can also serve as important predictors. For example, Kamakura, Ramaswami and Srivastava [14] report strong interdependencies between the types of financial services purchased. Although we are not interested in the amount of interdependency, it might be very helpful to use purchase information of

other products when predicting purchase decisions. This takes into account the possible information on the interrelationships.

Conceptual Model

The variables that can be used to predict the potential value of a customer in a marketing decision support system depend to a great extent on the availability of data. Spring et al. [21] report that most companies that use a customer database have information on the purchasing behavior of customers at their own company. Often they also have information on some socio-demographic characteristics. Subjective information on attitudes and lifestyle is typically not available. Therefore, despite the possible effect of this type of variables on the potential value of a customer, these variables, in general, cannot be included in a model for a marketing decision support system. Hence, in our conceptual model we will consider socio-demographic characteristics and the purchasing behavior at the own company as the determinants of potential value. The conceptual model is displayed in figure 2. Note that the information on purchases at the company is also part of the customers potential. We account for this in the estimation strategy.

<<Insert Figure 2 about here>>

3. EMPIRICAL MODELING

In this section, we will present the empirical implementation of our conceptual model. We start with a discussion of the data requirements. Next, we will discuss the empirical specification of the models for purchase behavior, for potential value, and a customer base segmentation based on these models.

Data requirements

Information about all of a customer's product purchases in the company's markets is needed to derive a customer's potential value. This information is usually not available, but a survey among customers is an easy way of obtaining this information. Besides complete information on purchase behavior, predictors for these purchase decisions are also needed. From the conceptual model we concluded that both socio-demographic and actual purchase information at the company can be useful predictors of purchase decisions. Actual purchase information is usually stored in the customer information file (CIF). Some companies also have socio-demographic information in their CIF, but otherwise such information can be obtained from external suppliers, such as CCI.

Estimation Procedure

Estimation of potential value can be carried out with models at different levels of aggregation of behavior. A model for purchasing behavior for each product or service uses the data at the lowest level of aggregation. The individual purchases can also be aggregated into an individual specific measure of potential value. This measure of potential value can be modeled with a linear regression model. When interest is restricted to a segmentation of the customer base into a high potential and a low potential segment, the data on potential value can be summarized with the segment memberships of each customer. This can be modeled with a probit model.

The models that use less aggregated information, in general, provide more information about the driving forces of potential value. However, such models do not necessarily result in a better performance in predicting the aggregated variables. A

model that is aimed solely at modeling the aggregate variable and not the underlying behavior, such as the probit model for segment membership, might be better.

At each level of aggregation of the data, it would be desirable to use different types of econometric models. At the lowest level of aggregation the dependent variable is the decision to purchase a certain product or service, which is a binary choice. Usually, a probit model is used to predict the purchases of the various services [10]. However, in many cases these purchase decisions are made simultaneously, or, at least, they are related. In our empirical application, which deals with the purchases of insurance policies, for example, the unobserved risk attitudes of the customers are likely to result in interdependencies across the decisions to purchase the different insurance policies. For this reason a multivariate probit model is also estimated. This model allows for correlations between the error terms in the probit equations for each service [5].

The (univariate) probit model for purchases of product j , $j=1, \dots, J$, by customer i is specified as follows:

$$y_{ij}^* = \mathbf{b}_j X_i + \sum_{k=1}^J \mathbf{g}_{jk} Z_{ik} + \mathbf{e}_{ij} \quad (1)$$

$$y_{ij} = 1 \text{ if } y_{ij}^* > 0 \quad (2)$$

$$y_{ij} = 0 \text{ if } y_{ij}^* \leq 0 \quad (3)$$

where for $i=1, \dots, N$, $j=1, \dots, J$:

$$y_{ij}^* = \text{unobserved variable}$$

y_{ij} = ownership of product or service j for customer i (1 = ownership, 0 = no ownership) (survey)

X_i = socio-demographic indicators (e.g. age, income) of customer i (CIF or external)

Z_{ik} = observed ownership of product or service k at company for customer i (CIF)

ϵ_{ij} = error

The main assumption underlying the regular probit model is that the errors are independent across individuals, but also across insurance types [10]. The multivariate probit model allows for correlations relating to the purchase decisions for the insurance types. Here the assumption is that the vector of errors, $\epsilon_{i1}, \dots, \epsilon_{ij}$, follows a multivariate normal distribution with an unrestricted covariance matrix [5]. As these correlations result in dependencies relating to the purchase decision for the various services, the multivariate probit model results in probabilities with which a customer purchases a certain portfolio of services.

In our empirical application both the multivariate probit model with an unrestricted covariance matrix and univariate probits for each type of insurance are used. The models are validated by comparing the hit rate of the models, i.e., the percentage of observations correctly predicted, with the hit rate of a naïve model. The models are tested for predictive accuracy with the test of Franses [8]. The estimation results for the purchase decisions can be used to predict potential value. However, the results can also serve a different purpose. Knowing which customers are more likely to purchase a particular service is also helpful in developing a target selection model for marketing activities for the service concerned.

Using information on the profitability of each product, a customer's potential value can be predicted with the estimation results of the multivariate probit model. A prediction for the potential value is obtained by multiplying the predicted probability of ownership of each possible service portfolio, by the expected profitability of such a portfolio. Thus we obtain the following equation to compute the potential value of customer i .

$$\text{Potential Value}_i = \sum_{k=1}^K \text{Prob}(\text{customer } i \text{ owns portfolio } k) * \text{Profit}_k \quad (4)$$

where:

$\text{Prob}(\text{customer } i \text{ owns portfolio } k)$ = Probability of customer i purchasing portfolio k

Profit_k = Profit margin of all services in portfolio k .

In the situation without dependence across the different services, this reduces to the more familiar probabilities that result from the traditional probit model:

$$\text{Prob}(y_{ij} = 1) = \text{Prob}(\mathbf{e}_{ji} > -\mathbf{b}_j X_i - \sum_{k=1}^J \mathbf{g}_{jk} Z_{ik}) \quad (5)$$

and the following formula for the potential value:

$$\text{Potential Value}_i = \sum_{j=1}^J \text{Prob}(y_{ij} = 1) * \text{Profit}_j \quad (6)$$

The above formulae for predicting the potential value of a customer use detailed information about purchase behavior of the different products. When you are solely interested in a customer's potential value itself, and not in the services that determine this potential value, a simple regression model can be used to predict the potential value of a customer. Predictions of potential value can then be based on an Ordinary Least Squares estimate of the following regression model:

$$\text{Potential value}_i = \mathbf{b}X_i + \sum_{k=1}^J \mathbf{g}_k Z_k + \mathbf{e}_i \quad (7)$$

From the resulting estimation results you can derive which customer characteristics determine potential value, but not how these characteristics influence the purchases of each type of service. Although this insight is lost, the regression model might still be the more appropriate model for predicting potential value as it is designed to model continuous variables.

The models for predicting potential value can be evaluated using well-known criteria like the Mean Absolute Prediction Error (MAPE). For comparison we also report these measures for the simplest possible prediction of a customer's potential value, which is the mean potential value in the estimation sample.

When interest is limited to a segmentation of the customer base into a high potential and a low potential segment, a suitable model that can be used is the probit model for segment membership. This method can also be easily generalized for the case with multiple segments with the ordered probit model [10]. The probit model for membership of the high potential value segment is defined as follows, [see also equations (1) to (3)]:

$$y_i^* = \mathbf{b}X_i + \sum_{k=1}^J \mathbf{g}_k Z_{ik} + \mathbf{e}_i \quad (8)$$

$$y_i = 1 \text{ if } y_i^* > 0 \quad (9)$$

$$y_i = 0 \text{ if } y_i^* \leq 0 \quad (10)$$

Here y_j^* is an unobserved variable, $y_i = 1$ indicates that individual i is in the high potential value segment, while $y_i = 0$ indicates otherwise.

In the empirical application we use a median split to segment the customer base into two equally sized parts. The estimation results for the probit model for service purchases and the regression model for potential value are also used to segment the customer database into two segments of equal size, at least in the estimation sample.

4. APPLICATION TO THE INSURANCE INDUSTRY

In this section, we present the application of our methodology to an insurance company in the Netherlands. We start with a short description of the data. Then we estimate and evaluate the models for each aspect of behavior we are interested in.

Data

We use data from an insurance company in the Netherlands. This company is a large direct writer and does not use agents. They sell all types of insurance policies, ranging from fire and theft insurances to life insurance. The company aims to have close relationships with their customers and hence possesses a customer database in which information on the purchasing behavior of customers at the company, and some other characteristics, such as age and relationship duration, are stored.

Data on the ownership of different insurance policies were collected by means of a telephone survey among a proportionally stratified sample of about 2300 customers of the insurance company. The bases for stratification are relationship duration, purchase

level of insurances and claiming behavior. Using this sampling methodology, we obtain a representative sample on these important characteristics. The survey also includes questions on age, education, household size, income, and home ownership. After deleting cases with missing values we obtained a final sample of 1612 customers. In line with the profile of customers of this company, our sample can be described as representing rather prosperous and well-educated people. A more detailed description of the sample characteristics is given in Appendix A.

Respondents were asked to indicate whether they had effected 12 types of insurance. To check the reliability of the answers, we compared the reported ownership with the available information from the customer database. It turned out that there was not a single case where ownership was not reported, meaning there were no discrepancies with the customer information file. This indicated that the answers on the ownership questions were reliable.

Table 1 presents ownership rates for each of these 12 insurance types. Because of the confidential nature of our data, we report the insurance types in alphabetical order. The insurance types are: car, damages, disability, funeral, furniture, health, house, liability, legal aid, life, travel, and continuous travel insurance. The reported ownership rates of these insurance types are sorted by ownership rates, so they cannot be linked to the actual insurance types. The numbering introduced here will be used throughout the paper.

<< Insert Table 1 about here >>

Estimation Results

For four insurance types the ownership rates were very close to 100%. To reduce modeling efforts and to save some space, it was assumed that all customers own these four types of insurance. The variation in potential value we wanted to explain therefore results from the remaining eight types of insurance. In order to capture non-linear effects of the explanatory variables of age, income, and education, we used dummies for the separate classes in our models. The evaluation of the predictions was carried out on a sample that was not used for estimation. We split our sample into an estimation sample with 1000 households. The remaining 612 households were used to validate the models and to evaluate the prediction performance.

Prediction of Purchases

The prediction results for behavior at the lowest level of aggregation, the purchases of each insurance type, are presented in Table 2. All functions are significant ($p < 0.05$), except the one for insurance 11 ($p < 0.10$). We do not report the parameter estimates for the models, but the general conclusion is that socio-demographic variables as well as purchase data from the CIF serve as predictors for ownership. Important socio-demographic predictors are age, income, marital status and the ownership of a house. Besides for the prediction of potential value, the ownership probabilities that result from the probit models can also be used to target direct mail campaigns for an insurance at customers who are more likely to own this insurance.

For each type of insurance, Table 2 presents the fraction of correct predictions in the validation sample for univariate probits, multivariate probit, and for a naïve model that predicts what is most often observed in the estimation sample. The p values in the

table correspond to a test of predictive performance, where significant p values imply dependence between realizations and predictions [8].

<<Insert Table 2 about here>>

From the table it is clear that for each type of insurance the models predict more than 50% correctly and the p values indicate that there are significant relationships between the predictions and the realizations for most insurance types. For some types of insurance the naïve model outperforms both probit models. However, on average, the hit rates for the probit models are substantially higher, with only a small difference between the two probit models. At first sight, it seems remarkable that the more complicated multivariate probit model does not perform better than the univariate probit model. However, the information about the correlations in the multivariate probit model, that is available through the observed insurance portfolio, is also used in the univariate probit models through the dummies of insurance ownership at the company. This already includes all the information in the data about the possible correlations that is available for prediction¹.

Prediction of Potential Value

The aim of our paper is not to predict ownership rates, but to estimate potential profitability of the customers and to develop CRM strategies, based on these estimates. From the insurance company we have information on the average contribution margins of each insurance type. Combining this information with the predicted ownership probabilities of the probit models, each customer's potential value can be predicted.

Table 3 reports the Mean Absolute Prediction Errors (MAPE) of the predicted potential values from the multivariate probit model and the regression model.² The MAPE of a naïve model that always predicts the mean is also reported for comparison. The MAPE for the three models are all very similar (within 0.15%) and better than a model without explanatory variables, which is the naïve model in the table.

<<Insert Table 3 about here>>

The small improvements of our model compared to a naïve prediction model for insurance ownership and potential value are to some extent disappointing. From a management perspective, however, the advantage of linking observed characteristics to the observed behavior is that a segmentation of the customer base can be based on the observed characteristics. Such a segmentation can then be used in a decision support system. A segmentation cannot be created with the naïve model, as it predicts the same potential value for each customer.

Market Segmentation and Implications

So far, we have discussed the estimation and prediction results for insurance ownership and customer profitability. The remaining question is whether these results can be used to construct a useful segmentation of the customer base.

<<Insert Table 4 about here>>

Our first segmentation is based on potential value only. We distinguish customers with a high and a low potential value using a median split in the estimation sample.

This segmentation is often used in marketing practice (e.g. heavy users vs. low users) [6]. Table 4 presents the average actual potential value for the high and low potential value segment for each model. Also reported in each cell are the number of customers and the standard deviation of potential profit. For reasons of confidentiality we have indexed profits, so average profits are 100. The low value segment has, on average, 4% to 5% lower profit levels, while the high value segment, on average, yields 4% higher profits for the segmentation of the multivariate probit model and the regression model. Surprisingly, the probit model for segment membership does worse in predicting segment membership. This was not expected a priori, as the probit model is specially designed to model binary outcomes. Here the loss of information due to aggregation becomes visible.

<<Insert Table 5 about here>>

In section 2, we discussed a segmentation based on customers' potential value and customer profitability. The results of this segmentation are shown in Table 5 for the customer potential segmentation based on the regression model. The most prominent aspect of the market segmentation for the insurance company under consideration is that it has a large segment of customers with a high potential value, but only a low current value (Segment II, top-left in the matrix). Our analysis identifies this segment as a segment at which one should target up-selling activities, since there are large potential gains in this segment that are not captured by the company. The fact that usually simple and less profitable insurance types are sold by direct writers explains the existence of this large segment.

The information of the customer base segmentation presented in table 5 can be stored in the CIF. This information can be used to direct customer contacts. For example, in call centers management might give customers in attractive segments priority, e.g. shorter waiting times, in the service delivery process compared to the customers in the less attractive segments.

5. DISCUSSION, RESEARCH LIMITATIONS AND FUTURE RESEARCH

Discussion

Our research mainly focused on the modeling of customer potential value. We discussed and compared different statistical methods to model this value: univariate probit, multivariate probit and regression analysis. With respect to the modeling of ownership our models perform somewhat better than the naïve model. However, multivariate probit and univariate probit have similar results. Given these results, it appears more appropriate to use univariate probit, as this technique is easily performed in most statistical packages. This technique does not predict very well, though, as only some specific insurance types could be predicted well with our data. These insurance types, such as legal aid and continuous travel insurances, are typically related to a customer's socio-demographic characteristics. Ownership of other insurance types with less specific characteristics is more difficult to predict.

With respect to the prediction of potential value, regression analysis appears to have the best predictive power. This is also reflected by the fact that when we predict segment membership (that is: low potential value vs. high potential value), regression analysis also appears to predict better than the other methods.

In general, there is no theoretical reason why one of the models should perform better than the other models. Modeling purchase incidence has the advantage that it provides more insight into the services that drive customer potential value, but it also requires the largest amount of modeling. Models for behavior at higher levels of aggregation require less modeling efforts, but they might suffer from the loss of information due to aggregation. This is the case with the probit model for segment membership in our empirical application. Model validation and comparison of predictive performance is therefore of major importance when deciding on which model to use as input in a decision support system.

Research Limitations and Future Research

Our methodology only considers current potential value, whereas ideally a manager prefers information on current and future potential value of customers. To incorporate future potential value, panel information is needed which was not available. In future research, a longitudinal estimation strategy can be developed. Moreover, as with any segmentation, you can think of finer market segmentations. In addition to the proposed segmentation, you might consider responsiveness to up-selling activities as a third characteristic to include in the segmentation. Finally, our model was developed to predict the value of current customers. Future research can develop models that predict the potential value of new customers.

ACKNOWLEDGEMENTS

The authors acknowledge the data support of a Dutch insurance company and helpful comments by Philip Hans Franses and Ernst Verwaal and the two editors of the special issue on CRM.

REFERENCES

- [1] G.A. Antonides and W.F. van Raaij, *Consumer Behavior A European Perspective* (John Wiley & Sons, Chichester, 1998)
- [2] G.R. Bitran and S.V. Mondschein, *A Comparative Analysis of Decision Making Procedures in the Catalog Sales Industry*, *European Management Journal* 15, No.2 (1997)
- [3] R.C. Blattberg and J. Deighton, *Managing Marketing by the Customer Equity Test*, *Harvard Business Review* 75, No. 4 (1996)
- [4] J.R. Bult and T. Wansbeek, *Optimal Selection for Direct Mail*, *Marketing Science* 14, No. 4 (1995)
- [5] S. Chib and Greenberg, E. (1998) *Analysis of multivariate probit models*, *Biometrika* 82, No. 2 (1998)
- [6] J.F. Engel, R.D. Blackwell and P.W. Miniard, *Consumer Behavior* (The Dryden Press, Forth Worth, 1995)
- [7] S. Fournier, S. Dobscha and D.G. Mick, *Preventing the Premature Death of Relationship Marketing*, *Harvard Business Review* 76, No. 1 (1998)
- [8] P.H. Franses, *A Test for the Hit Rate in Binary Response Models*, *International Journal of Market Research* 42, No. 2 (2000)
- [9] A.W.H. Grant and L.A. Schlesinger, *Realize Your Customers Full Profit Potential*, *Harvard Business Review* 73, No. 5 (September-October, 1995)

- [10] W.H. Greene, *Econometric Analysis*, 3rd edition (Prentice Hall, New Jersey, 1997)
- [11] J.R. Hauser and G.L. Urban, The Value Priority Hypotheses for Consumer Budget Plans, *Journal of Consumer Research* 12, No. 4 (1986)
- [12] J.L. Heskett, W.E. Sasser and L.A. Schlesinger, *The Service Profit Chain* (Free Press, New York, 1997)
- [13] J.C. Hoekstra and K.R.E. Huizingh, The Lifetime Value Concept in Customer Based Marketing, *Journal of Market Focused Management* 3, No. 3/4 (1999)
- [14] W.A. Kamakura, S.N. Ramaswami and R.K. Srivastava, Applying Latent Trait Analysis in the Evaluation of Prospects for Cross-selling of Financial Services, *International Journal of Research in Marketing* 8, No. 4 (1991)
- [15] B.D. Kim and S.O. Kim, Measuring Upselling Potential of Life Insurance Customers: Application of Stochastic Frontier Model, *Journal of Interactive Marketing* 13, No. 4 (1999)
- [16] R.M. Morgan and S.D. Hunt, The Commitment-Trust Theory of Relationship Marketing, *Journal of Marketing* 58, No. 3 (July, 1994)
- [17] D. Peppers and M. Rogers, *The One to One Manager: Real-World Lessons in Customer Relationship Management* (Doubleday, New York, 1999)
- [18] F.F. Reichheld, *Loyalty Based Management* (Harvard Business School Press, Boston, 1996)
- [19] R.T. Rust, V.A. Zeithaml and K. Lemon, *Driving Customer Equity: How Customer Lifetime Value is Reshaping Corporate Strategy* (The Free Press, New York, 2000)
- [20] D.C. Schmittlein and R.A. Peterson, Customer Base Analysis: An Industrial Purchase Process Application, *Marketing Science* 13, No. 1 (1994)

- [21] P.N. Spring, P.C. Verhoef, J.C. Hoekstra and P.S.H. Leeflang, The Commercial Use of Segmentation and Predictive Modeling Techniques for Database Marketing, Working Paper (University of Groningen, 2000)
- [22] V.A. Zeithaml, Service Quality, Profitability and the Economic Worth of Customers, Journal of the Academy of Marketing Science 28, No. 1 (2000)

Figure 1: Segmentation with Current Value and Customer Potential

		CURRENT VALUE	
		<i>Low</i>	<i>High</i>
CUSTOMER POTENTIAL VALUE	<i>High</i>	II	IV
	<i>Low</i>	I	III

Figure 2: Conceptual Model Underlying DSS

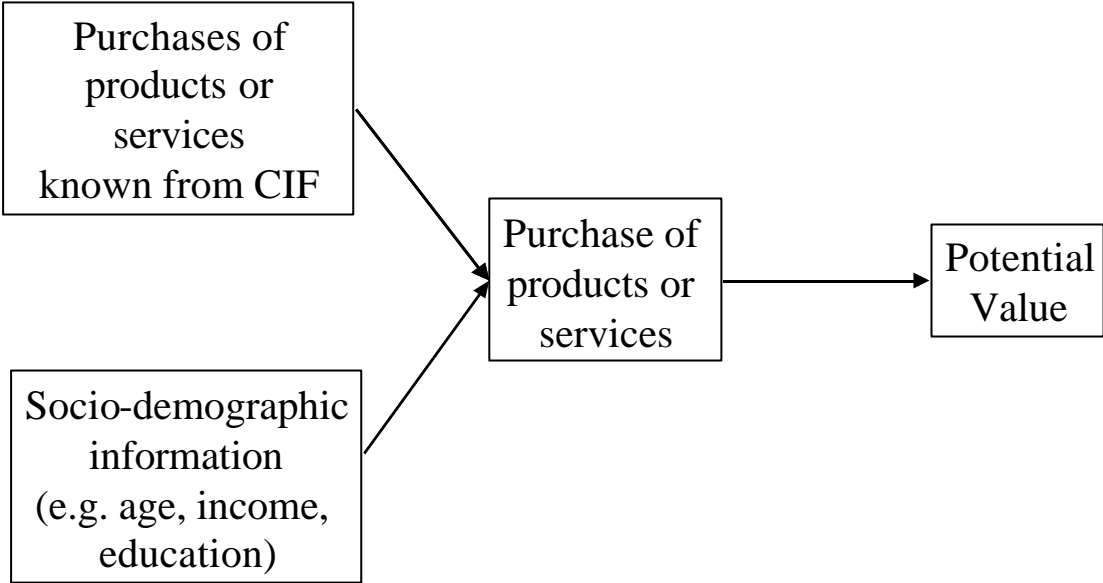


Table 1: Ownership rates for the 12 insurance types (N=1612).

Insurance	Ownership	Insurance	Ownership	Insurance	Ownership
	Rate (%)		Rate (%)		Rate (%)
1	98.7	5	88.8	9	57.1
2	98.0	6	71.0	10	50.7
3	97.6	7	65.0	11	42.3
4	96.3	8	63.8	12	40.4

Table 2. The fraction of correct predictions for our models and a naive model, with p values from the test of Franses (2000).

<i>Insurance type</i>	<i>Univariate Probit</i>	<i>Multivariate Probit</i>	<i>Naive model</i>
5	0.894 (0.000)	0.899 (0.000)	0.892
6	0.758 (0.000)	0.755 (0.000)	0.733
7	0.651 (0.001)	0.657 (0.000)	0.658
8	0.621 (0.268)	0.621 (0.224)	0.635
9	0.655 (0.000)	0.650 (0.000)	0.547
10	0.503 (0.463)	0.503 (0.411)	0.464
11	0.556 (0.457)	0.542 (0.721)	0.577
12	0.634 (0.000)	0.636 (0.000)	0.570

Table 3: Mean absolute prediction errors (MAPE) for our models and a naive model.

	Univariate	Multivariate	Regression	Naïve model
	probit	probit	model	
MAPE Profitability	19.5%	19.4%	19.4%	20.5%

Note: $MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - \hat{Y}_i|}{Y_i} * 100\%$

Table 4: Actual indexed profitability for different customer segments and the percentage of customers correctly classified

		Probit Choice	Multivariate Probit	Regression model	Probit Segment
High potential value segment	Mean	104.0	104.0	104.0	101.6
	Std. Dev.	20.1	19.7	19.9	20.3
	N	311	308	318	326
Low potential value segment	Mean	97.5	96.0	95.6	98.1
	Std. Dev.	21.4	21.2	20.9	21.3
	N	301	304	294	286
% correctly classified		53.1%	54.6%	55.9%	51.6%

Table 5: Actual indexed profitability for different customer segments

		Current Value	
		Low	High
High Potential Value	Mean	103.7	Mean 106.5
	Std. Dev.	19.9	Std. Dev. 20.2
	N	183	N 135
Low Potential Value	Mean	94.1	Mean 97.9
	Std. Dev.	21.3	Std. Dev. 20.9
	N	116	N 178

Appendix A
Sample Characteristics (N=1612)

Variable	Mean	Minimum	Maximum
Age	38.39	16	56
Female	0.414	0	1
Own house	0.803	0	1
Own cars	0.886	0	1
Number of Children	1.213	0	6
Single	0.146	0	1
Relationship duration	6.911	0	34
Education categories:			
Low education	0.105	0	1
Intermediate education	0.378	0	1
Higher education	0.518	0	1
Income categories:			
Low income	0.084	0	1
Middle income	0.330	0	1
High income	0.213	0	1
Very high income	0.226	0	1
Income Unknown	0.149	0	1

Biographies of authors

Peter C. Verhoef is a doctoral candidate in the Department of Marketing and Organization of the Faculty of Economics at the Erasmus University Rotterdam. In his dissertation he investigates the value of customer relationships. His other research interests include electronic commerce, waiting times and private labels. His work has been published in the European Journal of Marketing, Journal of Market Focused Management and Journal of Retailing and Consumer Services.

Bas Donkers received his PhD from Tilburg University in 2000. He currently has a post doc position at the Erasmus University Rotterdam. His research interests are in the modeling of individual decision making and, more general, the area of applied econometrics. His work has been published in the Journal of Risk and Uncertainty, Journal of Economic Psychology and Review of Income and Wealth.

Notes

¹ The functional form assumed for these correlations is different for the univariate and multivariate probit model.

² Prediction methods that add the profitability of all insurances the customer is predicted to own or predicting the profitability of the portfolio that is most likely according to the multivariate probit model, result in lower MAPEs.

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES *RESEARCH IN MANAGEMENT*

Publications in the Report Series Research* in Management

Impact of the Employee Communication and Perceived External Prestige on Organizational Identification

Ale Smidts, Cees B.M. van Riel & Ad Th.H. Pruyn

ERS-2000-01-MKT

Critical Complexities, from marginal paradigms to learning networks

Slawomir Magala

ERS-2000-02-ORG

Forecasting Market Shares from Models for Sales

Dennis Fok & Philip Hans Franses

ERS-2000-03-MKT

A Greedy Heuristic for a Three-Level Multi-Period Single-Sourcing Problem

H. Edwin Romeijn & Dolores Romero Morales

ERS-2000-04-LIS

Integer Constraints for Train Series Connections

Rob A. Zuidwijk & Leo G. Kroon

ERS-2000-05-LIS

Competitive Exception Learning Using Fuzzy Frequency Distribution

W-M. van den Bergh & J. van den Berg

ERS-2000-06-LIS

Start-Up Capital: Differences Between Male and Female Entrepreneurs, 'Does Gender Matter?'

Ingrid Verheul & Roy Thurik

ERS-2000-07-STR

The Effect of Relational Constructs on Relationship Performance: Does Duration Matter?

Peter C. Verhoef, Philip Hans Franses & Janny C. Hoekstra

ERS-2000-08-MKT

Marketing Cooperatives and Financial Structure: a Transaction Costs Economics Analysis

George W.J. Hendrikse & Cees P. Veerman

ERS-2000-09-ORG

* ERIM Research Programs:

LIS Business Processes, Logistics and Information Systems

ORG Organizing for Performance

MKT Decision Making in Marketing Management

F&A Financial Decision Making and Accounting

STR Strategic Renewal and the Dynamics of Firms, Networks and Industries

A Marketing Co-operative as a System of Attributes: A case study of VTN/The Greenery International BV,
Jos Bijman, George Hendrikse & Cees Veerman
ERS-2000-10-ORG

Evaluating Style Analysis
Frans A. De Roon, Theo E. Nijman & Jenke R. Ter Horst
ERS-2000-11-F&A

From Skews to a Skewed-t: Modelling option-implied returns by a skewed Student-t
Cyriel de Jong & Ronald Huisman
ERS-2000-12-F&A

Marketing Co-operatives: An Incomplete Contracting Perspective
George W.J. Hendrikse & Cees P. Veerman
ERS-2000-13- ORG

Models and Algorithms for Integration of Vehicle and Crew Scheduling
Richard Freling, Dennis Huisman & Albert P.M. Wagelmans
ERS-2000-14-LIS

Ownership Structure in Agrifood Chains: The Marketing Cooperative
George W.J. Hendrikse & W.J.J. (Jos) Bijman
ERS-2000-15-ORG

Managing Knowledge in a Distributed Decision Making Context: The Way Forward for Decision Support Systems
Sajda Qureshi & Vlatka Hlupic
ERS-2000-16-LIS

Organizational Change and Vested Interests
George W.J. Hendrikse
ERS-2000-17-ORG

Strategies, Uncertainty and Performance of Small Business Startups
Marco van Gelderen, Michael Frese & Roy Thurik
ERS-2000-18-STR

Creation of Managerial Capabilities through Managerial Knowledge Integration: a Competence-Based Perspective
Frans A.J. van den Bosch & Raymond van Wijk
ERS-2000-19-STR

Adaptiveness in Virtual Teams: Organisational Challenges and Research Direction
Sajda Qureshi & Doug Vogel
ERS-2000-20-LIS

Currency Hedging for International Stock Portfolios: A General Approach
Frans A. de Roon, Theo E. Nijman & Bas J.M. Werker
ERS-2000-21-F&A

Transition Processes towards Internal Networks: Differential Paces of Change and Effects on Knowledge Flows at Rabobank Group
Raymond A. van Wijk & Frans A.J. van den Bosch
ERS-2000-22-STR

Assessment of Sustainable Development: a Novel Approach using Fuzzy Set Theory
A.M.G. Cornelissen, J. van den Berg, W.J. Koops, M. Grossman & H.M.J. Udo
ERS-2000-23-LIS

Creating the N-Form Corporation as a Managerial Competence

Raymond vanWijk & Frans A.J. van den Bosch

ERS-2000-24-STR

Competition and Market Dynamics on the Russian Deposits Market

Piet-Hein Admiraal & Martin A. Carree

ERS-2000-25-STR

Interest and Hazard Rates of Russian Saving Banks

Martin A. Carree

ERS-2000-26-STR

The Evolution of the Russian Saving Bank Sector during the Transition Era

Martin A. Carree

ERS-2000-27-STR

Is Polder-Type Governance Good for You? Laissez-Faire Intervention, Wage Restraint, And Dutch Steel

Hans Schenk

ERS-2000-28-ORG

Foundations of a Theory of Social Forms

László Pólos, Michael T. Hannan & Glenn R. Carroll

ERS-2000-29-ORG

Reasoning with partial Knowledge

László Pólos & Michael T. Hannan

ERS-2000-30-ORG

Applying an Integrated Approach to Vehicle and Crew Scheduling in Practice

Richard Freling, Dennis Huisman & Albert P.M. Wagelmans

ERS-2000-31-LIS

Informants in Organizational Marketing Research: How Many, Who, and How to Aggregate Response?

Gerrit H. van Bruggen, Gary L. Lilien & Manish Kacker

ERS-2000-32-MKT

The Powerful Triangle of Marketing Data, Managerial Judgment, and Marketing Management Support Systems

Gerrit H. van Bruggen, Ale Smidts & Berend Wierenga

ERS-2000-33-MKT

The Strawberry Growth Underneath the Nettle: The Emergence of Entrepreneurs in China

Barbara Krug & László Pólos

ERS-2000-34-ORG

Consumer Perception and Evaluation of Waiting Time: A Field Experiment

Gerrit Antonides, Peter C. Verhoef & Marcel van Aalst

ERS-2000-35-MKT

Trading Virtual Legacies

Slawomir Magala

ERS-2000-36-ORG

Broker Positions in Task-Specific Knowledge Networks: Effects on Perceived Performance and Role Stressors in an Account Management System

David Dekker, Frans Stokman & Philip Hans Franses

ERS-2000-37-MKT

An NPV and AC analysis of a stochastic inventory system with joint manufacturing and remanufacturing
Erwin van der Laan
ERS-2000-38-LIS

Generalizing Refinement Operators to Learn Prenex Conjunctive Normal Forms
Shan-Hwei Nienhuys-Cheng, Wim Van Laer, Jan Ramon & Luc De Raedt
ERS-2000-39-LIS

Classification and Target Group Selection bases upon Frequent Patterns
Wim Pijls & Rob Potharst
ERS-2000-40-LIS

New Entrants versus Incumbents in the Emerging On-Line Financial Services Complex
Manuel Hensmans, Frans A.J. van den Bosch & Henk W. Volberda
ERS-2000-41-STR

Modeling Unobserved Consideration Sets for Household Panel Data
Erjen van Nierop, Richard Paap, Bart Bronnenberg, Philip Hans Franses & Michel Wedel
ERS-2000-42-MKT

The Interdependence between Political and Economic Entrepreneurship
ERS-2000-43-ORG
Barbara Krug

Ties that bind: The Emergence of Entrepreneurs in China
Barbara Krug
ERS-2000-44-ORG

What's New about the New Economy? Sources of Growth in the Managed and Entrepreneurial Economies
David B. Audretsch and A. Roy Thurik
ERS-2000-45-STR

Human Resource Management and Performance: Lessons from the Netherlands
Paul Boselie, Jaap Paauwe & Paul Jansen
ERS-2000-46-ORG

Average Costs versus Net Present Value: a Comparison for Multi-Source Inventory Models
Erwin van der Laan & Ruud Teunter
ERS-2000-47-LIS

A Managerial Perspective on the Logic of Increasing Returns
Erik den Hartigh, Fred Langerak & Harry Commandeur
ERS-2000-48-MKT

Fuzzy Modeling of Client Preference in Data-Rich Marketing Environments
Magne Setnes & Uzay Kaymak
ERS-2000-49-LIS

The Mediating Effect of NPD-Activities and NPD-Performance on the Relationship between Market Orientation and Organizational Performance
Fred Langerak, Erik Jan Hultink & Henry S.J. Robben
ERS-2000-50-MKT

Extended Fuzzy Clustering Algorithms
Uzay Kaymak & Magne Setnes
ERS-2000-51-LIS

Sensemaking from actions: Deriving organization members' means and ends from their day-to-day behavior

ERS-2000-52-MKT

Johan van Rekom, Cees B.M. van Riel & Berend Wierenga

Mining frequent itemsets in memory-resident databases

ERS-2000-53-LIS

Wim Pijls & Jan C. Bioch

Possible futures for the HR function in different market

ERS-2000-54-ORG

Roger Williams, Jaap Paauwe & Anne Keegan

Quantity versus Quality in Project Based Learning Practices

ERS-2000-55-ORG

Anne Keegan & J. Rodney Turner

Crew Scheduling for Netherlands Railways. "Destination: Customer"

ERS-2000-56-LIS

Leo Kroon & Matteo Fischetti

The Management of Innovation in Project Based Firms

ERS-2000-57-ORG

Anne Keegan and J. Rodney Turner

Learning by Experience in the Project-Based Organization

ERS-2000-58-ORG

J. Rodney Turner, Anne Keegan & Lynn Crawford